USING GENOMICS, METAGENOMICS AND OTHER "OMICS" TO ASSESS VALUABLE MICROBIAL ECOSYSTEM SERVICES AND NOVEL BIOTECHNOLOGICAL APPLICATIONS

EDITED BY: Diana Elizabeth Marco and Florence Abram
PUBLISHED IN: Frontiers in Microbiology and Frontiers in Bioengineering and Biotechnology

frontiers Research Topics



frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88945-814-1
DOI 10.3389/978-2-88945-814-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

USING GENOMICS, METAGENOMICS AND OTHER "OMICS" TO ASSESS VALUABLE MICROBIAL ECOSYSTEM SERVICES AND NOVEL BIOTECHNOLOGICAL APPLICATIONS

Topic Editors:
Diana Elizabeth Marco, Universidad Nacional de Córdoba and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
Florence Abram, National University of Ireland Galway, Ireland

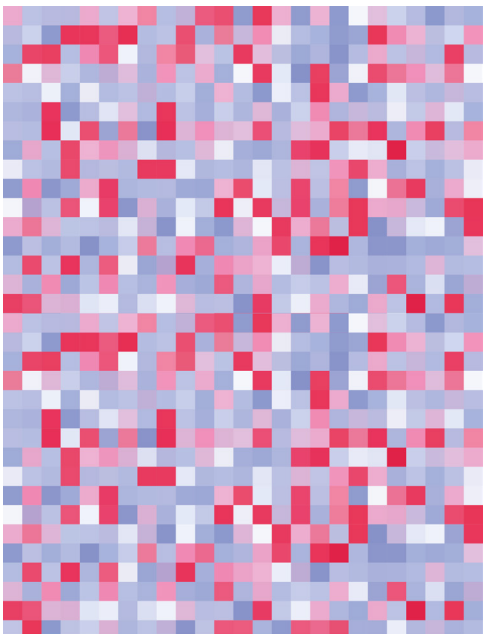


Image: tashechka/Shutterstock.com

Most ecosystem services and goods human populations use and consume are provided by microbial populations and communities. Indeed, numerous provisioning services (e.g. food and enzymes for industrial processes), regulating services (e.g. water quality, contamination alleviation and biological processes such as plant-microbial symbioses), and supporting services (e.g. nutrient cycling, agricultural production and biodiversity) are mediated by microbes.

The fast development of metagenomics and other meta-omics technologies is expanding our understanding of microbial diversity, ecology, evolution and functioning. This enhanced knowledge directly translates into the emergence of new applications in an unlimited variety of areas across all microbial ecosystem services and goods. The varied topics addressed in this Research Topic include the development of innovative industrial processes, the discovery of novel natural products, the advancement of new agricultural methods, the amelioration of negative effects of productive or natural microbiological processes, as well as food security and human health, and archeological conservation.

The articles compiled provide an updated, high-quality overview of current work in the field. This body of research makes a valuable contribution to the understanding of microbial ecosystem services, and expands the horizon for finding and developing new and more efficient biotechnological applications.

Citation: Marco, D. E., Abram, F., eds. (2019). Using Genomics, Metagenomics and Other "Omics" to Assess Valuable Microbial Ecosystem Services and Novel Biotechnological Applications. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88945-814-1

Table of Contents

09 ***Editorial: Using Genomics, Metagenomics and Other "Omics" to Assess Valuable Microbial Ecosystem Services and Novel Biotechnological Applications***
Diana E. Marco and Florence Abram

INDUSTRIAL, PRODUCTIVE AND AGRICULTURAL PROCESSES

14 ***Assessing the Effect of Pretreatments on the Structure and Functionality of Microbial Communities for the Bioconversion of Microalgae to Biogas***
Olivia Córdova, Rolando Chamy, Lorna Guerrero and Aminael Sánchez-Rodríguez

25 ***Linking Microbial Community Structure and Function During the Acidified Anaerobic Digestion of Grass***
Aoife Joyce, Umer Z. Ijaz, Corine Nzeteu, Aoife Vaughan, Sally L. Shirran, Catherine H. Botting, Christopher Quince, Vincent O'Flaherty and Florence Abram

38 ***Metabolic Adaptation of Methanogens in Anaerobic Digesters Upon Trace Element Limitation***
Babett Wintsche, Nico Jehmlich, Denny Popp, Hauke Harms and Sabine Kleinsteuber

48 ***Linking the Effect of Antibiotics on Partial-Nitritation Biofilters: Performance, Microbial Communities and Microbial Activities***
Alejandro Gonzalez-Martinez, Alejandro Margareto, Alejandro Rodriguez-Sanchez, Chiara Pesciaroli, Silvia Diaz-Cruz, Damia Barcelo and Riku Vahala

64 ***Characterization of Bacterial and Fungal Community Dynamics by High-Throughput Sequencing (HTS) Metabarcoding During Flax Dew-Retting***
Christophe Djemiel, Sébastien Grec and Simon Hawkins

80 ***Sequencing of the Cheese Microbiome and its Relevance to Industry***
Bhagya. R. Yeluri Jonnala, Paul L. H. McSweeney, Jeremiah J. Sheehan and Paul D. Cotter

92 **Zygosaccharomyces bailii** ***is a Potential Producer of Various Flavor Compounds in Chinese*** **Maotai-***Flavor Liquor Fermentation***
Yan Xu, Yan Zhi, Qun Wu, Rubing Du and Yan Xu

101 ***A Novel RAYM_RS09735/RAYM_RS09740 Two-Component Signaling System Regulates Gene Expression and Virulence in Riemerella anatipestifer***
Ying Wang, Ti Lu, Xuehuan Yin, Zutao Zhou, Shaowen Li, Mei Liu, Sishun Hu, Dingren Bi and Zili Li

114 ***Studying the Differences of Bacterial Metabolome and Microbiome in the Colon Between Landrace and Meihua Piglets***
Shijuan Yan, Cui Zhu, Ting Yu, Wenjie Huang, Jianfeng Huang, Qian Kong, Jingfang Shi, Zhongjian Chen, Qinjian Liu, Shaolei Wang, Zongyong Jiang and Zhuang Chen

128 ***Alfalfa Intervention Alters Rumen Microbial Community Development in Hu Lambs During Early Life***
Bin Yang, Jiaqing Le, Peng Wu, Jianxin Liu, Le L. Guan and Jiakun Wang

141 ***Exploring the Spatial-Temporal Microbiota of Compound Stomachs in a Pre-Weaned Goat Model***
Yu Lei, Ke Zhang, Mengmeng Guo, Guanwei Li, Chao Li, Bibo Li, Yuxin Yang, Yulin Chen and Xiaolong Wang

152 ***Comparative Genomics Reveals Evidence of Genome Reduction and High Extracellular Protein Degradation Potential in Kangiella***
Jiahua Wang, Ye Lu, Muhammad Z. Nawaz and Jun Xu

165 ***Recovering Genomics Clusters of Secondary Metabolites From Lakes Using Genome-Resolved Metagenomics***
Rafael R. C. Cuadrat, Danny Ionescu, Alberto M. R. Dávila and Hans-Peter Grossart

178 ***Functional Diversity of Transcriptional Regulators in the Cyanobacterium Synechocystis sp. PCC 6803***
Mengliang Shi, Xiaoqing Zhang, Guangsheng Pei, Lei Chen and Weiwen Zhang

188 ***Diversity of Gene Clusters for Polyketide and Nonribosomal Peptide Biosynthesis Revealed by Metagenomic Analysis of the Yellow Sea Sediment***
Yongjun Wei, Lei Zhang, Zhihua Zhou and Xing Yan

200 ***Genomic Reconstruction of Carbohydrate Utilization Capacities in Microbial-Mat Derived Consortia***
Semen A. Leyn, Yukari Maezato, Margaret F. Romine and Dmitry A. Rodionov

217 ***Genetic Diversity of Nitrogen-Fixing and Plant Growth Promoting* Pseudomonas *Species Isolated From Sugarcane Rhizosphere***
Hai-Bi Li, Rajesh K. Singh, Pratiksha Singh, Qi-Qi Song, Yong-Xiu Xing, Li-Tao Yang and Yang-Rui Li

237 ***Long-term Fertilization Structures Bacterial and Archaeal Communities Along Soil Depth Gradient in a Paddy Soil***
Yunfu Gu, Yingyan Wang, Sheng'e Lu, Quanju Xiang, Xiumei Yu, Ke Zhao, Likou Zou, Qiang Chen, Shihua Tu and Xiaoping Zhang

252 ***Maize Endophytic Bacterial Diversity as Affected by Soil Cultivation History***
David Correa-Galeote, Eulogio J. Bedmar and Gregorio J. Arone

261 ***Long-Term Rock Phosphate Fertilization Impacts the Microbial Communities of Maize Rhizosphere***
Ubiana C. Silva, Julliane D. Medeiros, Laura R. Leite, Daniel K. Morais, Sara Cuadros-Orellana, Christiane A. Oliveira, Ubiraci G. de Paula Lana, Eliane A. Gomes and Vera L. Dos Santos

272 ***Microbial Community and Functional Structure Significantly Varied Among Distinct Types of Paddy Soils but Responded Differently Along Gradients of Soil Depth Layers***
Ren Bai, Jun-Tao Wang, Ye Deng, Ji-Zheng He, Kai Feng and Li-Mei Zhang

288 ***Metagenomic Profiling of Soil Microbes to Mine Salt Stress Tolerance Genes***
Vasim Ahmed, Manoj K. Verma, Shashank Gupta, Vibha Mandhan and Nar S. Chauhan

ENVIRONMENTAL DETOXIFICATION AND BIOREMEDIATION

299 ***A Genomic Outlook on Bioremediation: The Case of Arsenic Removal***
Frédéric Plewniak, Simona Crognale, Simona Rossetti and Philippe N. Bertin

307 ***Proteomic Analysis of 2,4,6-Trinitrotoluene Degrading Yeast Yarrowia lipolytica***
Irina V. Khilyas, Guenter Lochnit and Olga N. Ilinskaya

318 ***Metagenomic Analysis of a Biphenyl-Degrading Soil Bacterial Consortium Reveals the Metabolic Roles of Specific Populations***
Daniel Garrido-Sanz, Javier Manzano, Marta Martín, Miguel Redondo-Nieto and Rafael Rivilla

331 ***Opportunistic Bacteria Dominate the Soil Microbiome Response to Phenanthrene in a Microcosm-Based Study***
Sean Storey, Mardiana Mohd Ashaari, Nicholas Clipson, Evelyn Doyle and Alexandre B. de Menezes

344 ***Taxon-Function Decoupling as an Adaptive Signature of Lake Microbial Metacommunities Under a Chronic Polymetallic Pollution Gradient***
Bachar Cheaib, Malo Le Boulch, Pierre-Luc Mercier and Nicolas Derome

361 ***pH Stress-Induced Cooperation Between* Rhodococcus ruber *YYL and Bacillus cereus MLY1 in Biodegradation of Tetrahydrofuran***
Zubi Liu, Zhixing He, Hui Huang, Xuebin Ran, Adebanjo O. Oluwafunmilayo and Zhenmei Lu

372 ***The Role of Enriched Microbial Consortium on Iron-Reducing Bioaugmentation in Sediments***
Yuanyuan Pan, Xunan Yang, Meiying Xu and Guoping Sun

382 ***Enhancing Nitrate Removal From Freshwater Pond by Regulating Carbon/Nitrogen Ratio***
Rong Chen, Min Deng, Xugang He and Jie Hou

391 ***Predicting Species-Resolved Macronutrient Acquisition During Succession in a Model Phototrophic Biofilm Using an Integrated 'Omics Approach***
Stephen R. Lindemann, Jennifer M. Mobberley, Jessica K. Cole, L. M. Markillie, Ronald C. Taylor, Eric Huang, William B. Chrisler, H. S. Wiley, Mary S. Lipton, William C. Nelson, James K. Fredrickson and Margaret F. Romine

406 ***Niche Partitioning of the N Cycling Microbial Community of an Offshore Oxygen Deficient Zone***
Clara A. Fuchsman, Allan H. Devol, Jaclyn K. Saunders, Cedar McKay and Gabrielle Rocap

424 ***FixK$_2$ is the Main Transcriptional Activator of* Bradyrhizobium diazoefficiens nosRZDYFLX *Genes in Response to Low Oxygen***
María J. Torres, Emilio Bueno, Andrea Jiménez-Leiva, Juan J. Cabrera, Eulogio J. Bedmar, Socorro Mesa and María J. Delgado

440 ***Comparative Analysis of the Microbiota Between Sheep Rumen and Rabbit Cecum Provides New Insight Into Their Differential Methane Production***
Lan Mi, Bin Yang, Xialu Hu, Yang Luo, Jianxin Liu, Zhongtang Yu and Jiakun Wang

ECOSYSTEM MONITORING AND SPECIES CONSERVATION

454 ***Amplicon-Based Sequencing of Soil Fungi From Wood Preservative Test Sites***
Grant T. Kirker, Amy B. Bishell, Michelle A. Jusino, Jonathan M. Palmer, William J. Hickey and Daniel L. Lindner

470 ***Quantitative Detection of Active Vibrios Associated With White Plague Disease in Mussismilia braziliensis Corals***
Luciane A. Chimetto Tonon, Janelle R. Thompson, Ana P. B. Moreira, Gizele D. Garcia, Kevin Penn, Rachelle Lim, Roberto G. S. Berlinck, Cristiane C. Thompson and Fabiano L. Thompson

480 ***Engineering Strategies to Decode and Enhance the Genomes of Coral Symbionts***
Rachel A. Levin, Christian R. Voolstra, Shobhit Agrawal, Peter D. Steinberg, David J. Suggett and Madeleine J. H. van Oppen

491 ***The Divergence in Bacterial Components Associated With Bactrocera dorsalis Across Developmental Stages***
Xiaofeng Zhao, Xiaoyu Zhang, Zhenshi Chen, Zhen Wang, Yongyue Lu and Daifeng Cheng

501 ***Functional Characteristics of the Flying Squirrel's Cecal Microbiota Under a Leaf-Based Diet, Based on Multiple Meta-Omic Profiling***
Hsiao-Pei Lu, Po-Yu Liu, Yu-bin Wang, Ji-Fan Hsieh, Han-Chen Ho, Shiao-Wei Huang, Chung-Yen Lin, Chih-hao Hsieh and Hon-Tsen Yu

514 ***Differential Proteomic Profiles of* Pleurotus ostreatus *in Response to Lignocellulosic Components Provide Insights into Divergent Adaptive Mechanisms***
Qiuyun Xiao, Fuying Ma, Yan Li, Hongbo Yu, Chengyun Li and Xiaoyu Zhang

FOOD QUALITY AND SAFETY

528 ***Investigation of Citrinin and Pigment Biosynthesis Mechanisms in* Monascus purpureus *by Transcriptomic Analysis***
Bin Liang, Xin-Jun Du, Ping Li, Chan-Chan Sun and Shuo Wang

539 ***Identification and Characterization of* als *Genes Involved in D-Allose Metabolism in Lineage II Strain of Listeria monocytogenes***
Lu Zhang, Yan Wang, Dongxin Liu, Lijuan Luo, Yi Wang and Changyun Ye

HUMAN HEALTH AND DISEASES

548 ***Distinct Microbial Signatures Associated With Different Breast Cancer Types***
Sagarika Banerjee, Tian Tian, Zhi Wei, Natalie Shih, Michael D. Feldman, Kristen N. Peck, Angela M. DeMichele, James C. Alwine and Erle S. Robertson

568 ***Oropharyngeal and Sputum Microbiomes are Similar Following Exacerbation of Chronic Obstructive Pulmonary Disease***
Hai-Yue Liu, Shi-Yu Zhang, Wan-Ying Yang, Xiao-Fang Su, Yan He, Hong-Wei Zhou and Jin Su

579 ***Oral Administration of Recombinant Saccharomyces boulardii Expressing Ovalbumin-CPE Fusion Protein Induces Antibody Response in Mice***
Ghasem Bagherpour, Hosnie Ghasemi, Bahare Zand, Najmeh Zarei, Farzin Roohvand, Esmat M. Ardakani, Mohammad Azizi and Vahid Khalaj

588 ***Meta-Analysis of Aedes aegypti Expression Datasets: Comparing Virus Infection and Blood-Fed Transcriptomes to Identify Markers of Virus Presence***
Kiyoshi Ferreira Fukutani, José Irahe Kasprzykowski, Alexandre Rossi Paschoal, Matheus de Souza Gomes, Aldina Barral, Camila I. de Oliveira, Pablo Ivan Pereira Ramos and Artur Trancoso Lopo de Queiroz

PRESERVATION OF MUSEUM OBJECTS AND ARCHEOLOGICAL REMAINS

601 ***Microbial Community Analyses of the Deteriorated Storeroom Objects in the Tianjin Museum Using Culture-Independent and Culture-Dependent Approaches***
Zijun Liu, Yanhong Zhang, Fengyu Zhang, Cuiting Hu, Genliang Liu and Jiao Pan

613 ***Identification of Fungal Communities Associated With the Biodeterioration of Waterlogged Archeological Wood in a Han Dynasty Tomb in China***
Zijun Liu, Yu Wang, Xiaoxuan Pan, Qinya Ge, Qinglin Ma, Qiang Li, Tongtong Fu, Cuiting Hu, Xudong Zhu and Jiao Pan

NEW METHODOLOGICAL DEVELOPMENTS

622 ***PCR Primer Design for 16S rRNAs for Experimental Horizontal Gene Transfer Test in Escherichia coli***
Kentaro Miyazaki, Mitsuharu Sato and Miyuki Tsukuda

629 ***Clean Low-Biomass Procedures and Their Application to Ancient Ice Core Microorganisms***
Zhi-Ping Zhong, Natalie E. Solonenko, Maria C. Gazitúa, Donald V. Kenny, Ellen Mosley-Thompson, Virginia I. Rich, James L. Van Etten, Lonnie G. Thompson and Matthew B. Sullivan

644 ***viGEN: An Open Source Pipeline for the Detection and Quantification of Viral RNA in Human Tumors***
Krithika Bhuvaneshwar, Lei Song, Subha Madhavan and Yuriy Gusev

657 ***Intriguing Interaction of Bacteriophage-Host Association: An Understanding in the Era of Omics***
Krupa M. Parmar, Saurabh L. Gaikwad, Prashant K. Dhakephalkar, Ramesh Kothari and Ravindra Pal Singh

frontiers in Microbiology

EDITORIAL
published: 12 February 2019
doi: 10.3389/fmicb.2019.00151

Editorial: Using Genomics, Metagenomics and Other "Omics" to Assess Valuable Microbial Ecosystem Services and Novel Biotechnological Applications

Diana E. Marco[1]* *and Florence Abram*[2]*

[1] Faculty of Exact, Physical and Biological Sciences, CONICET, Córdoba National University, Córdoba, Argentina, [2] Functional Environmental Microbiology, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland

Keywords: genomics, metagenomics, meta-omics, microbial ecology, ecosystem services, biotechnology, industrial processes, natural products

OPEN ACCESS

Edited by:
Jean Armengaud, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), France

Reviewed by:
Christopher Staley, University of Minnesota Twin Cities, United States

***Correspondence:**
Diana E. Marco
dmarco@agro.unc.edu.ar
Florence Abram
florence.abram@nuigalway.ie

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *07 November 2018*
Accepted: *22 January 2019*
Published: *12 February 2019*

Citation:
Marco DE and Abram F (2019) Editorial: Using Genomics, Metagenomics and Other "Omics" to Assess Valuable Microbial Ecosystem Services and Novel Biotechnological Applications. Front. Microbiol. 10:151. doi: 10.3389/fmicb.2019.00151

Editorial on the Research Topic

Using Genomics, Metagenomics and Other "Omics" to Assess Valuable Microbial Ecosystem Services and Novel Biotechnological Applications

Most ecosystem services [defined as the benefits people obtain from ecosystems, Millennium Ecosystem Assessment (2005)] and goods human populations rely on are provided by microbial populations and communities. Indeed, numerous provisioning services (e.g., food and enzymes for industrial processes); regulating services (e.g., water quality, contamination alleviation, and biological processes such as plant-microbial symbioses); and supporting services (e.g., nutrient cycling, agricultural production, and biodiversity), are mediated by microbes. Thus, to preserve and protect these ecosystems, currently facing climate change and anthropologic activities-related challenges, as well as to harness some of the naturally microbial occurring processes, a deep understanding of the microbiology underpinning ecosystem functioning is required. To this end, experimental strategies deploying metagenomics, and other meta-omics technologies have been implemented. The resulting enhanced knowledge directly translates into the emergence of new applications in an unlimited variety of areas across all microbial ecosystem services and goods. To compile an updated overview of these new developments, we called for contributions to this Research Topic. Our expectations were amply rewarded by a high number of submitted manuscripts, for which we are greatly indebted to the authors.

We grouped the corresponding articles under different research themes as follows: (i) development of innovative industrial processes, (ii) discovery of novel natural products, (iii) advancement of new agricultural methods, (iv) amelioration of negative effects of productive or natural microbiological processes, (v) food security and human health, (vi) archeological conservation, and (vii) methodological developments. Some theme overlap is naturally anticipated due to the complex nature of the microbial processes investigated. For example, the improvement of agricultural practices not only presents an economic significance but also may contribute to the amelioration of some pressing environmental issues. The diversity of research showcased in this special issue demonstrates the tremendous potential of omics methods for advancing knowledge and underpinning the development of novel biotechnologies.

Microbial processes have been harnessed by human populations to produce many goods like food and beverages from ancient times. More recently, a variety of industries based on microbial

functions began to develop and today microbial-based applications from industrial enzymes to new drugs discovery are widespread (Okafor and Okeke, 2017). The application of omics methods to microbial communities driving milk fermentation for cheese production (Jonnala et al.), flax fiber production (Djemiel et al.), or nutrient acquisition from diet of goat, lamb and piglets (Chen et al.; Lei et al.; Yang et al.), allows for the formulation of new and more efficient production methods. Similarly, optimization avenues for industrial processes of economic and environmental importance can be designed based on knowledge gained from omics investigations.

For example, insights from metatranscriptomics analysis led to the optimisation of biogas production during the anaerobic digestion of microalgae, Córdova et al.; Joyce et al. used metaproteomics in conjunction with 16S rRNA profiling of DNA and cDNA to investigate the anaerobic digestion of perennial grass to produce second-generation biofuels. Microbial groups involved in the anaerobic process were identified, and the functional importance of Clostridia was highlighted by the authors. Wintsche et al. investigated the effects of trace elements depletion during the anaerobic digestion of distillers grains. Using *mcrA* gene amplicon sequencing and metaproteomics, activity shifts within methanogenic communities and the importance of *Methanosarcina* for reactor performance stabilization under critical conditions were highlighted. Tackling the issue of contaminant prevalence in wastewater, González-Martínez et al. investigated the effect of antibiotics exposure in autotrophic nitrogen removal systems for wastewater treatment combining 16S rRNA gene amplicon sequencing and metatranscriptomics.

Omics and meta-omics methodologies also hold great potential for facilitating the screening of secondary microbial metabolites for biotechnological and pharmaceutical industries (Wang et al.; Cuadrat et al.). Using metabolomics to study mutants of *Synechocystis* sp., Shi et al. identified the functions of several novel transcriptional regulators involved in response to diverse environmental stresses (heat, heavy metals), tolerance to ethanol and carbohydrate transport and metabolism. Using comparative genomics, Leyn et al. could unveil a hierarchical carbon flow from cyanobacteria to heterotrophs within benthic microbial-mat derived consortia. The authors reconstructed carbohydrate utilization pathways and identified glycohydrolytic enzymes, carbohydrate transporters and pathway-specific transcriptional regulators in the heterotrophic members of the bacterial consortia. This study revealed novel functional roles of 171 genes, and the utilization capabilities for 40 carbohydrates and their derivatives by the mat, opening the avenue for potential biotechnological applications.

Another important service provided by microbial communities is biofertilization, *via* the synthesis of plant nutrients or phytohormones, the mobilization of soil compounds, the protection of plants under stressful conditions, the defense against plant pathogens (García-Fraile et al., 2015), and biological nitrogen fixation which converts bio-unavailable N_2 gas to plant-available ammonia (NH_4^+) (Bedmar et al., 2013). All these microbial activities allow for a more environmentally friendly agriculture by diminishing the use of chemical fertilizers and toxic compounds. Understanding the structure and functioning of nitrogen-fixing and plant growth promoting microbiomes using omics methods led to improvements in crop management including sugarcane (Li et al.) rice, (Bai et al.; Gu et al.) and maize (Correa-Galeote et al.; Silva et al.), while avoiding or diminishing the use of artificial fertilizers. Using a functional metagenomics approach, Ahmed et al. screened for microbial salt tolerant genes that could be used for producing bioactive compounds to improve crop production under high saline conditions.

Other valuable services provided by microorganisms are environmental bioremediation and amelioration of negative consequences of soil and water contamination from different human activities (Shah et al., 2011). Although many microbes with a bioremediation potential have been isolated and characterized, in most cases a single microorganism cannot completely degrade a given pollutant or be effective in naturally prevalent *in situ* mixed contaminations (Dangi et al., 2018). A variety of pollutants such as arsenic, byphenil, phenanthrene, nitrate, and others can be degraded by microbial consortia, and in this context, the use of omics methodologies, can lead to a better understanding of the mechanisms underlying environmental detoxification. For example, Garrido-Sanz et al. used metagenomics to model the biodegradation of byphenil in contaminated soils and could assign reactions and pathways to specific bacterial groups. A metatrascriptomics study by Liu et al. showed that cooperation among strains (elicited by low soil pH) within microbial consortia improved tetrahydrofuran remediation efficiency compared to single microbial strains activity. Nitrate (NO_3^-) contamination (mainly from agriculture) of soils and freshwater bodies is a major environmental issue, causing many wildlife and human health problems. High NO_3^- water concentrations contributes to eutrophication and cause damage to the hemoglobin of aquatic organisms when under its reduced form, nitrite (NO_2^-). Drinking water containing high levels of NO_3^- and NO_2^- lead to methemoglobinemia (Greer and Shannon, 2005), while NO_3^- can be transformed in the digestive tract in carcinogenic nitrosamines (Craddock and Henderson, 1986). Denitrification is a key microbe mediated process of the nitrogen cycle occurring in low oxygen environments that converts NO_3^- to inert nitrogen gas (N_2) and thus ameliorates the effects of environmental nitrogen pollution. However, most denitrifying bacteria are not able to complete the pathway and emit nitrous oxide (N_2O) and nitric oxide (NO) as intermediate products. While NO contributes to acid rain, N_2O is a greenhouse gas with many fold greater potential for global warming compared with that of CO_2, and a main cause of ozone layer depletion (Bates et al., 2008). Thus, knowing the structure and functioning of denitrifying microbiomes is of paramount importance to device strategies for dealing with nitrate contamination, and the use omics methods is making an important contribution to this end. By using metagenomics to study genes involved in the denitrification pathway of freshwater ponds contaminated withNO_3^-, Chen et al. proposed a strategy for treating wastewater effluents by regulating the C/N ratio through the addition of extra organic carbon, to obtain higher denitrification efficiency. By using metagenomics,

metatranscriptomics, and metaproteomics, Lindemann et al. investigated the flow of nitrogen and other elements within phototrophic microbial consortia. Among other interesting results, the authors identified bacterial genomes encoding for nitrate and nitrite reductases needed for denitrification. They also reported that niche partitioning around nitrogen sources may structure the community when microorganisms directly compete for limiting phosphate. Using metagenomics and a phylogenetic placement approach, Fuchsman et al. characterized microbial genes for anoxic N cycling in metagenomes from samples recovered from an oxycline in the Eastern Tropical North Pacific oxygen deficient zone and as such provided an overview of the diverse microbial players driving this process. As previously noted, biofertilization offers an environmentally sustainable option for enhancing crop production. This holds particularly true for legumes, such as soybeans, which establish symbiotic relationships with rhizobia. However, while rhizobia, located in root nodules, fix atmospheric N_2 for the plants, thus reducing the need for fertilization, they also commonly perform incomplete denitrification. This results in N_2O production in root nodules and the corresponding emissions are increased during flooding (causing low oxygen conditions) (Tortosa et al., 2015). As an example of this situation, in Argentina, one of the main soybean-producers countries, about 20,000,000 ha were sown with soybean in 2016 (FAOSTAT), of which more than 700,000 ha were flooded (BCBA, 2017) (BCBA Report, 2017). By using the soybean endosymbiont *Bradyrhizobium diazoefficiens* and *in vitro* transcription (IVT) activation assays, Torres et al. were able to dissect the fine regulatory mechanisms involved in the control of the key steps in N_2O reduction to N_2 in response to low oxygen. The authors envisage that their findings should help to establish action plans for the development of practical strategies for N_2O emission mitigation from legume crops. Another important agricultural source of greenhouse gases is livestock production, especially ruminants, that produce methane (CH_4) a greenhouse gas 23 times more potent than CO_2 (IPCC, 2014). CH_4 is a byproduct of microbial feed fermentation inside the rumen, but is also produced by other non-ruminant herbivores like rabbits. Using 16S rRNA gene amplicon sequencing, Mi et al. could correlate microbial community structure differences with lower methane yields in rabbits when compared to sheep. Hydrogen utilization pathways were found to differ between the two animal species. Indeed, the authors reported a lower relative abundance of hydrogen-producing microbes and methanogens in rabbits, as well as an increased abundance of homoacetogens converting hydrogen to acetate.

A recent development in the application of omics technologies is in ecosystem health monitoring for species conservation and wildlife protection (Antwis et al., 2017). Using amplicon-based DNA sequencing of the internal transcribed spacer 1 (ITS1) region, Kirker et al. demonstrated that soil fungal community composition is impacted by long-term exposure to wood preservatives. Corals are among the most endangered species, and their cover and diversity around the world is declining fast. Chimetto Tonon et al. developed a qPCR assay for the monitoring of coral pathogens useful to determine their impact on coral reef ecosystems. In a different approach, Levin et al. focused on *Symbiodinium*, the coral photosymbiont, whose stress-induced loss causes coral bleaching. Using available sequencing data from *Symbiodinium* the authors developed a testable expression construct model that incorporates endogenous *Symbiodinium* promoters, terminators, and genes of interest to enhance the photosymbiont stress tolerance and thus, that of coral reefs.

Multi-omics approaches are also currently deployed to assist risk management in food safety and quality (Cocolin et al., 2017). Using a comparative transcriptomics analysis of *Monascus purpureus* (a yeast used as food colorant that also produces citrinin, a compound with nephrotoxic, hepatotoxic, and carcinogenic activities) and a mutant strain, Liang et al. were able to identify the mechanisms underlying pigment and citrinin biosynthesis. These findings will inform the construction of genetically engineered *Monascus purpureus* strains unable to produce citrinin and optimized for pigment synthesis. *Listeria monocytogenes* is an important food-borne pathogen that causes listeriosis, a dangerous disease with human life compromising consequences. Using transcriptome analysis and sequence alignment, Zhang et al. identified six genes related to D-allose metabolism only present in the genomes of lineage II strains. This finding will benefit isolation strategy and epidemiological research of *L. monocytogenes*.

Recognizing the utmost importance of microbiomes to human health, the Human Microbiome Project (Nelson and White, 2010) was launched in 2007 to provide unprecedented insights into our microbiota. While initially most of the information was derived from 16S rRNA amplicon sequencing and metagenomics, the second phase of the project, launched in 2014, called the Integrative Human Microbiome Project (iHMP), aims to create integrated longitudinal datasets from microbiome and host using multiple omics technologies (https://hmpdacc.org/ihmp/). In that context, Banerjee et al. investigated microbial diversity in four major types of breast cancer using whole genome and transcriptome amplification and a pan-pathogen microarray (PathoChip) strategy. The authors detected unique and common viral, bacterial, fungal, and parasitic signatures for each of the breast cancer types. This information will underpin better prognosis, treatment strategies and clinical outcomes. Liu et al. investigated the microbiome of sputum and oropharyngeal swabs in patients with chronic obstructive pulmonary disease (COPD) using 16S rRNA and ITS amplicon sequencing, and found that the two sample types generated rather similar taxonomic profiles. The finding from this work will contribute to the design of easier methodology for medical sampling in the context of COPD patients.

Omics methodologies have also been recently applied to the field of museum objects and archeological remains preservation. Using a combination of culture-independent and culture-dependent methods Liu et al. investigated the microbial communities responsible for the biodeterioration of antique museum objects and identified fungal and bacterial taxa responsible for the deterioration. The findings will inform the future planning of biocide treatment of museum antiques. In another example, Liu et al. determined the fungal community

structure of a wooden tomb from the Western Han Dynasty (206 B.C.−25 A.D.) in China. ITS1 gene amplicon sequencing identified a total of 114 genera distributed across five fungal phyla, with a dominant member, *Hypochnicium* sp. WY- DT1. This fungus was further demonstrated to possess the ability to degrade cellulose and lignin and therefore represents a serious threat to the preservation of wooden archeological remains.

Finally, a set of articles reports on new methodological developments, to deal with omics analysis related challenges. Miyazaki et al. proposed a method to tackle the lack of specificity of 16S rRNA for *Escherichia coli*, a species previously reported as able to harbor foreign 16S rRNA. To circumvent this problem, the authors designed a new primer set for 16S rRNA genes with no overlap with potential mismatch sites formerly detected. Zhong et al. proposed an *in silico* decontamination methodology for the investigation of microorganisms present in ice cores dated 20–30,000 years from the Tibetan Plateau. A series of controls were used to assess contaminant microbial diversity and abundances, which were removed *in silico* from the field samples data. As sequencing methods invariably lead to extensive datasets, more sophisticated and user-friendly methods for data analysis are required. Bhuvaneshwar et al. describe an open source bioinformatics pipeline (viGEN), which allows for the detection and quantification of viral RNA, and variants from viral transcripts. This pipeline can be used to provide novel biological insights into microbial infections and tumorigenesis. Parmar et al. review the use of genomics, transcriptomics, proteomics, and metabolomics methodologies as well as the associated bioinformatics tools to infer phylogenetic affiliation and function of bacteriophages and their impact on diverse microbial communities.

Finally, it is worth highlighting that meta-omics technologies are now commonly used in combination, as each omics provides a different and complimentary level of information on the ecosystem under investigation (Meiring et al., 2011). Indeed several studies from this special issue (e.g., Joyce et al.; González-Martínez et al.; Lindemann et al.) and in recent literature, combine multiple omics to better address microbiome structure and functioning in the context of environmental services, goods, and biotechnological applications. However, multi-omics datasets integration remains challenging and new modeling and statistical tools like multi-layer network theory and artificial intelligence methodologies are being developed to this end (Haas et al., 2017).

We believe that the articles published in this Research Topic provide an updated, high-quality overview of current work in the field. This body of research makes a valuable contribution to the understanding of microbial ecosystem services, and expands the horizon for finding and developing new and more efficient biotechnological applications.

AUTHOR CONTRIBUTIONS

DM drafted the manuscript, FA revised the draft and both authors agreed to the final version. DM proposed the Research Topic theme and the articles were edited by DM and FA.

ACKNOWLEDGMENTS

We would like to thank all the contributing authors for their interest in our Research Topic. DM is a research member of the Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Argentina. FA leads the Functional Environmental Microbiology research group and is a member of the Ryan Institute at the National University of Ireland Galway.

REFERENCES

Antwis, R. E., Griffiths, S. M., Harrison, X. A., Aranega-Bou, P., Arce, A., Bettridge, A. S., et al. (2017). Fifty important research questions in microbial ecology. *FEMS Microbiol. Ecol.* 93:5. doi: 10.1093/femsec/fix044

Bates, B. C., Kundzewicz, Z. W., Wu, S., and Palutikof, J. P. (eds.) (2008). *Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change* (Geneva: IPCC Secretariat), 210.

BCBA(Bolsa de Cereales de Buenos Aires) (2017). *Producción de Soja en Argentina*.

Bedmar, E. J., Bueno, E., Correa, D., Torres, G., et al. (2013). "Ecology of denitrification in soils and plant-associated bacteria," in *Beneficial Plant-microbial Interactions: Ecology and Applications*, eds M. González and J. Gonzalez-López (Boca Raton, FL: CRC Press), 164–182.

Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M. F., Jagadeesan, B., et al. (2017). Next generation microbiological risk assessment meta-omics: the next need for integration. *Intl. J. Food Microbiol.* 287, 10–17. doi: 10.1016/j.ijfoodmicro.2017.11.008

Craddock, V. M., and Henderson, A. R. (1986). Effect of N-nitrosamines carcinogenic for oesophagus on O6-alkyl-guanine-DNA-methyl transferase in rat oesophagus and liver. *J. Cancer Res. Clin. Oncol.* 111, 229–236.

Dangi, A. K., Sharma, B., Hill, R. T., and Shukla, P. (2018). Bioremediation through microbes: systems biology and metabolic engineering approach. *Crit. Rev. Biotechnol.* 39, 79–98. doi: 10.1080/07388551.2018.1500997

García-Fraile, P., Menéndez, E., and Rivas, R. (2015). Role of bacterial biofertilizers in agriculture and forestry. *AIMS Bioeng.* 2, 183–205. doi: 10.3934/bioeng.2015.3.183

Greer, F. R., and Shannon, M. (2005). Infant methemoglobinemia: the role of dietary nitrate in food and water. *Pediatrics* 116, 784–786. doi: 10.1542/peds.2005-1497

Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S., and Ralser, M. (2017). Designing and interpreting 'multi-omic'experiments that may change our understanding of biology. *Curr. Opin. Syst. Biol.* 6, 37–45. doi: 10.1016/j.coisb.2017.08.009

IPCC (2014). "Climate change 2014: synthesis report," in *Contribution of Working Groups I, II and III to the Fifth Assessment Report of Intergovernmental Panel on Climate Change*, eds R. K. Pachauri and L. A. Meyer (Geneva: Geneva Press), 151 p.

Meiring, T. L., Bauer, R., Scheepers, I., Ohlhoff, C., Tuffin, M. I., and Cowan, D. A. (2011). "Metagenomics and beyond: current approaches and integration with complementary technologies", in *Metagenomics: Current Innovations and Future Trends*, ed D. Marco (Norfolk: Caister Academic Press), 1–19.

Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well-being: Synthesis.* Washington, DC: Island Press.

Nelson, K. E., and White, B. A. (2010). "Metagenomics and its applications to the study of the human microbiome," in *Metagenomics: Theory, Methods and Applications*, ed D. Marco (Norfolk: Caister Academic Press), 171–182.

Okafor, N., and Okeke, B. C. (2017). *Modern Industrial Microbiology and Biotechnology*. Boca Raton, FL: CRC Press.

Shah, V., Jain, K., Desai, C., and Madamwar, D. (2011). "Metagenomics and integrative '-omics' technologies in microbial bioremediation: current trends and potential applications," in *Metagenomics: Current Innovations and Future Trends*, ed D. Marco (Norfolk: Caister Academic Press), 211–240.

Tortosa, G., Hidalgo, A., Salas, A., Bedmar, E. J., Mesa, S., and Delgado, M. J. (2015). Nitrate and flooding induce N_2O emissions from soybean nodules. *Symbiosis* 67, 125–133. doi: 10.1007/s13199-015-0341-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Marco and Abram. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 June 2018
doi: 10.3389/fmicb.2018.01388

Assessing the Effect of Pretreatments on the Structure and Functionality of Microbial Communities for the Bioconversion of Microalgae to Biogas

Olivia Córdova[1]*, *Rolando Chamy*[1], *Lorna Guerrero*[2] and *Aminael Sánchez-Rodríguez*[3]*

[1] Laboratorio de Biotecnología Ambiental, Escuela de Ingeniería Bioquímica, Facultad de Ingeniería, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, [2] Department of Chemical and Environmental Engineering, Universidad Técnica Federico Santa, Valparaíso, Chile, [3] Microbial Systems Ecology and Evolution, Department of Biological Sciences, Universidad Técnica Particular de Loja, Loja, Ecuador

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Seung Gu Shin, Pohang University of Science and Technology, South Korea
Jiangxin Wang, Shenzhen University, China

***Correspondence:**
Olivia Córdova
olivia.cordova.v@mail.pucv.cl
Aminael Sánchez-Rodríguez
asanchez2@utpl.edu.ec

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: 02 February 2018
Accepted: 06 June 2018
Published: 26 June 2018

Citation:
Córdova O, Chamy R, Guerrero L and Sánchez-Rodríguez A (2018) Assessing the Effect of Pretreatments on the Structure and Functionality of Microbial Communities for the Bioconversion of Microalgae to Biogas. Front. Microbiol. 9:1388. doi: 10.3389/fmicb.2018.01388

Microalgae biomethanization is driven by anaerobic sludge associated microorganisms and is generally limited by the incomplete hydrolysis of the microalgae cell wall, which results in a low availability of microalgal biomass for the methanogenic community. The application of enzymatic pretreatments, e.g., with hydrolytic enzymes, is among the strategies used to work around the incomplete hydrolysis of the microalgae cell wall. Despite the proven efficacy of these pretreatments in increasing biomethanization, the changes that a given pretreatment may cause to the anaerobic sludge associated microorganisms during biomethanization are still unknown. This study evaluated the changes in the expression of the metatranscriptome of anaerobic sludge associated microorganisms during *Chlorella sorokiniana* biomethanization without pretreatment (WP) (control) and pretreated with commercial cellulase in order to increase the solubilization of the microalgal organic matter. Pretreated microalgal biomass experienced significant increases in biogas the production. The metatranscriptomic analysis of control samples showed functionally active microalgae cells, a bacterial community dominated by γ- and δ-proteobacteria, and a methanogenic community dominated by *Methanospirillum hungatei*. In contrast, pretreated samples were characterized by the absence of active microalgae cells and a bacteria population dominated by species of the Clostridia class. These differences are also related to the differential activation of metabolic pathways e.g., those associated with the degradation of organic matter during its biomethanization.

Keywords: biogas, microalgae, *Chlorella*, methane, bioconversion, enzymatic pretreatment

INTRODUCTION

The advantages of using microalgae as substrate for biogas production (biomethanization) come from their biological and biochemical features, such as their ability to capture CO_2 and use it to sustain growth, their high productivity in relation to other biomasses and to the possibility of converting all fractions of microalgae organic matter into biofuels (Sialve et al., 2009; González-Fernández et al., 2012; Bohutskyi and Bouwer, 2013). However, biofuels production from

microalgae is not yet a system scalable to an industrially viable one (Zamalloa et al., 2011). This is mainly because microalgae cell wall is difficult to degrade by hydrolytic bacteria (such those commonly found in anaerobic sludge associated bacteria). Therefore, in the absence of available microalgae organic matter to feed the anaerobic digestion process, biogas production is deficient (González-Fernández et al., 2012).

One solution to this biotechnological problem is to apply a pretreatment to the microalgae cultures (Angelidaki and Batstone, 2010; Mendez et al., 2013; Passos et al., 2014) with the purpose of increasing the availability of soluble organic matter and thus improving biogas production yield (Bohutskyi and Bouwer, 2013). Different pretreatments may be applied: physical pretreatments (by applying a physical force and/or heat) or enzymatic pretreatments (by adding enzymatic raw extracts or commercial enzymes). Enzymatic pretreatments aim at increasing the selective permeability of the microalgae cell wall to release inner compounds as well as to solubilize cell wall constituents (González-Fernández et al., 2012; Mahdy et al., 2016). Enzymatic pretreatments change the way in which microalgae organic matter is made available in the medium and lead therefore to a new configuration of the microalgal biomass (respect to the organic matter configuration prior the pretreatment). The concept of biomass configuration will be used in this study to refer to the way in which organic matter from the microalgal biomass becomes available in the medium.

Bareither et al. (2013) characterized the microbial diversity (bacteria and archaea) during the biodegradation of urban solid waste in two conditions (solid waste and leachate of solid waste) and correlated it to methane production. Authors concluded that microbial communities were not similar between conditions. Their results support the hypothesis that the identity of functionally active species in anaerobic sludge associated microbial communities changes, not only for each substrate type, but also for the same substrate under different conditions. There are also studies that report changes in the structure, e.g., changes in the species diversity of microbial communities when using different substrates for biogas production under the same operating conditions in the laboratory (Lee et al., 2009; Kampmann et al., 2012). Aforementioned studies assessed changes in microbial communities during biogas production at a rather low resolution: it means without providing information on specific activated/repressed pathways across conditions (substrates and/or pretreatments). One could hypothesize that changes in the configuration of the microalgal biomass drive structural i.e., species being present, and also functional changes i.e. pathways being activated/repressed in the anaerobic sludge associated microbial communities during biomethanization. However, such hypothesis is poorly addressed in the literature.

Addressing changes in anaerobic sludge associated microbial communities during biogas production at a better resolution is now possible thanks to current developments of the so called "omic" technologies. The development of what has been termed "omic" techniques, particularly those that apply to genetic material isolated directly from environmental samples, i.e., metagenomic, metatranscriptomic, allows for the evaluation of structural (changes in the relative abundance of species), and functional dynamics (changes in genetic expression) of microbial communities (Jansson et al., 2012). Meta-omics studies have even been possible in the context of bioreactors, generating knowledge on how the configuration of a reactor and its operating conditions influence the microbial community (Zhang et al., 2010; Vanwonterghem et al., 2014). Meta-transcriptomic studies allow the identification of near full length transcripts being expressed by a given microbial community under a set of experimental conditions (Moset et al., 2015; Nolla-Ardèvol et al., 2015; Stolze et al., 2015). Such studies can be used to identify differentially expressed genes across conditions.

In the present study, we describe the impact that the enzymatic pretreatment of microalgal biomass has on the anaerobic sludge associated microbial communities during biogas production. Pretreatment impact is analyzed at the level of species composition as well as on what respect to the activation/repression of metabolic pathways. We did so by reconstructing the metatranscriptome of anaerobic sludge associated microbial communities during the biomethanization of *Chlorella sorokiniana,* with and without enzymatic pretreatment to increase the solubility of organic matter and to achieve significant increases in biogas production.

MATERIALS AND METHODS

Microalgae Culture

A culture of *C. sorokiniana* (Shihira and Krauss, 1965) isolated from an effluent of anaerobic sludge digestors belonging to a waste water treatment plant in Spain, was donated by the University of Huelva, Spain. The Sueoka culture medium (Sueoka, 1960) was used to maintain this culture in the laboratory. The culture was grown on 5 L flasks under non-sterile conditions at a temperature of $21 \pm 2^{\circ}$C, with artificial lighting of F24-39 W and I = 127.60 μmol of photons/($m^2 \times s$), 24-h light photoperiod and aeration of 1.3–1.5 L/min of atmospheric air.

Chlorella sorokiniana biomass composition was characterized in what respect to total protein content by the Kjeldahl method which measures total organic nitrogen (Owusu-Apenten, 2002; Safi et al., 2013). Total lipids where determined by Soxhlet method (APHA-AWWA-WPCF, 1999), and carbohydrate by the Dubois method (Dubois et al., 1951). Recalcitrant material, measured as the insoluble fiber content of the sample, was determined using acid digestion followed by alkaline digestion (APHA-AWWA-WPCF, 1999).

Enzymatic Pretreatment

For enzymatic pretreatment application 400 mL of microalgal biomass was used at an enzyme/substrate ratio of 1%, pH 7 for 24 h at 37°C. The Ns22128 enzyme (cellulase) from Novozymes® was used for this purpose.

Cell Wall Rupture

Microalgae cell wall rupture was evaluated by SYTOX Green staining in pretreated cells (Sato et al., 2004). This probe has a high affinity for nucleic acids and, only penetrate cells whose cell membranes are damaged. In this way, probe fluorescence and the

microalgae autofluorescence were used to mark dead cells (due to rupture or damage) and live cells respectively.

Biochemical Methane Potential (BMP)

Methane production from *C. sorokiniana* cultures was evaluated using a biochemical methane potential test (BMP) (Angelidaki et al., 2009). The inoculum used came from an anaerobic sludge reactor fed with sludge from a waste water treatment plant located at "La Farfana", Santiago, Chile. Bottles of 100 mL capacity were used for the BMP test. All flasks were inoculated at 0.5 g. of volatile solids (VS) substrate/g. of VS inoculum ratio. Bottles containing only the inoculum were used as controls in order to correct for inoculum methane yields. We assessed the methane production from the inoculum, determined in blank assays with medium, and no microalgal biomass, which is subtracted from the methane production obtained with microalgal biomass assays. Enzyme control, biomass control, and inoculum control were performed for each BMP assay.

Bubbles were made in the bottles using a mix of gases (80% N and 20% CO_2) in order to ensure anaerobic conditions, and were then sealed and kept at 37°C. The test ended once the methane production had stopped.

The percentage of CH_4 in the biogas was determined by gas chromatography using a Perkin Elmer Clarus 500 chromatograph, oven temperature 80°C, detector TCD at 120°C, and injector 80°C. Helium was used as carrier with a Hayesep Column Q 4 m × 1/8″ OD (13 ft.). One milliliter of biogas was taken with a glass syringe and then injected into the port of the Gas Chromatograph. Determinations were performed by triplicates to estimate the average value of CH_4 percentage present in the biogas.

CH_4 production was quantified by displacement of a NaOH solution due to carbon dioxide absorption. The accumulated CH_4 production in time (accumulated CH_4 mL/g. VS of substrate) was normalized to mL/g. VS of substrate using Equation (1).

$$\frac{mL\ of\ CH_4}{g.\ VS\ of\ substrate} = \frac{mL\ of\ produced\ CH_4}{\frac{g.\ VS\ substrte}{L} \times\ mL\ of\ substrate\ in\ a\ bottle} \quad (1)$$

Methane Productivity Modeling

Methane productivity was modeled using the modified Gompertz model (Donoso-Bravo et al., 2010) based on the values observed during the BMP test according to Equation (2):

$$B = P \times \exp(-\exp\left(\frac{Rm \times e}{P}(\lambda - t) + 1\right)) \quad (2)$$

where B represents the accumulated volume of CH_4 produced at time t (in days), P the maximum CH_4 production potential (mL CH_4/g. VS of substrate), Rm the maximum production rate (mL CH_4/g. VS of substrate/day), λ the duration of the latency stage (in h), and t the incubation time (in days).

Analytical Methodology

All analyses were performed by triplicates and average values and their standard deviations estimated for every biological replicate during the BMP assay. Both microalgal biomass and inoculum were characterized according to standard methods (APHA-AWWA-WPCF, 1999) to quantify physical-chemical parameters such as: total solids (TS), VS, and volatile suspended solids (VSS). The pH was measured with a HI/111 Hanna Instrument pH meter with a sensitivity of ±1 mV, which corresponds to 0.01 units of pH.

Statistical Analysis

Statistical analyses were performed to determine if there were significant differences between the conditions with and without enzymatic pretreatment in relation to their BMP. The analyses were performed using the software Statistica 13 (StatSoft Inc., Tulsa, USA, 2016). Each time a variance analysis was used to check whether the normality and homoscedasticity assumptions had been met. When the assumptions had not been met, the data was properly transformed.

RNA Extraction and Quantification for Sequencing Libraries

Samples from both experimental conditions were taken directly from the BMP test bottles. Samples were first centrifuged at 10,000 rpm for 10 min to discard the supernatant and the pellet was stored at −80°C until the moment of analysis. Total RNA extraction was performed using the PowerSoil RNA Isolation® (MOBIO) extraction kit following the manufacturer's instructions. RNA samples were then treated with the DNase Max Kit (MOBIO) to remove possible genomic DNA contamination. The samples were freeze dried for 6 h and send out for library preparation and Illumina sequencing at Molecular Research LP MRDNA Laboratory (Shallowater, Texas, USA). Total RNA concentration was determined using the Qubit® RNA Assay Kit (Life Technologies). RNA integrity value (RIN) was determined with the Agilent RNA 6000 Nano Reagents and RNA Nano Chips in Agilent 2100 Bioanalyzer (Agilent Technologies). Between 0.5 and 1.5 μg of the total RNA were used to remove the DNA using the Baseline-ZERO™ DNase (Epicentre) kit according to the manufacturer's instructions. In order to only retain the total mRNA fraction, the rRNA was eliminated from the RNA samples which were already free of DNA, using the Ribo-Zero™ Magnetic Gold Kit (Bacteria; Illumina). These samples were used library preparation with the TruSeq™ RNA LT Sample Preparation Kit (Illumina) according to the manufacturer's instructions. The final concentration of all libraries was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies) and their average size was determined with an Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries were pooled in an equimolar proportion of 2 nM, and 8 pM of the pool was pair-end sequenced for 300 cycles using a HiSeq 2500 system (Illumina).

Metatranscriptome Sequencing and Gene Expression Differential Analysis

Sequence quality control was performed with the FastQC toolkit (Andrews, 2010). Reads were subject to "reads trimming" before analysis, in order to eliminate stretches of low quality. Subsequently, a *de novo* metatranscriptome assembly was

performed using the Trinity bioinformatics tool (Grabherr et al., 2011). Transcripts quantification was done using the RSEM tool, which implements an EM algorithm (Expectation Maximization) to maximize the verisimilitude that a fragment comes from a given transcript and then calculates digital values of gene expression (Li and Dewey, 2011). The detection of differentially expressed genes between the samples without pretreatment (WP) and with pretreatment (EP) was done using DESeq (Anders and Huber, 2010).

Taxonomic Analysis and Functional Annotation of Microbial Communities Under Conditions With/Without Enzyme Pretreatment

Taxonomic annotations at the species level of each reconstructed transcript was done by BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We also downloaded the sequence of the rRNA *16S* gene of the species for which at least one differentially expressed transcript was detected. These species were considered as "active species" in each of the experimental conditions.

Prediction of Activated Key Proteins/Enzymes and of Their Corresponding Metabolic Pathways During Anaerobic Digestion

To identify metabolic pathways that were active at both experimental conditions (with and WP) during anaerobic digestion, the Kyoto Encyclopedia of Genes and Genomes-KEGG (http://www.genome.jp/kegg/) platform was used. We first mapped differentially expressed transcript to their corresponding genes in KEGG using the BlastKOALA tool. Genes where then mapped to metabolic pathways within KEGG. Finally, the XPathway was used based to compare metabolic pathways across experimental conditions. The XPathway is able to detect and quantify the metabolic differentiation between the two conditions (Temate-tiagueu et al., 2016).

RESULTS

Microalgal Culture Characterization

The characterization of the *C. sorokiniana* biomass showed the following relative composition: 45.5% proteins, 26.2% lipids, 23.7% carbohydrates, and 4.70% of raw fiber (insoluble material) at pH 6.9. The concentration of TS (g/L) was 6.90 ± 0.10, VS (g/L) of 6.50 ± 0.10, and a Chemical Oxygen Demand (g/L) of 13.8 ± 1.30.

Biochemical Methane Potential (BMP) and Productivity

The cellulase pretreated microalgal biomass showed a methane production of 537 mL of accumulated CH_4/g. VS of substrate, thus achieving a 75% increase in relation to the non-pretreated biomass with 307 mL of accumulated CH_4/g. VS of substrate (**Figure 1**). Additionally, differences were observed between experimental conditions in the behavior of the accumulated methane curves i.e., in the latency stage, the methane production rate slope and in the biochemical methane potential. According to the values inferred by the modified Gompertz model, the maximum methane production rate (*Rm*) in the pretreated biomass was on average 2.65 times that one observed in the non-pretreated biomass (**Table 1**).

The average methane percentage found in the produced biogas during the BMP test was of 66.7 ± 1.5% which falls within known methane percentages in biogas (60–70%).

Molecular Analysis of Anaerobic Sludge Microbial Communities

After quality control of the raw data (see section Materials and Methods for details), a total of 29,618,400 reads were retained. From those retained reads we were able to reconstruct a total of 96,193 transcripts from all samples. An average of 32,552 transcripts were reconstructed from control samples (without enzymatic pretreatment) while 38,921 was the average of transcripts reconstructed for samples with enzymatic pretreatment. A subset of 15,088 transcripts were only expressed in samples without enzymatic

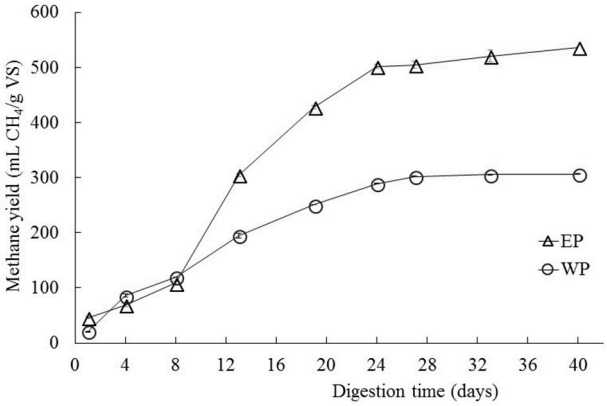


FIGURE 1 | Biochemical methane potential of *C. sorokiniana* when subjected to enzymatic pretreatment (cellulase) under conditions of 1% enzyme-substrate concentration, pH 7, for 24 h (EP data series). The biochemical methane potential is also shown for control samples (WP data series) that were not treated enzymatically. All test was performed by triplicates. Average values ± standard deviations are plotted in each case.

TABLE 1 | Methane productivity inferred from observed values by fitting a Gompertz Model.

Pretreatment	Condition	λ (h)	R_m (mL CH_4/g. VS/d)	P (mL CH_4/g. VS)	R^2
Enzymatic with cellulase	1% e/s-pH 7–24 h	3.31 ± 0.24	30.67 ± 3.27	545.68 ± 42.70	0.991
Biomass without pretreatment	–	1.60 ± 0.24	11.56 ± 0.15	317.66 ± 0.90	0.965

λ, latency period; Rm, maximum production rate; P, maximum CH_4 production potential. R^2 value for adjustment between observed and modeled values. e/s, enzyme-substrate ratio. All test performed by triplicates. Average values ± standard deviations are provided.

pretreatment while 11,686 transcripts showed an expression only in the samples with enzymatic pretreatment. From this information, 227 differentially expressed genes across conditions were detected. All sequencing data generated during this study can be accessed at the NCBI Trace Archive (study ID SRP139287).

Taxonomic and Functional Annotations of Differentially Expressed Genes Across Experimental Conditions

Taxonomic and functional annotation was only performed for the 227 differentially expressed genes across experimental conditions (from here on we refer to control samples as WP and to samples with enzymatic pretreatment as EP ones). Differentially expressed transcripts can be seen as the functionally active fraction that better represent the species with the highest responsiveness to the experimental conditions being tested (with and without enzymatic pretreatment). The unanalyzed transcripts (non-differentially expressed) have a high probability of being part of constitutive metabolic pathways that are unrelated to the cellular responses triggered/repressed by the experimental conditions. Based on the above, it should be noted that the analysis of the relative abundance of differentially expressed transcripts allowed us to identify the structure of the functionally active microbial community by providing information on the existing taxa, as well as differentially active metabolic pathways between the conditions.

Functionally Active Taxa in Control Samples

A total of 47 differentially expressed transcripts were identified for the condition WP (**Figure 2**), 39 of which were taxonomically annotated at the species level. For the remaining eight transcripts annotation at the species level was not possible.

From the taxonomic annotation of the 39 expressed transcripts in the WP condition, a predominant first cluster (C-I) that represents the 35% of the microbial community becomes evident. C-I is made up of sulfate-reducing bacteria. A second cluster (C-II) of extreme environment thermophilic bacteria was found, composed of *Thermosipho africanus*, *Defluviitoga tunisiensis* (which are closely related), and of *Wenzhouxiangella marina*. Additionally, three archaea species were detected in C-II: *Methanosaeta concilii* and *Methanospirillum hungatei* JF-1, which presented the highest quantity of annotated transcripts, belonging to the Methanomicrobia class and finally *Candidatus methanoplasma* of the Thermoplasmata class.

As expected, we got evidence of active microalgal cells which were represented by four species: *Chlorella* sp., *C. sorokiniana*, *Chlorella vulgaris*, and *Micractinium reisseri*, all belonging to the Chlorellaceae family.

Functionally Active Taxa in Samples Subjected to an Enzymatic Pretreatment

A total of 55 differentially expressed transcripts were identified in the EP condition, 50 of which were taxonomically annotated at the species level (**Figure 3**). Annotation at the species level was not possible for the remaining five transcripts.

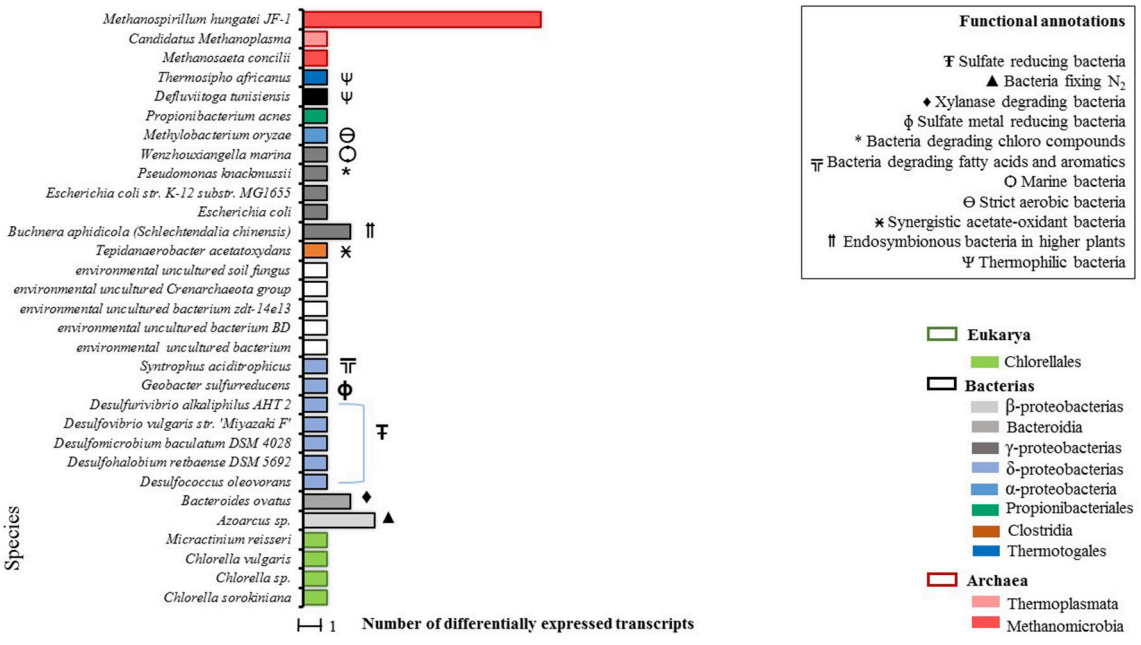


FIGURE 2 | Total count of differentially expressed transcripts by species in WP samples (microalgal biomass without enzymatic pretreatment). Phylogenetic annotations are indicated at the level of classes (color-coded legend) as well as species functional annotations for transcripts with a taxonomic assignment (BLAST based; inset box). For blank horizontal bars, identification at the species level was not possible.

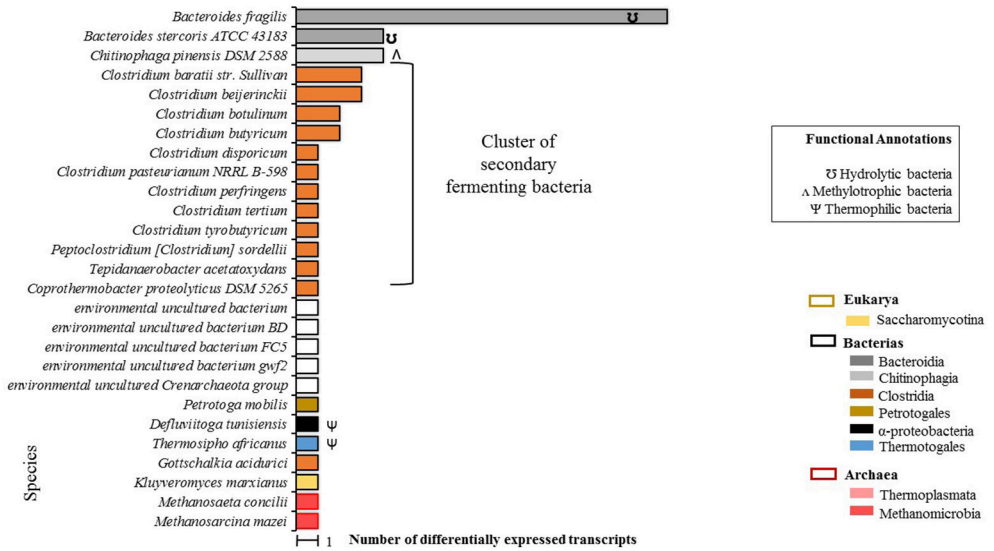


FIGURE 3 | Total count of differentially expressed transcripts by species in EP samples (microalgal biomass with enzymatic pretreatment). Phylogenetic annotations are indicated at the level of classes (color-coded legend) as well as species functional annotations for transcripts with a taxonomic assignment (BLAST based; inset box). For blank horizontal bars, identification at the species level was not possible. Species belonging to secondary fermenting bacteria have been clustered together.

A predominant bacterial cluster (C-I) was found which represented 50% of the microbial community. C-I was made up of secondary Clostridia-class fermentation bacteria. Additionally, a second cluster (C-II) of extreme environment thermophilic bacteria was foun composed of *T. africanus*, *Petrotoga molbilis,* and *D. tunisiensis* species, which represented the 14% of the microbial community. Two archaea species were detected: *M. concilii* and *Methanosarcina mazei*, both of the Methanomicrobia class. It should be noted that under this condition, no transcripts were identified for microalgae species, which could confirm that no living microalgal cells were actually present after the enzymatic pretreatment.

Metabolic Pathways Activated Across Experimental Conditions

Control Samples

The number of active metabolic pathways identified for both conditions was low. However, we were able to map key enzymes of these metabolic pathways which provided strong evidence of the actual processes that were triggered in the microbial community in response to the experimental conditions being assayed. **Table 2** shows the results of the prediction of key proteins/enzymes for the activation of metabolic pathways in the WP condition.

Metabolic pathway activation was detected for processes involved in the processing of environmental information, carbohydrate and lipid metabolism, and in the activation of bacterial defense mechanisms. This shows some of the functional behavior of the microbial community which biomethanized a microalgal biomass with no damage or rupture to the cell wall. There was also evidence of bacterial *quorum sensing* (cellular process, ko02014). *Quorum sensing* detection could be indicative of genetic expression regulation in response to fluctuations in cell population density. Bacteria produce and release chemical signals (autoinducers) which become more concentrated as cell density increases (Waters and Bassler, 2005). *Quorum sensing* activation was detected in *T. africanus* (Bacteria, Thermotogae) and in *D. tunisiensis* (Bacteria, Petrotogae) both of which are phylogenetically related. The activation of *quorum sensing* was also detected in a third bacterium, although this could not be identified at the species level.

Bacterial organic material degradation was also detected. The activation of carbohydrate metabolism was recorded through galactose degradation (carbohydrate metabolism, ko00052) for *Escherichia coli* (Bacteria, γ-proteobacteria). In addition, lipid metabolism activation through fatty acids degradation by *Defluviitoga oleovarans* (Bacteria, δ-proteobacteria) and amino acids degradation (Metabolism of other amino acids ko 00430, ko00250) through the hydrolytic action of *Propionibacteria acnes* (bacteria, Propionibacteria) were also detected. Bacterial defense mechanisms activation was shown for *Geobacter sulfurreducens* (Bacteria, δ-proteobacteria) for which the synthesis of β-lactam of resistance was found (resistance to drugs, via β-lactam ko1501). For the archaea *M. hungatei* JF-1, methane metabolism activation was detected (ko, 00680) only for the pathway $CO_2 \rightarrow CH_4$. Additionally, cell mobility activation for archae was apparently achieved through the flagellar protein pathway FlaB (ko02040). No prediction of proteins/enzymes was achieved that could indicate metabolic pathways activation in the microalgal species identified in the study.

Samples Subjected to an Enzymatic Pretreatment

Some similarities were observed in EP samples with respect to WP samples mainly in relation to *quorum sensing* activation.

TABLE 2 | Key proteins/enzymes prediction for metabolic pathways activation by enzyme mapping through the KEGG online platform of differentially expressed transcripts detected on WP samples.

Protein/Enzyme	KEGG id	Organism from anaerobic sludge	E-value	KEGG pathway (id)
Peptide/nickel transport system substrate-binding protein	K02035	*T. africanus*	$6e^{-50}$	Quorum sensing (ko02014)
Peptidoglycan-associated lipoprotein	K03640	*Geobacter*	$1e^{-35}$	Transporters (ko02000)
F-type H+-transporting ATPase subunit alpha	K02111	*S. aciditrophicus*	$7e^{-152}$	Oxidative phosphorylation (ko00190)
Membrane fusion protein, multidrug efflux system	K03585	*G. sulfurreducens*	$2e^{-21}$	beta-Lactam resistance (ko01501)
7,8-Dihydropterin-6-yl-methyl-4-(beta-D-ribofuranosyl)aminobenzene 5′-phosphate synthase	K06897	*D. vulgaris*	$7e^{-81}$	Folate biosynthesis (ko00790)
DNA-binding protein HU-beta	K03530	Non-annotated	$1e^{-22}$	DNA repair and recombination proteins (ko03400)
Cold shock protein (beta-ribbon, CspA family)	K03704	*G. sulfurreducens*	$5e^{-28}$	Transcription factors (ko03000)
Large subunit ribosomal protein L4	K02926	*S. aciditrophicus*	$6e^{-100}$	Ribosome (ko03010)
Archaeal flagellin FlaB	K07325	*M. hungatei*	$2e^{-82}$	Secretion system (ko02040)
Beta-galactosidase	K01190	*E. coli*	$8e^{-33}$	Glycan degradation (ko00511)
Branched-chain amino acid transport system substrate-binding protein	K01999	Non-annotated	$4e^{-52}$	Quorum sensing (ko02014)
Prolyl-tRNA synthetase	K01881	Non-annotated	$1e^{-19}$	Aminoacyl-tRNA biosynthesis (ko00970)
Butyryl-CoA dehydrogenase	K00248	*D. oleovorans*	$4e^{-85}$	Fatty scid degradation (ko00071)
5,10-Methylenetetrahydromethanopterin reductase	K00320	*M. hungatei*	0.0	Methanogenesis, CO_2 => methane (ko00680)
Alanine dehydrogenase	K00259	*P. acnes*	0.0	Alanine, aspartate and glutamate metabolism (ko00250)

However, differences with WP samples were observed for metabolic pathways involved in carbohydrates and lipids metabolism, as well as for the activated bacterial defense systems (**Table 3**).

Regarding bacterial *quorum sensing* activation (cell process, ko02014), it was identified for *T. africanus* and *P. mobilis* (Bacteria, Petrotogae). It should be noted that for the WP condition, *quorum sensing* activation was also recorded for *T. africanus* and for a Petrotogae bacteria (*D. tunisiensis*). Regarding organic matter degradation, we collected evidence supporting that carbohydrate metabolism activation was through the degradation pathway of starch and sucrose (carbohydrate metabolism, ko00050) for *D. tunisiensis* (Bacteria, Petrotogae). For the same bacteria, co-factor metabolism activation was through the Pantothenate pathway and CoA biosynthesis (ko00770). Through this pathway cofactors play a key role in the biosynthesis and decomposition of fatty acids as well as in the biosynthesis of polyketides (secondary metabolites) and non-ribosomal peptides (Begley et al., 2001). Amino sugars degradation pathway and nucleotide sugar metabolism (ko00520) was also identified in *P. mobilis* (Bacteria, Petrotogae).

Like in the WP condition, the activation of bacteria defense mechanisms was recorded in the EP condition. In this case the activation of the defense mechanism was via specific restriction enzymes and hydrolytic enzymes (ko02048) identified in *D. tunisiensis*. For the same species, there was evidence of environmental information processing, via the ABC membrane transport pathway, which connects the ATP hydrolysis to the active transport of a wide variety of substrates such as ions, sugars, lipids, sterols, peptides, proteins (ko 02010). No prediction of proteins/enzymes was achieved that could indicate metabolic pathways activation in the archaea species detected under this condition.

DISCUSSION

Functional changes observed in microbial communities under both experimental conditions provide insights on how the community of both bacteria and archaea "restructures" based on whether or not an enzymatic pretreatment is performed. Such changes could mediate the community response to a new configuration of microalgal biomass. *De novo* metatranscriptome analysis of WP and EP turned out to be useful in providing a high resolution "picture" of the microbial community genetic expression without the need of an *a priori* knowledge of the genomes present (Moran, 2009; Vanwonterghem et al., 2014).

We collected phenotypic (methane production) a molecular (genetic expression) information suggesting that the anaerobic sludge associated microbial community managed to adapt to new configuration of organic matter that arose from the enzymatic pretreatment of the microalgal biomass. Changes in the configuration of organic matter when applying pretreatments to microalgal biomasses have been reported by Jiang et al. (2011). These authors applied pretreatments to the microalgal biomass with ultrasound and with the MFC (Microbial Fuel Cell) technique. After the pretreatments they analyzed the components being degraded by the bacterial fraction associated to an anaerobic sludge. They concluded that the components degraded differed among the pretreatments. For the pretreatment with ultrasound, mainly aromatic proteins were solubilized, while

TABLE 3 | Key proteins/enzymes prediction for metabolic pathways activation by enzyme mapping through the KEGG online platform of differentially expressed transcripts detected on EP samples.

Protein/Enzyme	KEGG id	Organism from anaerobic sludge	E-value	KEGG pathway (id)
Small acid-soluble spore protein D	K06421	*C. butyricum*	$1e^{-28}$	Unknown
Thiol peroxidase, atypical 2-Cys peroxiredoxin	K11065	*C. beijerinckii*	$5e^{-30}$	Oxidoreductases (ko01000)
Glycerol-3-phosphate acyltransferase PlsX	K03621	*D. tunisiensis*	0.0	Glycerolipid metabolism (ko00561)
ATP-binding protein involved in chromosome partitioning	K03593	*D. tunisiensis*	0.0	Mitochondrial biogenesis (ko03029)
Small acid-soluble spore protein D	K06421	*C. perfringens*	$1e^{-30}$	Unknown
Putative ABC transport system ATP-binding protein	K05833	*D. tunisiensis*	$3e^{-160}$	Transporters (ko02000)
Beta-glucosidase	K05349	*D. tunisiensis*	0.0	Starch and sucrose metabolism (ko00500)
Putative transposase	K07497	*W. marina*	$6e^{-43}$	Unknown
UDP-N-acetylglucosamine 2-epimerase	K01791	*P. mobilis*	$7e^{-71}$	Amino sugar and nucleotide sugar metabolism (ko00520)
Alpha-amylase	K01176	*D. tunisiensis*	$3e^{-127}$	Starch and sucrose metabolism (ko00500)
Putative ABC transport system substrate-binding protein	K01989	*D. tunisiensis*	0.0	Transporters (ko02000)
Phosphopantothenoylcysteine decarboxylase/phosphopantothenate cysteine ligase	K13038	*D. tunisiensis*	$3e^{-38}$	Pantothenate and CoA biosynthesis (ko00770)
Tyrosine-protein kinase Etk/Wzc	K16692	*D. tunisiensis*	0.0	Protein kinases (ko01001)
5-Methylcytosine-specific restriction enzyme B	K07452	*D. tunisiensis*	0.0	Prokaryotic defense system (ko02048)
HSP20 family protein	K13993	*C. pasteurianum*	$2e^{-32}$	Chaperones and folding catalysis (ko03110)
Branched-chain amino acid transport system permease protein	K01998	*P. mobilis*	$1e^{-139}$	Quorum sensing (ko02024)
Flagellin	K02406	*C. pasteurianum*	$2e^{-104}$	Flafellar assembly (ko02040)
Peptide/nickel transport system substrate-binding protein	K02035	*T. africanus*	$6e^{-50}$	Quorum sensing (ko02024)

for the pretreatment with the MCF technique carbohydrates were the main degraded substrates (Jiang et al., 2011). This situation can be explained by the fact that extracellular polymeric substances (EPS) is one of the main constituents of the total microalgal organic matter (Mishra and Jha, 2009). EPS are metabolic products that can be released to the extracellular medium and/or may be accumulated on the cell surface, providing protection to cells against a hostile environment. Therefore, once the pretreatment was applied, the microalgal cells were able to release different components depending on the type of pretreatment, thus changing the bioavailability of organic matter for bacteria.

A large number of reads could not be assigned to known sequences, which made it hard to obtain a complete transcript analysis. This is due to the large quantity of bacteria and archaea found in association to anaerobic sludge that have not yet being sequenced and/or annotated. Therefore, the conclusions we provide in the following paragraphs must still be taken with caution since they arose from a still incomplete meta-transcriptome. Despite this, the present study is a pioneering effort to shed light on the main changes that might occur at the metatranscriptomics level in the context of biomethanization and biomass pretreatment, when little a priori genomic information of the target microbial community is available.

Changes observed in the metatranscriptome of the studied microbial community could have been caused by multiple factors, for example, a change in bacteria energy source or in microalgal defense mechanisms activation, which brought about new interactions among microorganisms (Bochner, 2009). The following sections discuss some of these factors.

Energy Sources

In WP samples microalgal biomass comprised living cells which suggests little to no damage of the cell wall. In contrast, in EP samples microalgal cells were mostly dead in their vast majority due to significant cell wall damage. This could have determined the type of organic matter bacteria were assimilating in both samples. Analysis of active metabolic pathways suggests that in WP samples bacteria energy sources came mainly from sugar-based organic matter degradation, such as galactose, fatty acids, and some amino acids. Under this condition, hydrolytic bacteria populations must degrade a rigid cell wall. Activated enzymes found in WP samples provided evidence for fatty acids (Acyl-CoA dehydrogenase), carbohydrates (Beta galactosidase), and amino acids (Alanine dehydrogenase) degradation.

The activity of the beta-galactosidase enzyme (k01190) is fundamental for galactose metabolism which is commonly activated in conjunction to amino sugars and nucleotide sugars metabolism. Such activations are common in hydrolytic bacteria metabolizing for instance glucosamine, which has been reported as the main component of the rigid cell wall of the microalgae species *C. vulgaris* and *C. sorokiniana* (Takeda, 1991; Templeton et al., 2012).

In EP samples, bacteria energy sources seemed to come from sugary organic matter degradation such as sucrose and starch. We got evidence for the activation in *D. tunisiensis* of the cytoplasmic α-amilase, an enzyme of the starch hydrolase

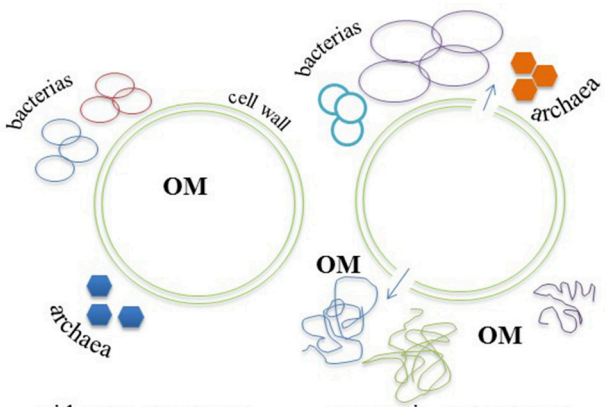
family that is key for the degradation of this compound (Janecek, 1997). Starch has been reported as an intracellular element, which shows that in EP samples energy came from intracellular organic components freed to the medium as consequence of cell wall breakdown.

Dominant Taxa, Key Functional Roles in Microbial Structure

The microbial community in WP samples appeared to be dominated by an active fraction of γ- and δ-proteobacteria. The δ-proteobacteria group is full of sulfate reducing bacteria (SRB). SRB are anaerobic microorganisms that use sulfate as acceptor of terminal electrons from the degradation of organic compounds with the concomitant production of H_2S (Muyzer and Stams, 2008). SRB digest fermentation products such as acetate, butyrate, lactate and hydrogen (Gerardi, 2003). In EP samples, the microbial community was widely dominated by a large variety of Clostridium species (13 species). Bacteria from the Clostridium genus are characterized by their intensive fermentative metabolism. They can use numerous organic compounds as carbon and nitrogen sources. Based on the presence of this fermenting bacteria cluster of *Clostridium* genus, we can infer that the new organic matter configuration that resulted from biomass enzymatic pretreatment, provided favorable conditions for the growth of Clostridium species. It has been reported that Clostridium species can exhibit an opportunistic behavior i.e., high adaptability, when an increase in soluble organic matter in the medium is verified (Lee et al., 2008; Szymanowska-Powałowska et al., 2014).

Ecological Interactions

In both samples (WP and EP), transcripts were found that codify for enzymes associated with bacteria quorum sensing (QS 02024). Quorum sensing is a regulator system that allows bacteria to share information about cell density and adjust their genetic expression in relation to their interaction with the environment (Williams, 2017). Solely based on transcript analysis, it is not possible to determine the causes of QS activation, that is, if it was due to bacteria-bacteria or microalga-bacteria defense mechanisms. However, we can mention that some of the processes controlled by QS include virulence, competition, conjugation, antibiotic production, motility, sporulation, and biofilm formation as bacteria defense mechanisms (Waters and Bassler, 2005). In the case of bacteria-bacteria interactions, the defense mechanisms imply a cell-to-cell communication that leads to the expression and release of bioactive substances to the surroundings and that influence the behavior of other microorganisms found in the environment (Waters and Bassler, 2005). In the case of bacteria-microalga mechanisms, QS allows bacteria to detect microalgal cells. The detection signal is precise



Level	without pretreatment (live cells)	enzymatic pretreatment (dead cells)
► Energy sources	Galactose, fatty acids, amino acids	Sucrose, starch, amino acids
► Dominant taxas	δ-γ-proteobacterias	Clostridia
► Ecological interaction	Mechanism of bacterial defense, bacterial QS	Bacterial QS/antagonist effect
► Methanogenesis (identified pathway)	CO_2→ CH_4	$C_2H_3O_2^-$→ CH_4
► Biogas production (mL CH_4 accumulated / g VS)	307±17	537±21 (↑31.6%)

FIGURE 4 | Main differences observed between the biomethanization process of a microalgal biomass with and without enzymatic pretreatment. Differences were categorized into five levels. Cartoonish representations of biomethanization process with emphasis on the expected changes of OM configurations between both experimental conditions are included. OM, organic matter.

and regulated according to the microalga, its growth stage and biomass density (Mitsutani et al., 2001). QS acts as a bacteria inducer to produce and secrete "algicide" substances in the surrounding medium (Waters and Bassler, 2005). Microalgae are able to secrete compounds that imitate the QS detection signals of many Gram-negative bacteria, resulting in stimulant, or inhibitor effects. For example, Shehata et al. (2013) described the effect of *Chlorella vulgaris* on the growth of different Clostridium species. This might be the reason why in WP samples where Chlorella cells were still alive, no proliferation of Clostridium species was detected. The opposite was observed in EP samples where Chlorella living cells were wiped out.

Methanogenesis

The final biochemical phase of the anaerobic digestion process is methanogenesis. During methanogenesis methane is metabolized by methanogenetic archaea in the global carbon cycle (Lee et al., 2009). In this study, differences were observed in the abundance and diversity of methanogenetic archaea across experimental conditions. In WP samples three archaea species were detected: *C. methanoplasma*, *M. concilii* (with a very low number of transcripts), and *M. hungatei* JF-1 (with a high number of transcripts). We also recorded the activation of the methanogenic pathway that uses CO_2 as substrate in the WP samples.

For *M. hungatei*, syntrophic relations with other microorganisms have been described (Walker et al., 2012). A syntrophic relationship is a specific form of microbial mutualism that occurs between acetogenic bacteria or BSRs and archaea, which can use organic or inorganic substances as substrates for fermentation. In these associations, BSRs as *Desulfovibrio* species, act as secondary fermenters that are obligatorily bound by interspecific electrons to the metabolic activity of methanogenetic archaea (Kato and Watanabe, 2010). A syntrophic relationship between *M. hungatei* and BSRs would explain the large quantity of transcripts identified for this archaea in the WP samples.

Two pathways have been described for methane formation from acetate as substrate. The first is the acetoclastic pathway, carried out by Methanosarcinaceae or Methanosaetaceae. The second pathway involves a two-stage reaction in which acetate is first oxidized to H_2 and CO_2 which are then converted to methane. This reaction is performed by acetate oxidizing bacteria, such as the Clostridium species found in the EP samples in a syntrophic association with hydrogenotrophic methanogens (Methanomicrobia or Methanobacteria) (Karakashev et al., 2006).

Finally, it can be mentioned that the molecular tools used in this study allowed us to link activated metabolic pathways to a diversity of prokaryotes under two different experimental conditions that differed mainly in available energy sources, dominant taxa, ecological interactions, and metabolic pathways for methanogenesis. **Figure 4** shows a general scheme depicting the main differences found across experimental conditions and that we believe are directly linked to the applied enzymatic pretreatment.

AUTHOR CONTRIBUTIONS

OC conceived the study, designed, and performed the experiments, evaluated the data, and drafted the manuscript. AS-R performed bioinformatics analyzes and drafted the manuscript. RC and LG supervised the work and assisted in drafting the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

OC appreciates her scholarship funded by the CONICYT, Beca Nacional Doctorado. 21121012.

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available Online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Angelidaki, I., Alves, M., Bolzonella, D., Borzacconi, L., Campos, J. L., Guwy, A. J., et al. (2009). Defining the biomethane potential (BMP) of solid organic wastes and energy crops: a proposed protocol for batch assays. *Water Sci. Technol.* 59, 927–934. doi: 10.2166/wst.2009.040

Angelidaki, I., and Batstone, D. (2010). Anaerobic digestion. *Solid Waste Technol. Manag.* 1–2, 583–600. doi: 10.1002/9780470666883.ch37

APHA-AWWA-WPCF (1999). Standard Methods for the Examination of Water and Wastewater, 20th Edn. Washington, DC.

Bareither, C. A., Wolfe, G. L., McMahon, K. D., and Benson, C. H. (2013). Microbial diversity and dynamics during methane production from municipal solid waste. *Waste Manag.* 33, 1982–1992. doi: 10.1016/j.wasman.2012.12.013

Begley, T. P., Kinsland, C., and Strauss, E. (2001). The biosynthesis of coenzyme A in bacteria. *Vitam. Horm.* 61, 157–171. doi: 10.1016/S0083-6729(01)61005-7

Bochner, B. R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiol. Rev.* 33, 191–205. doi: 10.1111/j.1574-6976.2008.00149.x

Bohutskyi, P., and Bouwer, E. (2013). "Biogas production from algae and cyanobacteria through anaerobic digestion: a review, analysis, and research needs," in *Advanced Biofuels and Bioproducts*, ed W. J. Lee (New York, NY: Springer New York), 873–975.

Donoso-Bravo, A., Pérez-Elvira, S. I., and Fdz-Polanco, F. (2010). Application of simplified models for anaerobic biodegradability tests. Evaluation of pre-treatment processes. *Chem. Eng. J.* 160, 607–614. doi: 10.1016/j.cej.2010.03.082

Dubois, M., Gilles, K., Hamilton, J. K., Rebers, P. A., and Smith, F. (1951). A colorimetric method for the determination of sugars. *Nature* 168:167. doi: 10.1038/168167a0

Gerardi, M. (2003). *The Microbiology of Anaerobic Digesters*. Hoboken, NJ: John Wiley & Sons, Inc.

González-Fernández, C., Sialve, B., Bernet, N., and Steyer, J.-P. (2012). Impact of microalgae characteristics on their conversion to biofuel. Part II: focus on biomethane production. *Biofuels Bioprod. Biorefining* 6, 205–218. doi: 10.1002/bbb.337

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. 29, 644–652. doi: 10.1038/nbt.1883

Janecek, S. (1997). α-amylase family: molecular biology and evolution. *J. Prog. Biophys. Mol. Biol.* 67, 67–97. doi: 10.1016/S0079-6107(97)00015-1

Jansson, J. K., Neufeld, J. D., Moran, M. A., and Gilbert, J. A. (2012). Omics for understanding microbial functional dynamics. *Environ. Microbiol.* 14, 1–3. doi: 10.1111/j.1462-2920.2011.02518.x.

Jiang, J., Zhao, Q., Wei, L., Wang, K., and Lee, D. J. (2011). Degradation and characteristic changes of organic matter in sewage sludge using microbial fuel cell with ultrasound pretreatment. *Bioresour. Technol.* 102, 272–277. doi: 10.1016/j.biortech.2010.04.066

Kampmann, K., Ratering, S., Kramer, I., Schmidt, M., Zerr, W., and Schnell, S. (2012). Unexpected stability of Bacteroidetes and Firmicutes communities in laboratory biogas reactors fed with different defined substrates. *Appl. Environ. Microbiol.* 78, 2106–2119. doi: 10.1128/AEM.06394-11

Karakashev, D., Batstone, D. J., Trably, E., and Angelidaki, I. (2006). Acetate oxidation is the dominant methanogenic pathway from acetate in the absence of Methanosaetaceae. *Appl. Environ. Microbiol.* 72, 5138–5141. doi: 10.1128/AEM.00489-06

Kato, S., and Watanabe, K. (2010). Ecological and evolutionary interactions in syntrophic methanogenic consortia. *Microbes Environ.* 25, 145–151. doi: 10.1264/jsme2.ME10122

Lee, C., Kim, J., Hwang, K., O'Flaherty, V., and Hwang, S. (2009). Quantitative analysis of methanogenic community dynamics in three anaerobic batch digesters treating different wastewaters. *Water Res.* 43, 157–165. doi: 10.1016/j.watres.2008.09.032

Lee, S. Y., Park, J. H., Jang, S. H., Nielsen, L. K., Kim, J., and Jung, K. S. (2008). Fermentative butanol production by clostridia. *Biotechnol. Bioeng.* 101, 209–228. doi: 10.1002/bit.22003

Li, B., and Dewey, C. N. (2011). RSEM : accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323.

Mahdy, A., Ballesteros, M., and González-Fernández, C. (2016). Enzymatic pretreatment of *Chlorella vulgaris* for biogas production: influence of urban wastewater as a sole nutrient source on macromolecular profile and biocatalyst efficiency. *Bioresour. Technol.* 199, 319–325. doi: 10.1016/j.biortech.2015.08.080

Mendez, L., Mahdy, A., Timmers, R. A., Ballesteros, M., and González-Fernández, C. (2013). Enhancing methane production of *Chlorella vulgaris* via thermochemical pretreatments. *Bioresour. Technol.* 149, 136–141. doi: 10.1016/j.biortech.2013.08.136

Mishra, A., and Jha, B. (2009). Isolation and characterization of extracellular polymeric substances from micro-algae *Dunaliella salina* under salt stress. *Bioresour. Technol.* 100, 3382–3386. doi: 10.1016/j.biortech.2009.02.006

Mitsutani, A., Izumi, Y., Kitaguchi, H., Kat, J., Ven, S., and Ishida, Y. (2001). Analysis of algicidal proteins of a diatom-lytic marine bacterium *Pseudoalteromonas* sp. strain A25 by two-dimensional electrophoresis. *Phycologia* 40, 286–291. doi: 10.2216/i0031-8884-40-3-286.1

Moran, M. A. (2009). Metatranscriptomics: eavesdropping on complex microbial communities. *Microbe* 4, 329–335. doi: 10.1128/microbe.4.329.1

Moset, V., Poulsen, M., Wahid, R., Højberg, O., and Møller, H. B. (2015). Mesophilic versus thermophilic anaerobic digestion of cattle manure: methane productivity and microbial ecology. *Microb. Biotechnol.* 8, 787–800. doi: 10.1111/1751-7915.12271

Muyzer, G., and Stams, A. J. (2008). The ecology and biotechnology of sulphate-reducing bacteria. *Nat. Rev. Microbiol.* 6, 441–454. doi: 10.1038/nrmicro1892

Nolla-Ardèvol, V., Strous, M., and Tegetmeyer, H. E. (2015). Anaerobic digestion of the microalga Spirulina at extreme alkaline conditions: biogas production, metagenome and metatranscriptome. *Front. Microbiol.* 6:597. doi: 10.3389/fmicb.2015.00597

Owusu-Apenten, R. (2002). *Food Protein Analysis: Quantitative Effects on Processing*. New York, NY: CRC Press.

Passos, F., Astals, S., and Ferrer, I. (2014). Anaerobic digestion of microalgal biomass after ultrasound pretreatment. *Waste Manag.* 34, 2098–2103. doi: 10.1016/j.wasman.2014.06.004

Safi, C., Charton, M., Pignolet, O., Silvestre, F., Vaca-Garcia, C., and Pontalier, P. Y. (2013). Influence of microalgae cell wall characteristics on protein extractability and determination of nitrogen-to-protein conversion factors. *J. Appl. Phycol.* 25, 523–529. doi: 10.1007/s10811-012-9886-1

Sato, M., Murata, Y., Mizusawa, M., Iwahashi, H., and Oka, S. (2004). A simple and rapid dual-fluorescence viability assay for microalgae. *Microbiol Cult. Coll.* 20, 53–59.

Shehata, A., Schrödi, W., Neuhaus, J., and Krüger, M. (2013). Antagonistic effect of different bacteria on Clostridium botulinum types A, B, D and E *in vitro*. *Vet. Rec.* 172:47. doi: 10.1136/vr.101184

Shihira, I., and Krauss, R. W. (1965). *Chlorella Physiology and Taxonomy of Forty-One Isolates*. College Park, MD: University of Maryland.

Sialve, B., Bernet, N., and Bernard, O. (2009). Anaerobic digestion of microalgae as a necessary step to make microalgal biodiesel sustainable. *Biotechnol. Adv.* 27, 409–416. doi: 10.1016/j.biotechadv.2009.03.001

Stolze, Y., Zakrzewski, M., Maus, I., Eikmeyer, F., Jaenicke, S., Rottmann, N., et al. (2015). Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. *Biotechnol. Biofuels* 8:14. doi: 10.1186/s13068-014-0193-8

Sueoka, N. (1960). Mitotic replication of deoxyribonucleic acid in chlamydomonas reinhardi. *Proc. Natl. Acad. Sci. U.S.A.* 46, 83–91. doi: 10.1073/pnas.46.1.83

Szymanowska-Powałowska, D., Orczyk, D., and Leja, K. (2014). Biotechnological potential of *Clostridium butyricum* bacteria. *Braz. J. Microbiol.* 45, 892–901. doi: 10.1590/S1517-83822014000300019

Takeda, H. (1991). Sugar composition of the cell wall and the taxonomy of chlorella (Chlorophyceae)1. *J. Phycol.* 27, 224–232. doi: 10.1111/j.0022-3646.1991.00224.x

Temate-tiagueu, Y., Seesi, S. A., Mathew, M., Mandric, I., Rodriguez, A., Bean, K., et al. (2016). Inferring metabolic pathway activity levels from RNA-Seq data. *BMC Genomics* 17(Suppl 5):542. doi: 10.1186/s12864-016-2823-y

Templeton, D. W., Quinn, M., Van Wychen, S., Hyman, D., and Laurens, L. M. (2012). Separation and quantification of microalgal carbohydrates. *J. Chromatogr. A* 1270, 225–234. doi: 10.1016/j.chroma.2012.10.034

Vanwonterghem, I., Jensen, P. D., Ho, D. P., Batstone, D. J., and Tyson, G. W. (2014). Linking microbial community structure, interactions and function in anaerobic digesters using new molecular techniques. *Curr. Opin. Biotechnol.* 27, 55–64. doi: 10.1016/j.copbio.2013.11.004

Walker, C. B., Redding-Johanson, A. M., Baidoo, E. E., Rajeev, L., He, Z., Hendrickson, E. L., et al. (2012). Functional responses of methanogenic archaea to syntrophic growth. *ISME J.* 6, 2045–2055. doi: 10.1038/ismej.2012.60

Waters, C. M., and Bassler, B. L. (2005). Quorum sensing: communication in bacteria. *Annu. Rev. Cell Dev. Biol.* 21, 319–346. doi: 10.1146/annurev.cellbio.21.012704.131001

Williams, P. (2017). SGM Special Lecture Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* 153(Pt 12), 3923–3938. doi: 10.1099/mic.0.2007/012856-0

Zamalloa, C., Vulsteke, E., Albrecht, J., and Verstraete, W. (2011). The techno-economic potential of renewable energy through the anaerobic digestion of microalgae. *Bioresour. Technol.* 102, 1149–1158. doi: 10.1016/j.biortech.2010.09.017

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple "omics" analysis for microbial biology: application and methodologies. *Microbiology* 156, 287–301. doi: 10.1099/mic.0.034793-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Córdova, Chamy, Guerrero and Sánchez-Rodríguez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 March 2018
doi: 10.3389/fmicb.2018.00540

Linking Microbial Community Structure and Function During the Acidified Anaerobic Digestion of Grass

*Aoife Joyce[1], Umer Z. Ijaz[2], Corine Nzeteu[1,3], Aoife Vaughan[3], Sally L. Shirran[4], Catherine H. Botting[4], Christopher Quince[5], Vincent O'Flaherty[3] and Florence Abram[1]**

[1] Functional Environmental Microbiology, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland, [2] Environmental Omics Laboratory, School of Engineering, University of Glasgow, Glasgow, United Kingdom, [3] Microbial Ecology Laboratory, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland, [4] Biomedical Sciences Research Complex, University of St Andrews, Fife, United Kingdom, [5] Microbiology and Infection, Warwick Medical School, University of Warwick, Coventry, United Kingdom

OPEN ACCESS

Edited by:
Jean Armengaud, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), France

Reviewed by:
Heike Sträuber, Helmholtz-Zentrum für Umweltforschung (UFZ), Germany
Seung Gu Shin, Pohang University of Science and Technology, South Korea

***Correspondence:**
Florence Abram
florence.abram@nuigalway.ie

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: 17 October 2017
Accepted: 09 March 2018
Published: 21 March 2018

Citation:
Joyce A, Ijaz UZ, Nzeteu C, Vaughan A, Shirran SL, Botting CH, Quince C, O'Flaherty V and Abram F (2018) Linking Microbial Community Structure and Function During the Acidified Anaerobic Digestion of Grass. Front. Microbiol. 9:540. doi: 10.3389/fmicb.2018.00540

Harvesting valuable bioproducts from various renewable feedstocks is necessary for the critical development of a sustainable bioeconomy. Anaerobic digestion is a well-established technology for the conversion of wastewater and solid feedstocks to energy with the additional potential for production of process intermediates of high market values (e.g., carboxylates). In recent years, first-generation biofuels typically derived from food crops have been widely utilized as a renewable source of energy. The environmental and socioeconomic limitations of such strategy, however, have led to the development of second-generation biofuels utilizing, amongst other feedstocks, lignocellulosic biomass. In this context, the anaerobic digestion of perennial grass holds great promise for the conversion of sustainable renewable feedstock to energy and other process intermediates. The advancement of this technology however, and its implementation for industrial applications, relies on a greater understanding of the microbiome underpinning the process. To this end, microbial communities recovered from replicated anaerobic bioreactors digesting grass were analyzed. The bioreactors leachates were not buffered and acidic pH (between 5.5 and 6.3) prevailed at the time of sampling as a result of microbial activities. Community composition and transcriptionally active taxa were examined using 16S rRNA sequencing and microbial functions were investigated using metaproteomics. Bioreactor fraction, i.e., grass or leachate, was found to be the main discriminator of community analysis across the three molecular level of investigation (DNA, RNA, and proteins). Six taxa, namely Bacteroidia, Betaproteobacteria, Clostridia, Gammaproteobacteria, Methanomicrobia, and Negativicutes accounted for the large majority of the three datasets. The initial stages of grass hydrolysis were carried out by Bacteroidia, Gammaproteobacteria, and Negativicutes in the grass biofilms, in addition to Clostridia in the bioreactor leachates. Numerous glycolytic enzymes and carbohydrate transporters were detected throughout the bioreactors in addition to proteins involved in butanol and lactate production. Finally, evidence of the prevalence of stressful

conditions within the bioreactors and particularly impacting Clostridia was observed in the metaproteomes. Taken together, this study highlights the functional importance of Clostridia during the anaerobic digestion of grass and thus research avenues allowing members of this taxon to thrive should be explored.

Keywords: anaerobic digestion, cellulosic substrate, 16S rRNA profiling, metaproteomics, biomolecule co-extraction

INTRODUCTION

The development of the bioeconomy is critical in attaining several of the UN Sustainable Development Goals for 2030 (United Nations, 2015), including SDG7 (renewable energy), SDG8 (good jobs and economic growth) and SDG13 (climate action), as well as achieving 20% energy production in Europe from renewable sources by 2020 (Vega et al., 2014). In this context harvesting valuable bioproducts from various waste streams becomes a necessary element for the growth of a sustainable bioeconomy worldwide (Werner et al., 2011). Anaerobic digestion (AD) is a well-established sustainable technology for the treatment of a diverse range of wastewaters (Adulkar and Rathod, 2014; Mustafa et al., 2014; Yuan et al., 2014; Lackey et al., 2015; Wang et al., 2015; Zerrouki et al., 2015), as well as solid wastes (Wang et al., 2014; Michele et al., 2015; Nielfa et al., 2015; Ratanatamokul and Saleart, 2016) converting waste streams to energy produced in the form of biogas.

Recently, there has been a considerable increase in energy crops usage in the biogas industry (Vega et al., 2014) with corn representing one of the favorite feedstocks. However, using such crops for energy production is directly competing with other industries (Ranum et al., 2014). In addition, the crops require close management, as they have to be replanted and typically involve the use of fertilizers and pesticides for successful growth. Therefore alternative feedstocks need to be identified as bioproducts sources. Grass is such an alternative feedstock, as it grows naturally in many areas, with minimal labor necessary to sustain it. Grass is estimated to represent 70% of agricultural land worldwide and cover 40% of terrestrial surfaces (Cerrone et al., 2014). It can grow on soils unsuitable for other crops and its production has been estimated, in Ireland for example, to exceed livestock requirements by about 1.7 million tons dry solids each year (McEniry et al., 2013). As such, grass represents a promising second-generation biomass resource. Perennial ryegrass has been demonstrated to be a suitable feedstock for AD (Cysneiros et al., 2011), leading to the production of energy with the potential for recovery of process intermediates of high market values (Cerrone et al., 2014).

The natural process of AD is driven by the concerted, sequential and cooperative activities of several microbial trophic groups. Broadly, four main steps can be distinguished during the process: (i) hydrolysis where polymers are converted to monomers; (ii) acidogenesis leading to volatile fatty acid production; (iii) acetogenesis leading to acetate and H_2/CO_2 generation and finally; (iv) methanogenesis where acetate and H_2/CO_2 are converted to CH_4 (Narihiro and Sekiguchi, 2007). Even though microbial consortia clearly underpin AD, the relationship between process performance and microbial community composition and functioning have yet to be adequately characterized (Amha et al., 2018). Reactors are typically designed and operated on the basis of empirical relationships between reactor performance and process parameters, bypassing microbial processes undeniably at the core of AD. As a result, process instability and failures, due to, for example, the accumulation of free ammonia, volatile fatty acids, long chain fatty acids and low pH, are still common and poorly understood (Amha et al., 2018). Thus the advancement of the technology relies on a greater insight and understanding of the behavior of the AD microbiome. There is, however, limited knowledge of the functional activities of the microbial consortia present in AD systems (Abram et al., 2011; Abdul et al., 2014), and this is especially true for the AD of solid feedstocks.

Recent technological developments, and specifically the advancement of high-throughput omics, have allowed for the possibility of system approaches to be explored (Siggins et al., 2012a; Abram, 2015; Narayanasamy et al., 2015). Particularly, metaproteomics can be used to determine key metabolic pathways and functional activities occurring in a given ecosystem at the time of sampling. Proteins identified can support the characterization of microbial groups involved in specific functions via protein assignment. Metaproteomics has been applied to many diverse environments including marine, freshwater, soil, human biology as well as natural and bioengineered systems (Siggins et al., 2012b; Wilmes et al., 2015). It has also been previously employed to uncover key biochemical metabolic pathways occurring in anaerobic bioreactor treating wastewater (Abram et al., 2011; Siggins et al., 2012b; Hettich et al., 2013; Gunnigle et al., 2015a,b; Heyer et al., 2016) and more recently solid feedstocks (Kohrs et al., 2014; Lü et al., 2014; Theuerl et al., 2015; Heyer et al., 2016; Abendroth et al., 2017). Here, we report on the investigation of the microbial community structure and function in triplicate anaerobic bioreactors digesting grass. The leachates of the reactors were not buffered, in order to favor the accumulation of process intermediates as a result of methanogenesis inhibition via acidification. 16S rRNA amplicon profiling from DNA and cDNA samples was combined with metaproteomics in an effort to link the knowledge obtained from sequencing data (community structure) to the functional activities taking place at the time of sampling.

MATERIALS AND METHODS

Bioreactor Operation and Sampling

Triplicate leach-bed bioreactors (R1, R2, and R3), with a working volume of 4 L, were operated at 37°C in a semi-continuous

mode with a solid retention time of 7 days as previously described (Cysneiros et al., 2012). For the first batch, the triplicate bioreactors were seeded with 84 g volatile solids (VS) of pressed ensiled ryegrass and 126 g VS of anaerobic granular sludge from a full-scale mesophilic reactor (Carbery Milk Products, Ireland) to which 3.2 L of water supplemented with trace elements (0.2 mM $MnCl_2$, 0.2 mM H_3BO_3, 0.1 mM $ZnCl_2$, 0.06 mM $CuCl_2$, 0.01 mM $NaHSO_4$, 0.6 mM $CaCl_2$, 0.07 mM $NiCl_2$ and 0.1 mM SeO_2) were added. The leachate was recirculated in a down-flow mode using a peristaltic pump. At the end of each bioreactor run (7 days), 126 g VS of digestate were used to inoculate the next batch to which 84 g VS of ensiled pressed grass and 1.6 L of leachate from the previous batch supplemented with 1.6 L of freshly prepared leachate (water and trace elements) were added. VS analysis was performed gravimetrically according to the standard method of American Public Health Association [APHA] (2005). At the time of sampling, the bioreactors had been operated for 63 consecutive batches (each of 7 days duration). Duplicate 250 ml leachate and 50 g digestate samples were taken from each of the triplicate bioreactors (R1, R2, and R3) on the last day of the 63rd batch (day 7), when VS removal was 80, 81, and 73% for R1, R2, and R3, respectively. The pH of the reactors was allowed to fluctuate naturally in order to inhibit methane production and in turn favor process intermediates accumulation. A summary of the bioreactors' performance is presented in **Supplementary Table S1**. Chemical oxygen demand (COD) measurements were performed according to the Standing Committee of Analysts (1985). The leachate samples were centrifuged at 8,000 × *g* for 15 min at 4°C, prior to resuspension in 1 ml of 10 mM Tris Base, 0.1 mM EDTA and 5 mM $MgCl_2$ (resuspension buffer) and storage on ice before further use. Digestate samples were carefully drained, then immersed in 250 ml of resuspension buffer and placed in a sonication bath for 5 min to gently detach the grass biofilms. After grass removal, the digestate samples were filtered through two layers of muslin cloth twice before centrifugation at 8,000 × *g* for 15 min at 4°C. The resulting pellets were resuspended in 1 ml of resuspension buffer. Leachate and digestate samples were then centrifuged at 17,000 × *g* for 10 min before undergoing the following series of washes (Roume et al., 2012): twice with 1 ml of 0.9% NaCl, then twice with 1 ml 50 mM Tris-HCl and finally once with 1 ml resuspension buffer. The pellets were then flash frozen in liquid nitrogen and stored at −80°C until further use.

High-Throughput 16S rRNA Sequencing and Bioinformatic Analysis

DNA, RNA, and proteins were co-extracted from digestate and leachate samples using the RNA/DNA/Protein Purification kit from Norgen Biotek. Briefly, digestate and leachate cell pellets were resuspended in 500 µl lysis solution and 500 µl 1X Tris-EDTA to which 10 µl ml^{-1} β-mercaptoethanol were added. Cell lysis was carried out by bead beating for 30 s using zirconia beads (0.5 ml: 0.1 mm and 0.5 mm diameter in 1:1 ratio). The samples were then centrifuged at 17,000 × *g* for 30 min, and this step was repeated until no pellet was visible. The resulting supernatants were supplemented with 100 µl pure ethanol and loaded onto all-in-one chromatographic spin columns (Norgen Biotek). Purification and isolation of DNA, RNA and proteins were carried out following the manufacturer's recommendations. DNase treatment of RNA samples was performed using the Turbo DNA-free kit (Ambion). Control PCRs using DNase treated products as templates were carried out to ensure that no DNA remained in the RNA samples prior to cDNA generation using SuperScript III Reverse Transcriptase (Invitrogen), flash freezing in liquid nitrogen and storage at −80°C. Both cDNA and DNA samples were prepared for paired-end 16S rRNA sequencing using Illumina Miseq platform and Golay barcodes. Amplification of the 16S rRNA gene from DNA and cDNA samples was carried out in triplicate 25 µl reactions using 515F/806R primers (targeting the V4 region; Caporaso et al., 2011) and the Q5® High Fidelity DNA Polymerase kit (New England Biolabs) as follows: 1X Q5® reaction buffer, 200 µM dNTPs, 0.5 µM of each primer, 0.02 U μl^{-1} of Q5® TAQ polymerase and 500 ng of template. PCR conditions consisted of a hotstart at 98°C for 30 s, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 52°C for 30 s and elongation at 72°C for 30 s and a final elongation step at 72°C for 2 min. The replicate amplicons were then pooled and quantified using the Qubit dsDNA HS Assay kit (Life Technologies) following the manufacturers' instructions. Samples were normalized to 3 ng μl^{-1} and pooled together, prior to Illumina sequencing. A total of 24 samples were analyzed, corresponding to duplicate samples from both grass biofilms and leachate fractions from the triplicate bioreactors. Illumina sequencing was carried out by the Centre for Genomic Research (Liverpool, United Kingdom) and generated a total of 2.13 10^7 reads corresponding to 1.14 × 10^7 and 9.98 × 10^6 DNA and cDNA sequences, respectively. Only 20 reads were obtained for one of the duplicate grass biofilm cDNA samples from R1, which was therefore dropped from further analysis. Sequencing data were analyzed using the Illumina Amplicon Processing Workflow available at: http://userweb.eng.gla.ac.uk/umer.ijaz/bioinformatics/Illumina_workflow.html. Briefly, paired-end reads were trimmed, overlapped and assembled, prior to OTU clustering and chimera removal using the gold database from UCHIME. Phylogenetic trees and OTU assignments were carried out using MUSCLE. The DNA and cDNA sequences were deposited on NCBI's Sequence Read Archive under the accession number SRP119456.

Metaproteomics

Protein concentrations were determined using the Calbiochem Non-Interfering Protein Assay™ kit (Merck KGaA, Darmstadt, Germany), following the manufacturer's instructions. Protein samples were normalized to a concentration of 1.3 µg μl^{-1} and analyzed by GeLC MS/MS (Dzieciatkowska et al., 2014) as follows: 52 µg of each sample were loaded onto SDS-PAGE and the proteins separated along the length of the gels. The protein samples were fractionated to reduce complexity by excising the top, middle and bottom third of each lane, which were analyzed separately. In-gel digestion, protein reduction and alkylation as well as tryptic digestion were performed

prior to peptide extraction with 10% formic acid as previously described (Shevchenko et al., 1996). The resulting peptides were then concentrated using a SpeedVac concentrator (Thermo Savant) before separation on an Acclaim PepMap C18 trap and an Acclaim PepMap RSLC C18 column (Thermo Fisher Scientific), using a nanoLC Ultra 2D plus loading pump and nanoLC AS-2 autosampler (Eksigent, Redwood City, CA, United States). The peptides were then eluted with a gradient of acetonitrile, containing 0.1% formic acid (1–40% acetonitrile in 60 min, 40–99% in a further 10 min, followed by washing with 99% acetonitrile for 5 min before re-equilibration with 1% acetonitrile). The eluate was sprayed into a TripleTOF 5600 electrospray tandem mass spectrometer (Sciex, Foster City, CA, United States) and analysis was carried out in Information Dependent Acquisition (IDA) mode, performing 250 ms of MS followed by 100 ms MS/MS analyses on the 20 most intense peaks. MS/MS data were processed with ProteinPilot v4.5 software (Sciex) using the Paragon search algorithm. The resulting mass spectra were searched against the TrEMBL database, using the following search parameters: cysteine alkylation with iodoacetamide, 'Gel-based ID' for 'Special Factors,' 'Biological modifications' for 'ID focus,' and a 'Thorough' 'Search effort.' NCBI and Swiss-Prot databases searches were also carried out and led to similar results (data not shown). Generalist databases were chosen over custom-build databases, composed of representatives of species identified in the DNA and cDNA gene marker sequencing datasets, to avoid transferring the inherent PCR bias typically associated with 16S rRNA profiling to the metaproteomic analysis. In order to mitigate the number of false positive associated with the use of generalist databases a stringent confidence cut-off of 10% was applied. The mass spectrometry metaproteomics data along with the corresponding FDR analysis of each gel chunk were deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD007956. A summary of the number of MS/MS acquired, MS/MS assigned to peptides and the number of distinct peptides for each gel chunk is displayed in **Supplementary Table S2**. A threshold of unused Protscore (from ProteinPilot) of 2 (corresponding to protein detection with ≥99% confidence) and a minimum of two peptides were employed for protein identification. When protein assignment was ambiguous, i.e., when a protein was assigned to multiple species, the lowest common ancestor is reported. Analysis of clusters of orthologous groups (COGs) was carried out using MEGAN5[1] and an overview of metabolic pathways from which proteins were identified was generated using the Metaproteomics Data Analysis Workflow available at http://userweb.eng.gla.ac.uk/umer.ijaz/bioinformatics/Metaproteomics.html. In this pipeline, the enzyme commission (EC) numbers corresponding to the identified proteins are retrieved when available, MinPath (Ye and Doak, 2009) is used to construct parsimonious pathways and iPath2.0 (Yamada et al., 2011) is employed for pathway visualization. Krona plots were constructed using the Krona template (Ondov et al., 2011) and Circos plots using the Circos online tool developed by Krzywinski et al. (2009).

[1]http://ab.inf.uni-tuebingen.de/software/megan5/

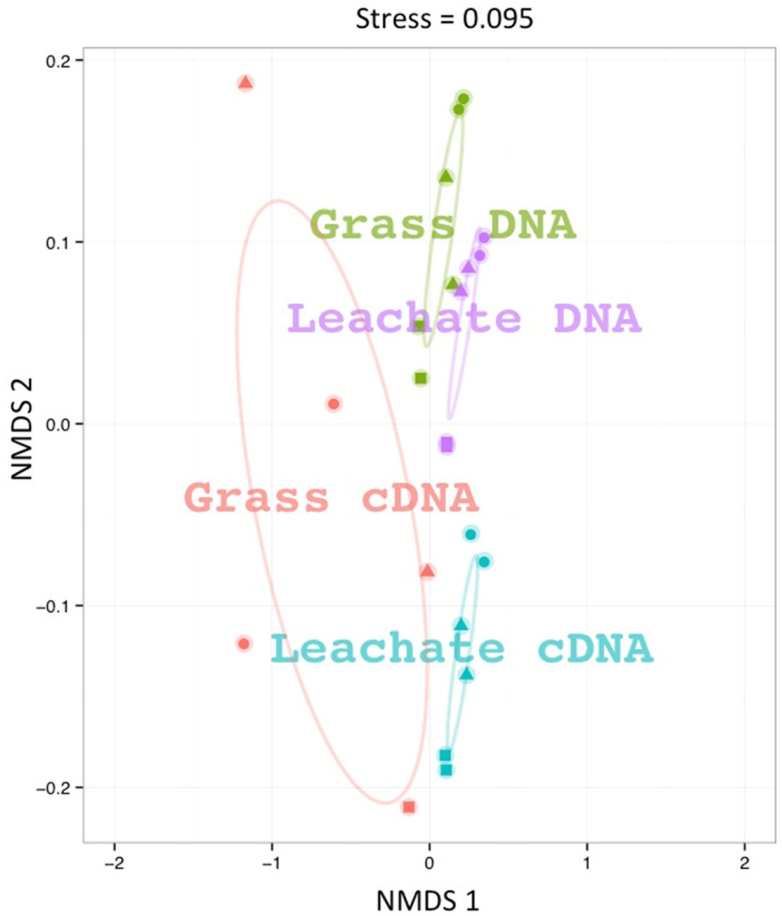
Statistical Analyses

Non-metric multidimensional scaling (NMDS) based on Bray–Curtis distances was performed using the statistical program R (R Core Team, 2017) to compare microbial community dissimilarities among grass and leachate samples in (i) DNA and cDNA; and (ii) metaproteomics datasets. The group labels are drawn at mean of the ordination values of the samples for that particular group, and the ellipses represent the 95% confidence interval of the standard error of ordination for a given group. To assess the statistical significance of the sample groupings, an analysis of similarity (ANOSIM) was carried out using R. Ratios were calculated, to compare grass and leachate datasets, using $(n_c/n)/(N_c/N)$, where n_c is the number of hits to a given category 'c' (i.e., taxonomic assignments) in a specific grass dataset (i.e., DNA, cDNA, or proteins), n is the total number of hits in all categories in the same grass dataset, N_c is the number of hits to that category in the corresponding leachate dataset and N is the number of hits in all categories in the same leachate dataset. Ratios smaller than 1 indicate an under-representation of a specific category 'c' among grass samples compared to leachate samples, with ratios greater than 1 corresponding to an over-representation of a specific category in the grass samples. Statistical over- and under-representation of a given taxonomic assignment between two datasets was determined by pairwise comparisons using two-tailed Fishers' exact test with confidence intervals at 99% significance ($P_{adj} < 0.05$).

RESULTS

Microbiome Composition, and Transcriptionally and Translationally Active Taxa

16S rRNA profiling revealed a total of 1549 operational taxonomic units (OTUs) across the 24 samples analyzed (duplicate grass and leachate DNA, and cDNA, from the triplicate bioreactors). NMDS was used to visualize microbial community dissimilarities between the samples (stress value: 0.095; **Figure 1**). NMDS ordination positions each sample as a function of its distance from all other data points. An NMDS plot stress value below 0.1 indicates that the two-dimensional representation is ideal for data interpretation (Rees et al., 2004). Sample clustering was visually uncovered as a function of (i) bioreactor fraction, i.e., grass or leachate; and (ii) nucleic acid fraction, i.e., DNA or cDNA (**Figure 1**). The observed sample groupings were found to be statistically significant using ANOSIM, with the exception of DNA samples for which grass and leachate bioreactor fractions could not be satisfactorily distinguished ($P_{adj} > 0.05$; **Figure 1**). ANOSIM analysis also indicated that the samples did not cluster as a function of bioreactors ($P_{adj} > 0.05$; data not shown). Taken together, the results suggest that molecular (i.e., DNA



Type
- Grass cDNA
- Grass DNA
- Leachate cDNA
- Leachate DNA

Reactor
- R1
- R2
- R3

ANOSIM	R value	P_{adj}
DNA: Grass vs leachate	0.237	0.102
cDNA: Grass vs Leachate	0.485	0.018
Grass: DNA vs cDNA	0.605	0.018
Leachate: DNA vs cDNA	0.476	0.020

FIGURE 1 | Non-metric multidimensional scaling (NMDS) plot based on Bray–Curtis distances of the 16S rRNA sequences from grass biofilm (DNA represented in green and cDNA in red) and leachate (DNA represented in purple and cDNA in blue) fractions from bioreactors R1 (squares), R2 (circles) and R3 (triangles). Analysis of similarity (ANOSIM) was carried out to assess the statistical significance of sample groupings and the corresponding *R*-values and corrected P_{adj}-values are displayed.

and cDNA) and bioreactor (i.e., grass and leachate) fractions were the main drivers of microbial community structure. Six taxa, namely Bacteroidia, Betaproteobacteria, Clostridia, Gammaproteobacteria, Methanomicrobia, and Negativicutes, in addition to sequences classified as unknown, accounted for up to 93% of OTUs in the DNA, and 98% in the cDNA, datasets (**Figures 2A,B**). Proteins were also assigned predominantly to these six phylogenetic classes (**Figure 2C**) with only a few proteins assigned to unknown species (**Figure 2C**), as protein-coding sequences from unknown microorganisms are unlikely to be present in the NCBInr database. In addition, the overall contribution of these six phylogenetic classes to protein assignment was less than the contribution to DNA and cDNA. This was attributed to a large proportion of proteins that shared a lowest common ancestor at a taxonomic level higher than class (indicated in brackets in **Figure 2C**). Unclassified OTUs accounted for up to 50% of the microbial community in cDNA samples, highlighting the likely contribution of yet unknown species to the AD of grass (**Figure 2B**). Overall, differences in the relative abundance of the six main taxa could be observed amongst the three molecular datasets (i.e., DNA, cDNA, and proteins). For example, Bacteroidia's relative abundance was higher in DNA and protein datasets compared to cDNA samples. Conversely, the reverse trend was observed for Negativicutes with an increased relative abundance amongst cDNA and protein datasets compared to DNA samples. Differences could also be seen between grass and leachate fractions at the three levels of molecular investigation (DNA, cDNA, and proteins, **Figures 2**, **3**). Bacteroidia and Clostridia were found to be over-represented in the leachate compared to grass biofilms in DNA, cDNA, and protein samples (**Figure 3**). Similarly, Gammaproteobacteria and Methanomicrobia were over-represented in the grass biofilm datasets across the three level of molecular information (**Figure 3**). Negativicutes were under-represented in the grass fraction in both DNA and cDNA datasets while a statistically significant increased number of proteins were assigned to this taxon in the grass fraction when compared to leachate samples (**Figure 3**). This observation is unlikely resulting from a bias in genome availability, as only 84 Negativicutes full genomes are currently available in the NCBI

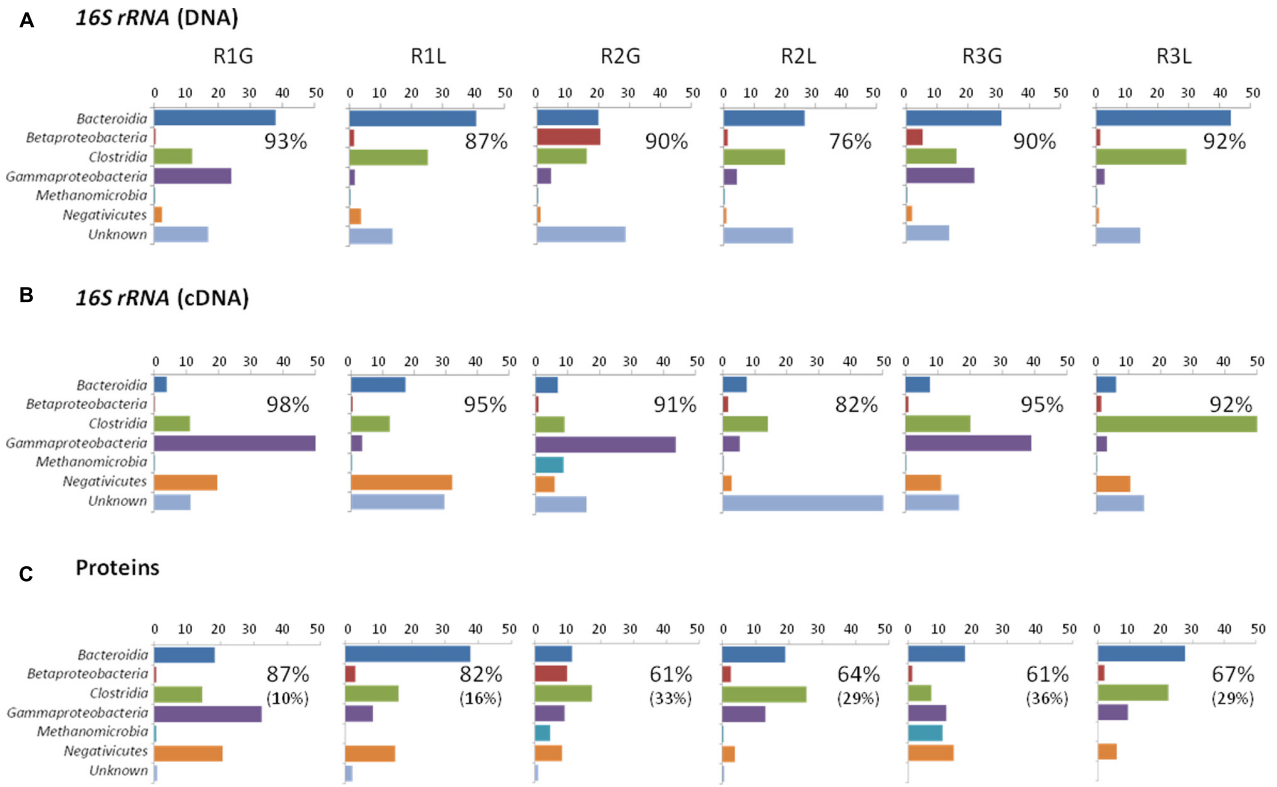


FIGURE 2 | Taxonomic composition, transcriptionally, and translationally active taxa of the grass biofilm and leachate communities from the triplicate reactors R1, R2, and R3. G stands for grass and L for leachate. Community composition analysis was based on taxonomic assignment of **(A)** 16S rRNA gene sequences from DNA samples. Transcriptionally and translationally active taxa analyses were based on taxonomic assignment of **(B)** 16S rRNA gene sequences from cDNA samples and **(C)** proteins. Percentage relative abundances are displayed. The single percentage number displayed in each panel corresponds to the total contribution of the seven taxonomic categories represented (six microbial taxa in addition to unknown) to the datasets. Numbers in brackets represent the percentage of proteins for which the corresponding assigned lowest common ancestor was of higher taxonomic level than class.

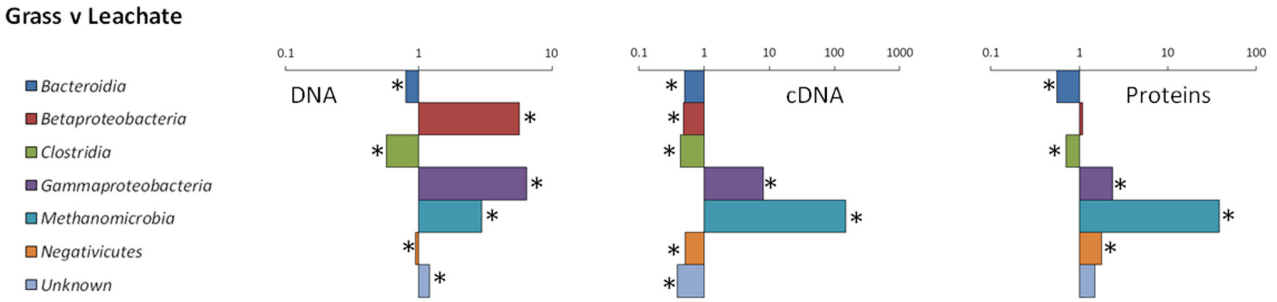


FIGURE 3 | Ratios of phylogenetic assignments in grass biofilm and leachate datasets. Ratios were calculated using $(n_c/n)/(N_c/N)$, where n_c is the number of hits (OTUs or proteins) to a given phylogenetic assignment in the grass datasets, n is the total number of hits in the corresponding grass datasets (DNA, cDNA, and proteins), N_c is the number of hits to that phylogenetic assignment in the leachate datasets, and N is the total number of hits in the corresponding leachate datasets. Statistically significant over- and under-representation of a given phylogenetic assignment between two datasets is represented by asterisks and was determined by pairwise comparisons using two-tailed Fishers' exact test with confidence intervals at 99% significance ($P_{adj} < 0.05$).

database against, for example, over 1880 Gammaproteobacteria genomes. Finally, Betaproteobacteria and OTUs classified as unknown displayed opposite trends in DNA and cDNA datasets where they were over-represented in the leachate and in the grass fractions, respectively (**Figure 3**). For these two microbial groups no statistically significant differential distribution between grass and leachate fractions could be identified in the protein samples.

Overview of Microbial Functions

A total of 1,830 proteins was detected across the 12 samples analyzed (duplicate leachate and grass extracts from the triplicate

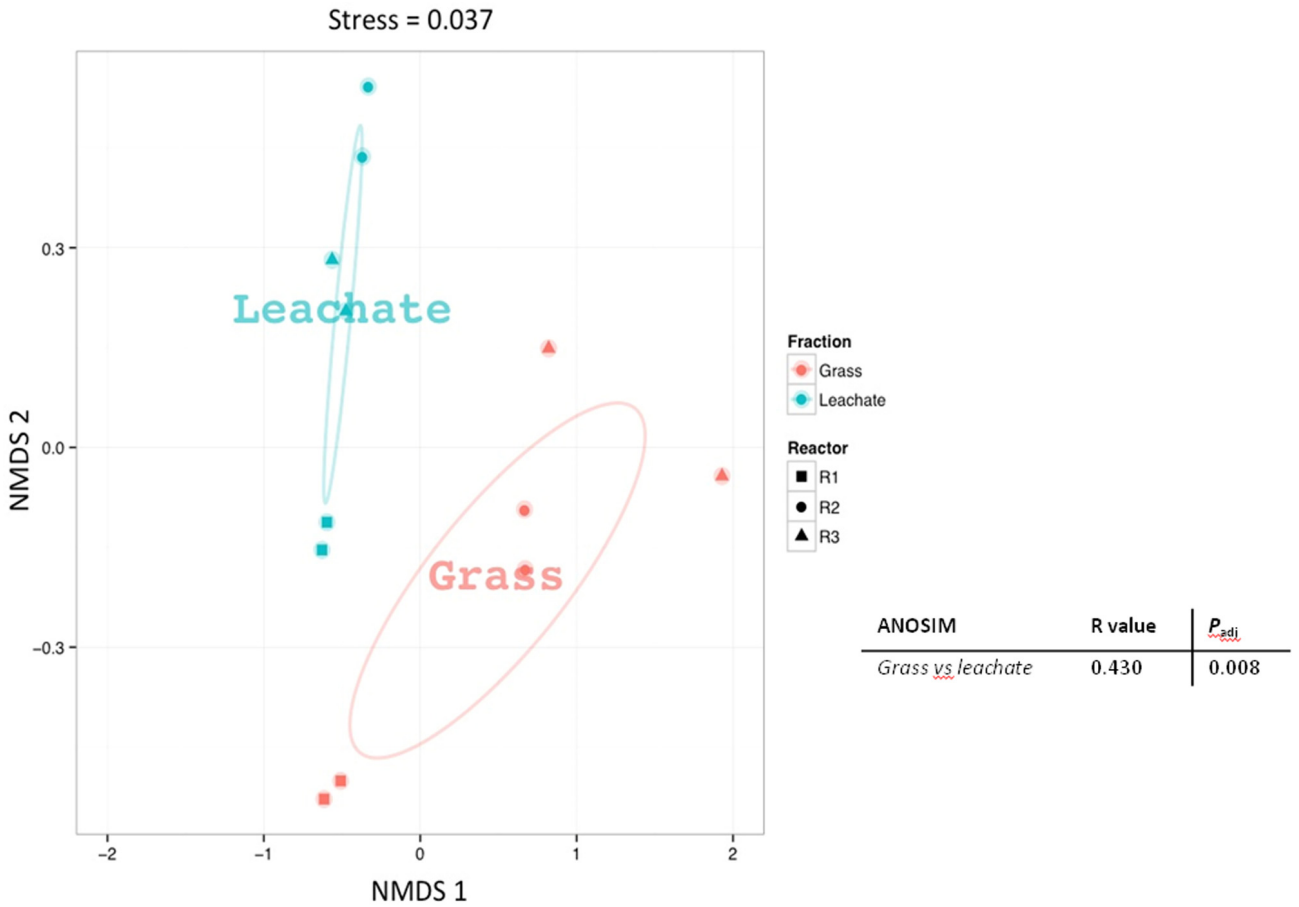


FIGURE 4 | Non-metric multidimensional scaling (NMDS) plot based on Bray–Curtis distances of the proteins extracted from grass biofilm (represented in red) and leachate (represented in blue) fractions from bioreactors R1 (squares), R2 (circles), and R3 (triangles). Analysis of similarity (ANOSIM) was carried out to assess the statistical significance of sample groupings and the corresponding *R*-values and corrected P_{adj}-values are displayed.

bioreactors; **Supplementary Table S3**), providing an overview of the metabolic pathways likely to be active at the time of sampling (**Supplementary Figure S1**). Grass is typically composed of hemicellulose (ranging from 35 to 50%), cellulose (25 to 40%), lignin (10 to 30%), free sugars (10 to 26%), and lipids (3%; Koch et al., 2010; Ellis et al., 2012). Evidence of the breakdown of grass components was observed in the metaproteomes. Proteins involved in carbohydrate metabolism (including cellulose and hemicellulose degradation, glycolysis and pentose phosphate pathway), energy metabolism (including TCA cycle, methanogenesis and lactate biosynthesis), lipid metabolism (including propionate metabolism, fatty acid β-oxidation and butanol production), amino acid metabolism (including glutamate fermentation to butyrate and nitrogen metabolism), nucleotide metabolism (including purine metabolism), as well as cofactors, and vitamins metabolism (including vitamin B6 biosynthesis) were detected in the triplicate bioreactors (**Supplementary Figure S1** and **Supplementary Table S3**). Sample dissimilarities were assessed with NMDS and ANOSIM, which indicated a clustering of the metaproteomic datasets as a function of bioreactor fraction, i.e., grass biofilms or leachate (stress value: 0.037 and $P_{adj} < 0.05$; **Figure 4**). Classifying microbial proteins into broad functional categories (i.e., cluster of orthologous genes; COG) did not, however, result in any statistically significant differential distribution between the grass and leachate metaproteomes (**Figure 5** and data not shown). The most abundant COG categories, collectively accounting for ∼80% of both grass and leachate proteins, were the following: translation (J), energy production and conversion (C), amino acid transport and metabolism (E), post-translational modification, protein turnover and chaperones (O) and carbohydrate transport and metabolism (G; **Figure 5**).

Grass Biodegradation

Evidence of cellulose (i.e., cellobiose hydrolysis) and hemicellulose (i.e., galactose, glycolate, xylan, and xylose catabolism) biodegradation was reflected in both grass and leachate metaproteomes (**Figure 6** and **Supplementary Table S3**). In addition, proteins involved in uronic acid degradation, a component of grass cell wall, could also be detected in both sample types (i.e., grass biofilms and leachate). A cellulose hydrolase (beta-glucosidase) assigned to

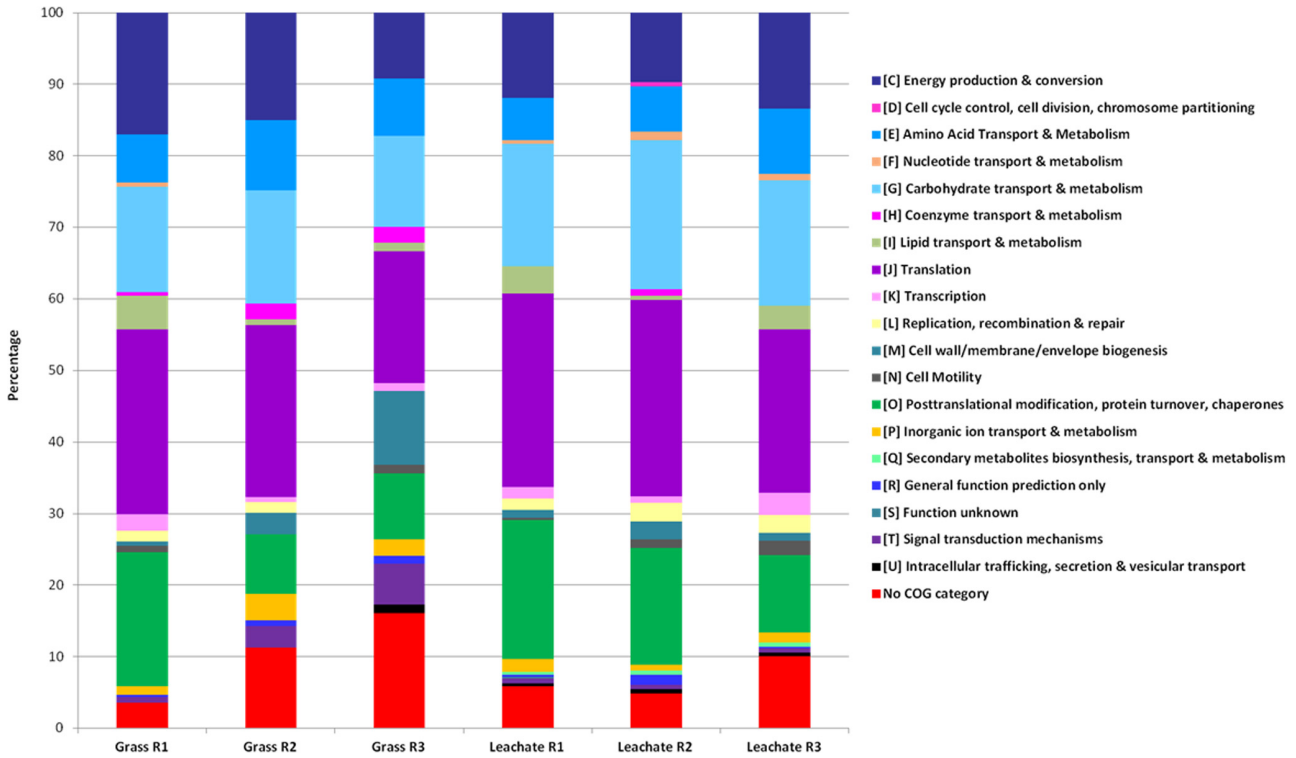


FIGURE 5 | Distribution of functional COG categories in the grass biofilm and leachate metaproteomes. R1, R2, and R3 stands for reactor 1, reactor 2, and reactor 3.

Bacteroidetes and a xylan hydrolase (beta-1,4-xylanase) assigned to *Clostridium sp*. could only be detected in the leachate samples (**Supplementary Table S3**). It is worth noting, however, that the non-detection of a protein does not necessarily imply that it was not expressed at the time of sampling. A galactose hydrolysing enzyme (beta-galactosidase), exclusively assigned to *E. coli* and a *Megasphaera elsdenii* protein involved in glycolate metabolism were detected in both bioreactor fractions (**Supplementary Table S3**). Xylose degradation was attributed to Bacteroidia in grass biofilms, in addition to Clostridia in the leachate samples (**Figure 6** and **Supplementary Table S3**). Numerous proteins with functions in carbohydrate transport were also detected in all the samples analyzed and mainly assigned to Clostridia and Spirochaetales. Glycolysis was found to take place in both grass and leachate fractions as indicated by a plethora of glycolytic enzymes predominantly assigned to Bacteroidia, Clostridia and Gammaproteobacteria (**Figure 6** and **Supplementary Table S3**).

Energy Production and Conversion

Many electron transfer proteins were found to be expressed at the time of sampling and were assigned to *Megasphaera* in both bioreactor fractions in addition to *Pseudomonas sp* in the grass biofilms and *Clostridium sp* in the leachate (**Figure 6** and **Supplementary Table S3**). Oxidoreductases were mainly assigned to *Megasphaera elsdenii*, *Prevotella*, and Clostridiales in all the samples analyzed (**Figure 6** and **Supplementary Table S3**). Numerous ATPases (ABC-type sugar transporter and F-type) and ATP synthases were detected in the triplicate bioreactors and assigned to Gammaproteobacteria, Negativicutes, Bacteroidia and Betaproteobacteria in all samples in addition to Clostridia specifically in the leachate fractions (**Figure 6** and **Supplementary Table S3**). Proteins with functions in the TCA cycle were expressed at the time of sampling and assigned to Gammaproteobacteria in the grass biofilms in addition to Bacteroidia in the leachate samples (**Figure 6** and **Supplementary Table S3**). Enzymes involved in the degradation of propionate to pyruvate and in the conversion of pyruvate to acetyl-coA/acetate were detected in both bioreactor fractions. Acetate could then partly be used as a substrate for methanogenesis as suggested by the detection of enzymes from the corresponding metabolic pathway and assigned to *Methanosarcina* in both leachate samples and grass biofilms (**Figure 6** and **Supplementary Table S3**). Lactaldehyde dehydrogenase, involved in the production of lactate, was assigned to *Dysgonomonas gadei* and detected in the leachate only (**Supplementary Table S3**). Of particular note, proteins involved in the production and conversion of energy that were assigned to Clostridia were only detected in the leachate samples suggesting a possibly more important role for this microbial group in that bioreactor fraction compared to grass biofilms (**Figure 6** and **Supplementary Table S3**).

Lipid and Amino Acid Metabolism

Fatty acid β oxidation and pyruvate fermentation to butanol were found to take place in both leachate and grass bioreactor fractions (**Figure 6** and **Supplementary Table S3**). Butanol

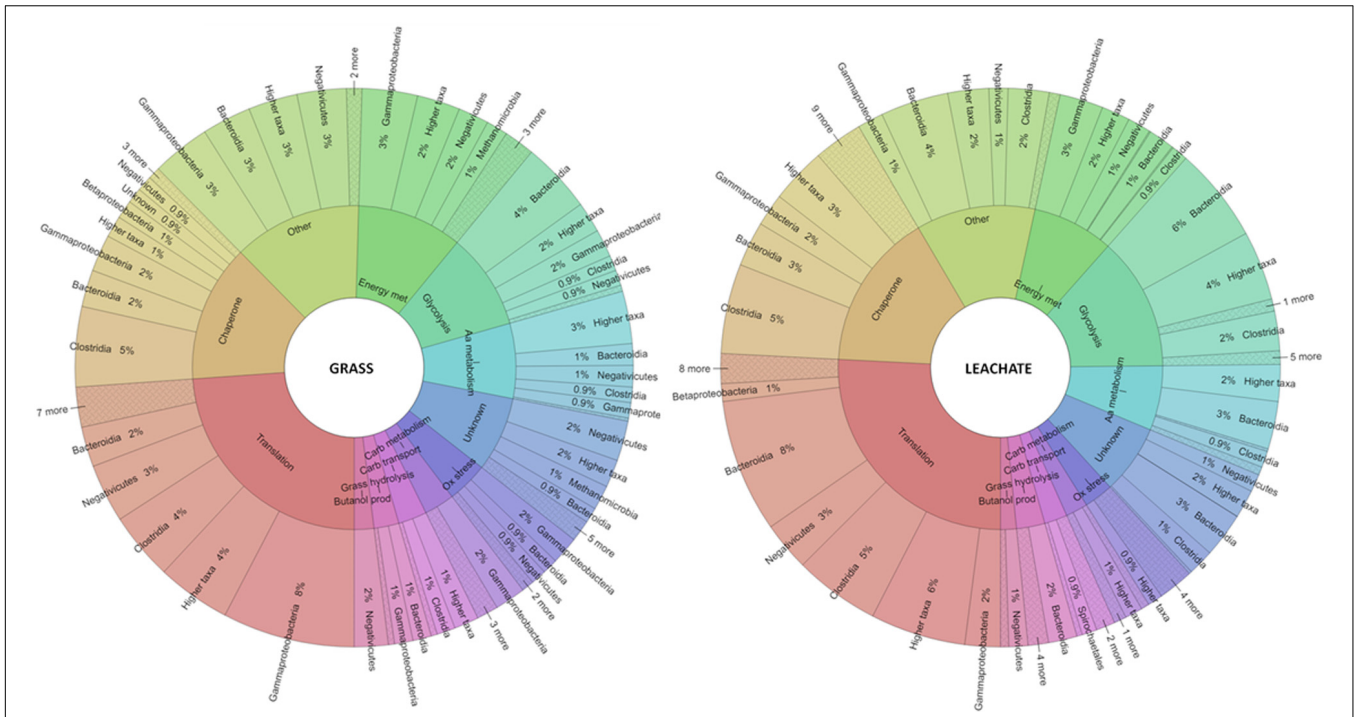


FIGURE 6 | Krona plots displaying phylogenetic classification of proteins detected at the class level on the outer ring and functional key words on the inner ring. aa, amino acid; carb, carbohydrate; met, metabolism; ox, oxidative; prod, production. For an interactive version of the Krona plot see Section "Data Accessibility."

production was exclusively attributed to *Megasphaera elsdenii* in the grass biofilms, in addition to Clostridia in leachate samples. Propionate degradation occurred in the leachate as evidenced by the detection of proteins involved in this metabolic pathway, namely methylmalonyl-CoA mutases assigned to Clostridia and propionyl-CoA carboxylases assigned to Bacteroidia (**Figure 6** and **Supplementary Table S3**). Proteins involved in lipid metabolism that were assigned to Clostridia were only detected in the leachate samples. Proteins assigned to this taxon with roles in amino acid metabolism were detected in both grass and leachate bioreactor fractions, where Clostridia were found to be involved in the fermentation of glutamate to butyrate (**Figure 6** and **Supplementary Table S3**). Bacteroidia and Negativicutes were also implicated in this process with the identification of 2-hydroxyglutaryl-CoA dehydratase and numerous glutamate dehydrogenases (**Supplementary Table S3**). Differences in the taxonomic assignment of proteins involved in amino acid transport and metabolism could be seen between the grass and leachate samples, with for example 17 and 12% of proteins assigned to Negativicutes and Gammaproteobacteria in the grass biofilms compared to 2% and none assigned to these taxa in the leachate fraction (see interactive view of the Krona plots).

Environmental Stresses

Chaperones were the second largest functional group detected in both bioreactor fractions, accounting for 14 and 16% of the proteins identified in the grass biofilms and leachate samples (**Figure 6** and interactive view of Krona plots). Chaperones are essential for protein folding but also for the re-folding of stress-denatured proteins. Numerous GroEL (60 kDa) and GroES (10 kDa) chaperonins were detected and assigned, amongst other taxa, to Clostridia and Bacteroidia (**Supplementary Table S3**). Recently, the co-expression of GroEL-GroES was found to be imperative for the production of a functional xylose isomerase in *Saccharomyces cerevisiae* (Temer et al., 2017). Xylose isomerases were detected in both bioreactor fractions and assigned exclusively to Clostridia and Bacteroidia (**Supplementary Table S3**). Trigger factor proteins, ClpB, DnaK, and heat shock protein 90 (Hsp90) were also found to be expressed in the bioreactors' grass biofilms and leachate samples (**Supplementary Table S3**). Trigger factors protect protein nascent chains from aggregation, and play important roles in the stabilization of partially folded proteins (Avellaneda et al., 2017). In addition to performing housekeeping functions, DnaK can either reverse or denature stress-induced protein aggregation (Ghazaei, 2017). Hsp90 and ClpB work in tandem with DnaK to fold or re-fold stress-denatured proteins (Schlieker et al., 2002; Nakamoto et al., 2014). Chaperones were mostly assigned to Clostridia in both grass biofilm and leachate datasets, accounting for over 30% of the proteins from that functional category (see interactive view of the Krona plots). It is worth noting that only 14 and 20% of proteins were assigned to Clostridia in the grass biofilm and leachate metaproteomes. Taken together these results suggest that members of this microbial taxon were experiencing some level of stress within the bioreactors at the time of sampling. Proteins involved in oxidative stress were detected in all the samples analyzed where they were mainly assigned to

Clostridia, Negativicutes, Bacteroidia, and Gammaproteobacteria (**Supplementary Table S3**). Desulfoferrodoxin, rubrerythrin, rubredoxin, and superoxide dismutase were detected in both bioreactor fractions in addition to alkyl hydroperoxide reductase assigned to Proteobacteria and glutathione peroxidase from *Megasphaera elsdenii* which were only detected in the leachate samples (**Supplementary Table S3**). Desulfoferrodoxin and superoxide dismutase catalyze the conversion of superoxide radicals to hydrogen peroxide, which is then reduced to water by rubrerythrin and rubredoxin with the latter involved in electron transfer during the oxidation process (Coulter and Kurtz, 2001; Staerck et al., 2017). Alkyl hydroperoxide reductase and glutathione peroxidase can also reduce a variety of hydroperoxides including hydrogen peroxide (Lu and Holmgren, 2014). These results suggest that microorganisms present in the bioreactors' leachates as well as in grass biofilms were undergoing oxidative stress at the time of sampling. This might not, however, reflect *in situ* conditions but might result from the sampling procedure and downstream analyses.

Other Functional Activities

Evidence of *Clostridium* sp. sporulation was detected in grass biofilms and leachate samples (**Supplementary Table S3**). Specifically, proteins involved in stage V of sporulation were detected in the two bioreactors' fractions. This sporulation stage is one of the latest of the process and corresponds to the spore outer coat deposition (Al-Hinai et al., 2015). Sporulation is typically triggered by unfavorable environmental conditions including nutrient depletion, accumulation of butyrate and/or butanol as well as oxidative stress (Dürre, 2014; Al-Hinai et al., 2015). Proteins involved in high affinity phosphate uptake were detected in the leachate metaproteomes and assigned exclusively to Betaproteobacteria and Gammaproteobacteria (**Supplementary Table S3**). This observation might suggest that phosphate is limiting in the leachate bioreactor fraction. Interestingly, an aminoacyl-histidine dipeptidase (PepD) was detected in one of the bioreactor leachate samples (**Supplementary Table S3**) and *pepD* has been shown to be up-regulated during phosphate starvation (Henrich et al., 1992), while its over-expression negatively impacts biofilm formation (Brombacher et al., 2003).

DISCUSSION

This study investigated microbial community structure and function during the AD of grass, under operating conditions favoring the accumulation of process intermediates. To this end a rigorous experimental strategy encompassing DNA and cDNA 16S rRNA profiling and metaproteomics was deployed on replicated bioreactors. Bioreactor fraction, i.e., grass or leachate, was found to be the main discriminator of community analysis across the three molecular levels of investigation (DNA, RNA, and proteins). Similar microbial groups and functions were detected across the two bioreactor fractions with varying abundance for the 16S rRNA datasets and with changes in phylogenetic assignments for the metaproteomes. Six main taxonomic classes, together with OTUs classified as unknown accounted for the large majority of the three datasets. An overview of the main microbial functions, occurring

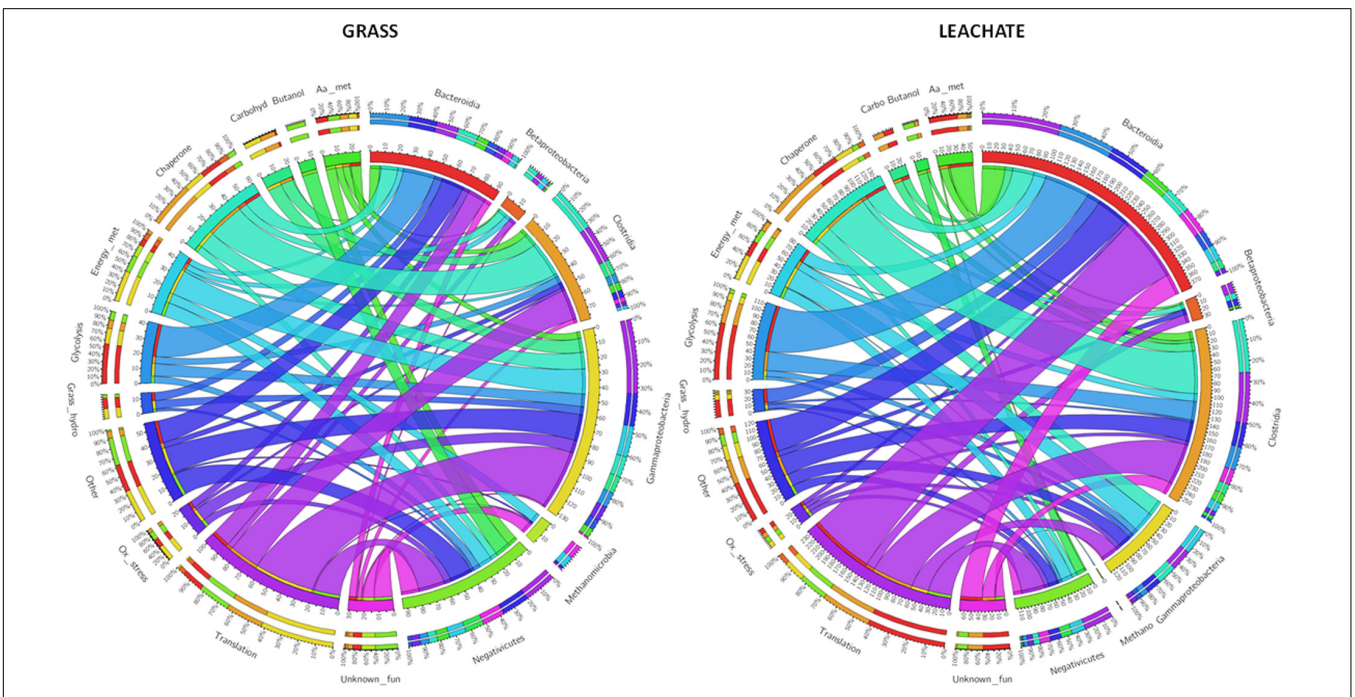


FIGURE 7 | Circos plots displaying broad functional activities and taxonomic assignments on the outer ring and proteins numbers on the inner ring. aa, amino acid; carbohyd, carbohydrate; fun, function; hydro, hydrolysis; met, metabolism; ox, oxidative.

in the bioreactors at the time of sampling, together with their corresponding phylogenetic assignments is presented in **Figure 7**. Proteins assigned to Gammaproteobacteria and Methanomicrobia represented a larger proportion of the metaproteomes from the grass biofilms when compared to leachate samples, while the reverse was observed for Bacteroidia and Clostridia (**Figure 7**). The same trends were reflected in the DNA and cDNA datasets (**Figure 3**). Taken together these results might indicate a possible preference for biofilm lifestyle as opposed to planktonic for Gammaproteobacteria and Methanomicrobia under the conditions experienced within the bioreactors. Focusing on the anaerobic process of grass acidification, using a similar experimental strategy as the one employed here, whereby acidification resulted from microbial activities, Abendroth et al. (2017) reported a dominance of Bacteroidetes (including Bacteroidia) at 37°C and of Firmicutes (including Clostridia) at 55°C in bioreactor leachates. Similar observations were reported in biogas plants using energy crops as feedstock, where Clostridia and Bacteroidales showed a higher abundance in thermophilic and mesophilic conditions, respectively (Kohrs et al., 2014). Conversely, the phylogenetic class Clostridia was proposed as a marker for biogas plants operated at mesophilic temperatures (Heyer et al., 2016). The prevalence of Clostridiales was also noted during the anaerobic conversion of office paper (mainly composed of cellulose and hemicellulose) to methane at 55°C, during which a rapid decrease from pH 7 to pH 5.8 resulting from hydrolytic and acidogenic microbial activities was reported (Lü et al., 2014). Furthermore, in agreement with the present study, the metaproteomes of the microbial communities involved in cellulose methanisation were dominated by proteins with roles in energy production and conversion (COG category C), carbohydrate transport and metabolism (G) and amino acid transport and metabolism (E; Lü et al., 2014). In addition, evidence of lactate and butanol production was suggested by the detection of enzymes involved in the corresponding metabolic pathways and assigned to *Clostridium* (Lü et al., 2014). Here, butanol production was exclusively attributed to *Megasphaera elsdenii* in the grass biofilms in addition to Clostridiales in the leachate samples (**Figure 7** and **Supplementary Table S1**). Additionally, a lactaldehyde dehydrogenase from *Dysgonomonas gadei* (Bacteroidia), indicative of lactate production, was detected in the leachate bioreactor fractions (**Supplementary Table S3**). Lactate fermentation was also reported in biogas plants with the detection of *E. coli* and Lactobacillales enzymes involved in this metabolic pathway (Kohrs et al., 2014; Heyer et al., 2016). Lactate is also likely degraded and/or involved into the production of medium chain carboxylates within the bioreactors (Zhu et al., 2015) but no evidence of such processes were obtained in the metaproteomes. Butyrate fermentation driven by Bacillales was found to take place during the AD of energy crops, while in the present study this process was mainly attributed to Clostridia, Bacteroidia, and Negativicutes (**Supplementary Table S3**). Clostridia, accounting for 14 and 20% of the grass and leachate metaproteomes, were involved in the initial stages of grass hydrolysis only in the leachate samples, while Bacteroidia, Gammaproteobacteria, and Negativicutes were implicated in this process in the grass biofilms (**Figure 7**). Overall a very similar distribution of functional activities was observed in grass biofilms and leachate samples (**Figures 5**, **7**). This, combined with differences in phylogenetic assignment distribution is indicative of functional redundancy whereby the same microbial functions are taking place throughout the replicated bioreactors but are driven by different microbial taxa. It is worth noting that sampling earlier in the bioreactor run might have led to a different conclusion. Evidence of environmental stress conditions prevailing within the bioreactors could be obtained in the metaproteomes. Clostridia seemed particularly affected as suggested by the expression of multiple chaperones, as well as proteins involved in oxidative stress response and sporulation (**Figure 7**). Even though several microbial groups of Methanomicrobia were identified in the DNA and cDNA datasets, including *Methanobacterium*, *Methanobrevibacter,* and *Methanosaeta*, only proteins from Methanosarcineae were detected in the bioreactors (**Supplementary Table S1**). This observation, together with the detection of stress response proteins, might point to the prevalence of inhospitable environmental conditions at the time of sampling. Indeed, Methanosarcineae have been shown to thrive under sub-optimal anaerobic bioreactor operating conditions (De Vrieze et al., 2012). Overall, this study emphasizes the importance of Clostridia in the AD of grass while highlighting that microbial members from this class were, at least at the time of sampling, experiencing somewhat stressful conditions. Thus, in an effort to optimize the process of grass AD, research avenues aiming at tailoring bioreactor environmental conditions to Clostridia should be explored.

DATA ACCESSIBILITY

16SrRNA sequence data were deposited on NCBI's Sequence Read Archive under the accession number SRP119456. Metaproteomic data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository (dataset identifier PXD007956). An interactive view of the Krona plots from **Figure 6** can be accessed here: https://htmlpreview.github.io/?https://github.com/FlorenceAbram/Grass-AD-study/blob/master/Krona_Metaproteomics.html.

AUTHOR CONTRIBUTIONS

FA designed the research. AV ran the bioreactors. AJ carried out the biomolecule co-extraction and prepared all samples for downstreams analyses. SS and CB performed the MS analysis. UI, AJ, FA, and CQ carried out the bioinformatic analyses. FA, AJ, and CN analyzed the data. FA and AJ wrote the paper with input from VO.

FUNDING

This research was funded by the Irish Higher Education Authority Program for Research in Third Level Institutions

Cycle 5: – PRTLI-5 ESI Ph.D. ENS Program. This work was also supported by the Wellcome Trust (grant number 094476/Z/10/Z for the TripleTOF 5600 mass spectrometer at the University of St Andrews), NERC (grant number NE/L011956/1), and a Royal Irish Academy Mobility Grant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00540/full#supplementary-material

FIGURE S1 | Overview of the metabolic pathways occurring at the time of sampling in the triplicate bioreactors (R1, R2, and R3). Red lines indicate reactions catalyzed by enzymes detected in the present study.

TABLE S1 | Summary of bioreactor R1, R2, and R3 operational performance.

TABLE S2 | Summary of number of MS/MS acquired, MS/MS assigned to peptides and number of distinct peptides for each sample.

TABLE S3 | Proteins identified in the 12 samples analyzed. Each technical replicates were combined to give a composite list for each of the triplicate bioreactor fractions: R1, R2, and R3 correspond to bioreactors 1, 2, and 3 and G and L stands for grass biofilm and leachate samples. Name, phylogenetic assignment (class and lowest common ancestor), COG category, predicted function and corresponding accession numbers are displayed for each protein.

REFERENCES

Abdul, R. S., Muller, E. L., and Wilmes, P. (2014). A hundred years of activated sludge: time for a rethink. *Front. Microbiol.* 5:47. doi: 10.3389/fmicb.2014.00047

Abendroth, C., Simeonov, C., Peretó, J., Antúnez, O., Gavidia, R., Luschnig, O., et al. (2017). From grass to gas: microbiome dynamics of grass biomass acidification under mesophilic and thermophilic temperatures. *Biotechnol. Biofuels* 10:171. doi: 10.1186/s13068-017-0859-0

Abram, F. (2015). Systems-based approaches to unravel multi-species microbial community functioning. *Comput. Struct. Biotechnol. J.* 13, 24–32. doi: 10.1016/j.csbj.2014.11.009

Abram, F., Enright, A. M., O'Reilly, J., Botting, C. H., Collins, G., and O'Flaherty, V. (2011). A metaproteomic approach gives functional insights into anaerobic digestion. *J. Appl. Microbiol.* 110, 1550–1560. doi: 10.1111/j.1365-2672.2011.05011.x

Adulkar, T. V., and Rathod, V. K. (2014). Ultrasound assisted enzymatic pretreatment of high fat content dairy wastewater. *Ultrason. Sonochem.* 21, 1083–1089. doi: 10.1016/j.ultsonch.2013.11.017

Al-Hinai, M. A., Jones, S. W., and Papoutsakis, E. T. (2015). The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiol. Mol. Biol. Rev.* 79, 19–37. doi: 10.1128/MMBR.00025-14

Amha, Y. M., Anwar, M. Z., Brower, A., Jacobsen, C. S., Stadler, L. B., Webster, T. M., et al. (2018). Inhibition of anaerobic digestion processes: applications of molecular tools. *Bioresour. Technol.* 247, 999–1014. doi: 10.1016/j.biortech.2017.08.210

American Public Health Association [APHA] (2005). *Standard Methods for the Examination of Water and Wastewater*, 21st Edn. Washington, DC: American Public Health Association.

Avellaneda, M. J., Koers, E. J., Naqvi, M. M., and Tans, S. J. (2017). The chaperone toolbox at the single-molecule level: from clamping to confining. *Protein Sci.* 26, 1291–1302. doi: 10.1002/pro.3161

Brombacher, E., Dorel, C., Zehnder, A. J., and Landini, P. (2003). The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in *Escherichia coli*. *Microbiology* 149, 2847–2857. doi: 10.1099/mic.0.26306-0

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Cerrone, F., Choudhari, S. K., Davis, R., Cysneiros, D., O'Flaherty, V., Duane, G., et al. (2014). Medium chain length polyhydroxyalkanoate (mcl-PHA) production from volatile fatty acids derived from the anaerobic digestion of grass. *Appl. Microbiol. Biotechnol.* 98, 611–620. doi: 10.1007/s00253-013-5323-x

Coulter, E. D., and Kurtz, D. M. (2001). A role for rubredoxin in oxidative stress protection in *Desulfovibrio vulgaris*: catalytic electron transfer to rubrerythrin and two-iron superoxide reductase. *Arch. Biochem. Biophys.* 394, 76–86. doi: 10.1006/abbi.2001.2531

Cysneiros, D., Banks, C. J., Heaven, S., and Karatzas, K. A. (2012). The role of phase separation and feed cycle length in leach beds coupled to methanogenic reactors for digestion of a solid substrate (Part 1): optimisation of reactors' performance. *Bioresour. Technol.* 103, 56–63. doi: 10.1016/j.biortech.2011.09.094

Cysneiros, D., Thuillier, A., Villemont, R., Littlestone, A., Mahony, T., and O'Flaherty, V. (2011). Temperature effects on the trophic stages of perennial rye grass anaerobic digestion. *Water Sci. Technol.* 64, 70–76. doi: 10.2166/wst.2011.673

De Vrieze, J., Hennebel, T., Boon, N., and Verstraete, W. (2012). *Methanosarcina*: the rediscovered methanogen for heavy duty biomethanation. *Bioresour. Technol.* 112, 1–9. doi: 10.1016/j.biortech.2012.02.079

Dürre, P. (2014). Physiology and sporulation in *Clostridium*. *Microbiol. Spectr.* 2:TBS-0010-2012.

Dzieciatkowska, M., Hill, R., and Hansen, K. C. (2014). GeLC-MS/MS analysis of complex protein mixtures. *Methods Mol. Biol.* 1156, 53–66. doi: 10.1007/978-1-4939-0685-7_4

Ellis, J. L., Dijkstra, J., France, J., Parsons, A. J., Edwards, G. R., Rasmussen, S., et al. (2012). Effect of high-sugar grasses on methane emissions simulated using a dynamic model. *J. Dairy Sci.* 95, 272–285. doi: 10.3168/jds.2011-4385

Ghazaei, C. (2017). Role and mechanism of the Hsp70 molecular chaperone machines in bacterial pathogens. *J. Med. Microbiol.* 66, 259–265. doi: 10.1099/jmm.0.000429

Gunnigle, E., Nielsen, J. L., Fuszard, M., Botting, C. H., Sheahan, J., O'Flaherty, V., et al. (2015a). Functional responses and adaptation of mesophilic microbial communities to psychrophilic anaerobic digestion. *FEMS Microbiol. Ecol.* 91:fiv132. doi: 10.1093/femsec/fiv132

Gunnigle, E., Siggins, A., Botting, C. H., Fuszard, M., O'Flaherty, V., and Abram, F. (2015b). Low-temperature anaerobic digestion is associated with differential methanogenic protein expression. *FEMS Microbiol. Lett.* 362:fnv059. doi: 10.1093/femsle/fnv059

Henrich, B., Backes, H., Klein, J. R., and Plapp, R. (1992). The promoter region of the *Escherichia coli pepD* gene: deletion analysis and control by phosphate concentration. *Mol. Gen. Genet.* 232, 117–125. doi: 10.1007/BF00299144

Hettich, R. L., Chongle, P., Chourey, K., and Giannone, R. J. (2013). Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal. Chem.* 85, 4203–4214. doi: 10.1021/ac303053e

Heyer, R., Benndorf, D., Kohrs, F., De Vrieze, J., Boon, N., Hoffmann, M., et al. (2016). Proteotyping of biogas plant microbiomes separates biogas plants according to process temperature and reactor type. *Biotechnol. Biofuels* 9:155. doi: 10.1186/s13068-016-0572-4

Koch, K., Lübken, M., Gehring, T., Wichern, M., and Horn, H. (2010). Biogas from grass silage - Measurements and modeling with ADM1. *Bioresour. Technol.* 101, 8158–8165. doi: 10.1016/j.biortech.2010.06.009

Kohrs, F., Heyer, R., Magnussen, A., Benndorf, D., Muth, T., Behne, A., et al. (2014). Sample prefractionation with liquid isoelectric focusing enables in depth microbial metaproteome analysis of mesophilic and thermophilic biogas plants. *Anaerobe* 29, 59–67. doi: 10.1016/j.anaerobe.2013.11.009

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Lackey, J. C., Peppley, B., Champagne, P., and Maier, A. (2015). Composition and uses of anaerobic digestion derived biogas form wastewater treatment facilities in North America. *Waste Manag. Res.* 33, 767–771. doi: 10.1177/0734242X15589781

Lü, F., Bize, A., Guillot, A., Monnet, V., Madigou, C., Chapleur, O., et al. (2014). Metaproteomics of cellulose methanisation under thermophilic conditions reveals a surprisingly high proteolytic activity. *ISME J.* 8, 88–102. doi: 10.1038/ismej.2013.120

Lu, J., and Holmgren, A. (2014). The thioredoxin antioxidant system. *Free Radic. Biol. Med.* 66, 75–87. doi: 10.1016/j.freeradbiomed.2013.07.036

McEniry, J., Crosson, P., Finneran, E., McGee, M., Keady, T. W. J., and O'Kiely, P. (2013). How much grassland biomass is available in Ireland in excess of livestock requirements? *Irish J. Agric. Food Res.* 52, 67–80.

Michele, P., Giuliana, D., Carlo, M., Sergio, S., and Fabrizio, A. (2015). Optimization of solid state anaerobic digestion of the OFMSW by digestate recirculation: a new approach. *Waste Manag.* 35, 111–118. doi: 10.1016/j.wasman.2014.09.009

Mustafa, N., Elbeshbishy, E., Nakhla, G., and Zhu, J. (2014). Anaerobic digestion of municipal wastewater sludges using anaerobic fluidized bed reactor. *Bioresour. Technol.* 172, 461–466. doi: 10.1016/j.biortech.2014.09.081

Nakamoto, H., Fujita, K., Ohtaki, A., Watanabe, S., Narumi, S., Maruyama, T., et al. (2014). Physical interaction between bacterial heat shock protein (Hsp) 90 and Hsp70 chaperones mediates their cooperative action to refold denatured proteins. *J. Biol. Chem.* 289, 6110–6119. doi: 10.1074/jbc.M113.524801

Narayanasamy, S., Muller, E. E., Sheik, A. R., and Wilmes, P. (2015). Integrated omics for the identification of key functionalities in biological wastewater treatment microbial communities. *Microb. Biotechnol.* 8, 363–368. doi: 10.1111/1751-7915.12255

Narihiro, T., and Sekiguchi, Y. (2007). Microbial communities in anaerobic digestion processes for waste and wastewater treatment: a microbiological update. *Curr. Opin. Biotechnol.* 18, 273–278. doi: 10.1016/j.copbio.2007.04.003

Nielfa, A., Cano, R., Pérez, A., and Fdez-Polanco, M. (2015). Co-digestion of municipal sewage sludge and solid waste modeling of carbohydrate, lipid and protein content influence. *Waste Manag. Res.* 33, 241–249. doi: 10.1177/0734242X15572181

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12:385. doi: 10.1186/1471-2105-12-385

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. Available at: https://www.R-project.org/

Ranum, P., Peña-Rosas, J. P., and Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Ann. N. Y. Acad. Sci.* 1312, 105–112. doi: 10.1111/nyas.12396

Ratanatamokul, C., and Saleart, T. (2016). Effects of sludge recirculation rate and mixing time on performance of a prototype single-stage anaerobic digester for conversion of food wastes to biogas and energy recovery. *Environ. Sci. Pollut. Res. Int.* 23, 7092–7098. doi: 10.1007/s11356-015-4448-0

Rees, G. N., Baldwin, D. S., Watson, G. O., Perryman, S., and Nielsen, D. L. (2004). Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. *Antonie Van Leeuwenhoek* 86, 339–347. doi: 10.1007/s10482-004-0498-x

Roume, R., Muller, E. E. L., Cordes, T., Renaut, J., Hiller, K., and Wilmes, P. (2012). A biomolecular isolation framework for eco-systems biology. *ISME J.* 7, 110–121. doi: 10.1038/ismej.2012.72

Schlieker, C., Bukau, B., and Mogk, A. (2002). Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: implications for their applicability in biotechnology. *J. Biotechnol.* 96, 13–21. doi: 10.1016/S0168-1656(02)00033-0

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* 68, 850–858. doi: 10.1021/ac950914h

Siggins, A., Enright, A.-M., Abram, F., Botting, C., and O'Flaherty, V. (2012a). Impact of trichloroethylene exposure on the microbial diversity and protein expression in anaerobic granular biomass at 37°C and 15°C. *Archaea* 2012:940159. doi: 10.1155/2012/940159

Siggins, A., Gunnigle, E., and Abram, F. (2012b). Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol. Ecol.* 80, 265–280. doi: 10.1111/j.1574-6941.2011.01284.x

Staerck, C., Gastebois, A., Vandeputte, P., Calenda, A., Erald Larcher, G., Gillmann, L., et al. (2017). Microbial antioxidant defense enzymes. *Microb. Pathog.* 110, 56–65. doi: 10.1016/j.micpath.2017.06.015

Standing Committee of Analysts (1985). *Chemical Oxygen Demand: Methods for the Examination of Waters and Associated Materials*, 2nd Edn. London: Her Majesty's Stationery Office, 46.

Temer, B., dos Santos, L. V., Negri, V. A., Galhardo, J. P., Magalhães, P. H. M., José, J., et al. (2017). Conversion of an inactive xylose isomerase into a functional enzyme by co-expression of GroEL-GroES chaperonins in *Saccharomyces cerevisiae*. *BMC Biotechnol.* 17:71. doi: 10.1186/s12896-017-0389-7

Theuerl, S., Kohrs, F., Benndorf, D., Maus, I., Wibberg, D., Schlüter, A., et al. (2015). Community shifts in a well-operating agricultural biogas plant: how process variations are handled by the microbiome. *Appl. Microbiol. Biotechnol.* 99, 7791–7803. doi: 10.1007/s00253-015-6627-9

United Nations (2015). *A/70/L1 Transforming Our World: The 2030 Agenda for Sustainable Development*. New York, NY: United Nations.

Vega, G. C. C., ten Hoeve, M., Birkved, M., Sommer, S. G., and Bruun, S. (2014). Choosing co-substrates to supplement biogas production from animal slurry – A life cycle assessment of environmental consequences. *Bioresour. Technol.* 171, 410–420. doi: 10.1016/j.biortech.2014.08.099

Wang, F., Hidaka, T., Uchida, T., and Tsumori, J. (2014). Thermophilic anaerobic digestion of sewage sludge with high solids content. *Water Sci. Technol.* 69, 1949–1955. doi: 10.2166/wst.2014.111

Wang, H., Qu, Y., Li, D., Zhou, X., and Feng, Y. (2015). Evaluation of an integrated continuous stirred microbial electrochemical reactor: wastewater treatment, energy recovery and microbial community. *Bioresour. Technol.* 195, 89–95. doi: 10.1016/j.biortech.2015.06.039

Werner, J. J., Knights, D., Garcia, M. L., Scalfone, N. B., Smith, S., Yarasheski, K., et al. (2011). Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4158–4163. doi: 10.1073/pnas.1015676108

Wilmes, P., Heintz-buschart, A., and Bond, P. L. (2015). A decade of metaproteomics?: where we stand and what the future holds. *Proteomics* 15, 3409–3417. doi: 10.1002/pmic.201500183

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., and Bork, P. (2011). iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* 39, W412–W415. doi: 10.1093/nar/gkr313

Ye, Y., and Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* 5:e1000465. doi: 10.1371/journal.pcbi.1000465

Yuan, X., Wen, B., Ma, X., Zhu, W., Xang, X., Cheng, S., et al. (2014). Enhancing the anaerobic digestion of lignocellulose of municipal solid waste using microbial pre-treatment method. *Bioresour. Technol.* 154, 1–9. doi: 10.1016/j.biortech.2013.11.090

Zerrouki, S., Rihani, R., Bentahar, F., and Belkacemi, K. (2015). Anaerobic digestion of wastewater from fruit juice industry: experiments and modeling. *Water Sci. Technol.* 72, 123–134. doi: 10.2166/wst.2015.193

Zhu, X., Tao, Y., Liang, C., Li, X., Wei, N., Zhang, W., et al. (2015). The synthesis of n-caproate from lactate: a new efficient process for medium-chain carboxylates production. *Sci. Rep.* 5:14360. doi: 10.1038/srep14360

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Joyce, Ijaz, Nzeteu, Vaughan, Shirran, Botting, Quince, O'Flaherty and Abram. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 March 2018
doi: 10.3389/fmicb.2018.00405

Metabolic Adaptation of Methanogens in Anaerobic Digesters Upon Trace Element Limitation

*Babett Wintsche[1], Nico Jehmlich[2], Denny Popp[1], Hauke Harms[1] and Sabine Kleinsteuber[1]**

[1] Department of Environmental Microbiology, Helmholtz Centre for Environmental Research—Helmholtz-Zentrum für Umweltforschung (UFZ), Leipzig, Germany, [2] Department of Molecular Systems Biology, Helmholtz Centre for Environmental Research—Helmholtz-Zentrum für Umweltforschung (UFZ), Leipzig, Germany

Anaerobic digestion (AD) is a complex multi-stage process relying on the activity of highly diverse microbial communities including hydrolytic, acidogenic and syntrophic acetogenic bacteria as well as methanogenic archaea. The lower diversity of methanogenic archaea compared to the bacterial groups involved in AD and the corresponding lack of functional redundancy cause a stronger susceptibility of methanogenesis to unfavorable process conditions such as trace element (TE) deprivation, thus controlling the stability of the overall process. Here, we investigated the effects of a slowly increasing TE deficit on the methanogenic community function in a semi-continuous biogas process. The aim of the study was to understand how methanogens in digester communities cope with TE limitation and sustain their growth and metabolic activity. Two lab-scale biogas reactors fed with distillers grains and supplemented with TEs were operated in parallel for 76 weeks before one of the reactors was subjected to TE deprivation, leading to a decline of cobalt and molybdenum concentrations from 0.9 to 0.2 mg/L, nickel concentrations from 2.9 to 0.8 mg/L, manganese concentrations from 38 to 18 mg/L, and tungsten concentrations from 1.4 to 0.2 mg/L. Amplicon sequencing of *mcrA* genes revealed *Methanosarcina* (72%) and *Methanoculleus* (23%) as the predominant methanogens in the undisturbed reactors. With increasing TE limitation, the relative abundance of *Methanosarcina* dropped to 67% and a slight decrease of acetoclastic methanogenic activity was observed in batch tests with ^{13}C-methyl-labeled acetate, suggesting a shift toward syntrophic acetate oxidation coupled to hydrogenotrophic methanogenesis. Metaproteome analysis revealed abundance shifts of the enzymes involved in methanogenic pathways. Proteins involved in methylotrophic and acetoclastic methanogenesis decreased in abundance while formylmethanofuran dehydrogenase from *Methanosarcinaceae* increased, confirming our hypothesis of a shift from acetoclastic to hydrogenotrophic methanogenesis by *Methanosarcina*. Both *Methanosarcina* and *Methanoculleus* increased the abundance of N5-methyltetrahydromethanopterin-coenzyme M methyltransferase and methyl-coenzyme M reductase. However, these efforts to preserve the ion motive force for energy conservation were seemingly more successful in *Methanoculleus*.

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Tim Magnuson, Idaho State University, United States
Stefan Junne, Technische Universität Berlin, Germany

***Correspondence:**
Sabine Kleinsteuber
sabine.kleinsteuber@ufz.de

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *30 November 2017*
Accepted: *21 February 2018*
Published: *13 March 2018*

Citation:
Wintsche B, Jehmlich N, Popp D, Harms H and Kleinsteuber S (2018) Metabolic Adaptation of Methanogens in Anaerobic Digesters Upon Trace Element Limitation. Front. Microbiol. 9:405. doi: 10.3389/fmicb.2018.00405

We conclude that both methanogenic genera use different strategies to stabilize their energy balance under TE limitation. *Methanosarcina* switched from TE expensive pathways (methylotrophic and acetoclastic methanogenesis) to hydrogenotrophic methanogenesis. *Methanoculleus* showed a higher robustness and was favored over the more fastidious *Methanosarcina*, thus stabilizing reactor performance under TE limitation.

Keywords: methanogenic pathways, metaproteome, trace metals, biogas process, *mcrA*, *Methanosarcina*, *Methanoculleus*

INTRODUCTION

Anaerobic digestion (AD) is a widespread and effective way for recycling organic waste and biomass residues while producing biogas as renewable energy carrier. Biogas production is a scalable, technically simple and low-cost technology and has therefore a huge potential for renewable energy supply in developing countries (Surendra et al., 2014). In a renewable energy system, it can contribute to all energy sectors (electricity, heating, and mobility) and complement fluctuating renewable energy sources such as wind and solar power (Patterson et al., 2011; Raboni et al., 2015).

AD is a complex multi-stage process relying on the activity of highly diverse microbial communities (Weiland, 2010). The process can be divided in four main phases: hydrolysis, acidogenesis, acetogenesis and methanogenesis. Methanogenesis is exclusively performed by distinct groups of archaea. Seven phylogenetic orders of methanogens (all belonging to the phylum *Euryarchaeota*) have been described so far (*Methanobacteriales*, *Methanococcales*, *Methanomassiliicoccales*, *Methanomicrobiales*, *Methanosarcinales*, *Methanocellales,* and *Methanopyrales*) and all but *Methanopyrales* are ascertainable in biogas processes (Borrel et al., 2013).

Methane can be produced via three different pathways. Hydrogenotrophic methanogens produce methane from carbon dioxide and hydrogen or formate. This pathway is performed by the cultivated methanogens of the orders *Methanobacteriales, Methanococcales, Methanomicrobiales, Methanocellales,* and *Methanopyrales* as well as some members of the *Methanosarcinales*. Methylotrophic methanogens can grow on methylated compounds like methanol or methylamines by dismutation (Whitman et al., 2006). Acetate is directly dismutated to methane and carbon dioxide by acetoclastic methanogens. Cultivated acetoclastic and methylotrophic methanogens are all members of the order *Methanosarcinales*. The recently described genus *Methanomassiliicoccus* belonging to the order *Methanomassiliicoccales* (class *Thermoplasmata*) is an exception (Borrel et al., 2013). It is capable of reducing methanol with hydrogen (Dridi et al., 2012) and might also use methylamines as methanogenic substrate (Poulsen et al., 2013). The reduction of methanol to methane with hydrogen was also described for *Methanosphaera stadtmanae,* which belongs to the *Methanobacteriales* (Miller and Wolin, 1985). Recent findings from metagenome analyses suggest that the actual metabolic and phylogenetic diversity of methanogens might be much higher and comprise a new class of *Euryarchaeota* ("*Methanofastidiosa*"—Nobu et al., 2016) or even other archaeal phyla ("*Bathyarchaeota*"—Evans et al., 2015; "*Verstraetearchaeota*"—Vanwonterghem et al., 2016).

Compared to the bacterial groups involved in AD, the lower diversity and the lack of functional redundancy among methanogenic archaea causes the susceptibility of methanogenesis to unfavorable process conditions such as trace element (TE) deprivation, thus determining the stability of the whole process (Demirel, 2014). The need for TE and the effects of TE limitation on methanogens and reactor performance have been addressed by various studies and reviews (Park et al., 2010; Demirel and Scherer, 2011; Choong et al., 2016). Cobalt, molybdenum, nickel, selenium and tungsten next to iron are known as essential TE for methanogens as shown by studies on their elemental composition (Scherer et al., 1983), their metallo-enzymes (Glass and Orphan, 2012; Choong et al., 2016) and the effect of stimulation by TE (Takashima et al., 1990). For instance, nickel is one of the most important TE for methanogens (Diekert et al., 1981) and was shown to enhance acetate utilization rates (Speece et al., 1983) and increase methane yields in maize silage-fed batch reactors by about 27% (Evranos and Demirel, 2015). Changes of AD reactor performance due to changing TE supplementation are mainly explained on the basis of the methanogenesis step (Park et al., 2010; Demirel and Scherer, 2011; Ariunbaatar et al., 2016; Choong et al., 2016).

Further studies are required to understand how methanogens react to TE deprivation specifically by adapting their metabolism and energy balance especially under limiting conditions. Here, we investigated the effects of a slowly increasing TE deficit on the methanogenic community function in a semi-continuous AD process. After parallel operation of two lab-scale reactors that were well supplied with TE, the TE supplementation of one reactor was stopped, resulting in a decline of TE concentrations to insufficient levels. As shown in our previous study (Wintsche et al., 2016), the slowly decreasing TE supply did not affect reactor efficiency, although shifts of the methanogenic community composition and presumably shifts in the methanogenic pathways were indicated by community fingerprinting of metabolic marker genes and their transcripts. The aim of the present study was to use metaproteomics and metabolite analyses with ^{13}C-labeled tracers to understand in more detail how methanogens cope with TE limitation and sustain their growth and metabolic activity leading to AD reactor stability.

MATERIALS AND METHODS

Laboratory-Scale Biogas Reactors and Sampling

Two identical continuous stirred tank reactors (working volume: 10 L) designated R1 and R2 were operated under mesophilic conditions for 93 weeks as described by Wintsche et al. (2016). The feedstock was dried distillers grains with solubles and the reactors were supplemented with a commercial iron additive and a TE mixture containing cobalt, nickel, molybdenum and tungsten as described by Schmidt et al. (2013). The reactors were operated at an organic loading rate of 5 g_{VS} L^{-1} d^{-1} (VS – volatile solids) resulting in a hydraulic retention time of 25 d. Both reactors were operated in parallel for 76 weeks before starting the experimental period in which the TE supply to R2 was altered by omitting the TE solution and reducing the supply of the iron additive from 2.57 to 0.86 g per day. This altered feeding scheme led to a decline of cobalt and molybdenum concentrations from around 0.9 to 0.2 mg/L, nickel concentrations from 2.9 to 0.8 mg/L, manganese concentrations from 38 to 18 mg/L, and tungsten concentrations from 1.4 to 0.2 mg/L from week 65 to 84. For a detailed description of the reactor setup, operational conditions and detailed measurements and modeling of TE depletion, see Wintsche et al. (2016).

Samples for batch experiments with ^{13}C-labeled acetate and proteome analysis were taken at four sampling times (week 65, 77, 80, and 84). Samples for DNA extraction were taken in week 74, 77, 80 and 84. The first sample was taken before the TE supplementation was stopped to ensure comparability for both undisturbed reactors. The next samples were taken one, 4 and 8 weeks after omitting the TE supply of R2.

Methanogenic Community Analysis

The methanogenic communities of both reactors at the four sampling times were analyzed by amplicon sequencing of *mcr*A genes. Reactor samples were stored at −20°C until DNA extraction. DNA was extracted with PowerSoil DNA Isolation Kit (MoBio Laboratories Inc., USA) according to the manufacturers' instructions. PCR amplification of *mcrA* genes was performed as described previously (Steinberg and Regan, 2008). Amplicons were sequenced using the 454 pyrosequencing platform GS Junior (Roche) according to Ziganshin et al. (2013). Raw sequences were analyzed with QIIME 1.9.1 Virtual Box release (Caporaso et al., 2010) as described by Popp et al. (2017). Briefly, sequences were quality filtered and chimeric sequences were removed. Sequences were clustered into operational taxonomic units based on 97% sequence identity and were taxonomically classified against a custom database compiled of *mcrA* sequences deposited in the Functional Gene Repository (Fish et al., 2013) using the RDP Classifier 2.2 (Wang et al., 2007). De-multiplexed raw sequences were deposited under the EMBL-EBI study accession number PRJEB21972 (http://www.ebi.ac.uk/ena/data/view/PRJEB21972).

Batch Experiments With ^{13}C-Labeled Acetate

Labeling experiments at four sampling times (I – week 65, II – week 77, III – week 80, IV – week 84) were done by transferring 1.7 L sludge from each reactor into 2-L Duran bottles purged with nitrogen. The bottles were closed, the headspace purged with biogas (61% CH_4, 39% CO_2, 50 ppm H_2S, 50 ppm H_2, 50 ppm O_2) and connected to a gas sampling bag. The bottles were incubated for 3 days at 37°C without feeding to reduce the high organic carbon pool within the samples. Bottles were swiveled daily.

^{13}C-labeled acetate (0.5 M) was applied as sodium salt. Carboxyl-labeled acetate (Sigma-Aldrich, isotopic purity 99 atom % ^{13}C) and methyl-labeled acetate (Sigma-Aldrich, isotopic purity 99 atom % ^{13}C) were fed in separate batch cultures. All solutions were prepared with sterile anoxic distilled water in glass vials. The closed vials were purged with nitrogen. Five 50-mL serum bottles for each labeled substrate and each reactor sample were prepared. All bottles were filled with 25 mL reactor sludge and closed airtight in an anaerobic chamber (97% N_2 and 3% H_2 atmosphere); then the headspaces were purged with biogas (composition as described above) outside the anaerobic chamber. The batch cultures were fed with 500 μL of ^{13}C-acetate solution via a syringe. Immediately after feeding and then every 2 h, one bottle per substrate was processed for gas and proteome analyses. The produced gas was released via a cannula and the volume measured by a U-tube manometer as described by Porsch et al. (2015). Gas composition was determined in triplicates by gas chromatography according to Sträuber et al. (2015). Analyses and calculation of labeled gas ratios (^{13}C-CO_2 to ^{12}C-CO_2 and ^{13}C-CH_4 to ^{12}C-CH_4) were done by gas chromatography mass spectrometry (MS) according to Popp et al. (2016). For proteome analysis, 500 μL of the sludge were centrifuged at maximum speed and the supernatant was discarded. The pellet was stored at −20°C until protein extraction.

Protein Extraction and Preparation

The methanogenic communities of both reactors at the four sampling weeks were analyzed using metaproteomics. For reactor R2, 10 batch cultures per sampling week were sampled for protein extraction. For the control reactor R1, 10 batch cultures in week 65 and three batch cultures each in week 77, 80, and 84 were analyzed for their metaproteome (see Supplementary Material, Data Sheet 3). To each sample pellet, 5 mL sodium dodecyl sulfate (SDS) buffer (1.25% w/v SDS, 0.1 M Tris/HCl pH 6.8, 20 mM dithiotreitol) was added and incubated for 1 h at room temperature. Afterwards, samples were centrifuged (30 min at 10,000 × g and 4°C) and the supernatant was collected and filtered through a nylon mesh with a pore size of 0.45 mm. The filtrate was mixed with the equal volume of phenol solution (10 g/mL) and incubated at room temperature for 15 min. Samples were centrifuged and the phenol phase was collected. The water phase was again mixed with the equal volume of phenol solution, incubated 15 min at room temperature with shaking and then centrifuged (12 min at 10,000 × g and 4°C). Both phenol

phases were pooled and washed twice with the equal volume of Millipore water for 15 min. After centrifugation (12 min at 10,000 × g and 4°C), the water phase was discarded and the proteins in the phenol phase were precipitated over night at −20°C with ice-cold ammonium acetate (100 mM ammonium acetate in methanol, five-fold, stored at −20°C). Protein pellets were obtained by centrifugation (12 min at 10,000 × g and 4°C). Protein pellets were resuspended in 20 µL SDS sample buffer (2% w/v SDS, 2 mM β-mercaptoethanol, 4% v/v glycerol, 40 mM Tris/HCl pH 6.8, 0.01% w/v bromophenol blue), heated at 90°C for 4 min and separated for 10 min by electrophoresis in a 12% SDS polyacrylamide gel (4% stacking gel, 12% separating gel). After electrophoresis, the gels were stained with colloidal Coomassie brilliant blue (Merck). The gel area containing the protein mixture of each sample was cut out in one piece, destained, dehydrated and proteolytically cleaved overnight at 37°C by trypsin (Promega). Extracted peptides were desalted using C18 ZipTip columns (Merck Millipore). Peptide lysates were dissolved in 0.1% formic acid and analyzed by liquid chromatography MS.

Mass Spectrometry-Based Proteome Analyses

The peptide lysates were separated on a UHPLC system (Ultimate 3000, Dionex/Thermo Fisher Scientific, Idstein, Germany). Five microliter samples were first loaded for 5 min on the pre-column (µ-pre-column, Acclaim PepMap, 75 µm inner diameter, 2 cm, C18, Thermo Scientific) at 4% mobile phase B (80% acetonitrile in Nanopure water with 0.08% formic acid), 96% mobile phase A (Nanopure water with 0.1% formic acid), then eluted from the analytical column (PepMap Acclaim C18 LC Column, 25 cm, 3 µm particle size, Thermo Scientific) over a 150 min non-linear gradient of mobile phase B (4–55% B).

MS was performed on an Orbitrap Fusion MS (Thermo Fisher Scientific, Waltham, MA, USA) with a TriVersa NanoMate (Advion, Ltd., Harlow, UK) source in LC chip coupling mode. The MS was set at cycle time of 3 s used for MS/MS scans with higher energy collision dissociation (HCD) at normalized collision energy of 28%. MS scans were measured at a resolution of 120,000 in the scan range of 350–2,000 m/z. MS ion count target was set to 4×10^5 at an injection time of 100 ms. Ions for MS/MS scans were isolated in the quadrupole with an isolation window of 1.6 Da and were measured with a resolution of 15,000 in the scan range of 350–1,400 m/z. The dynamic exclusion duration was set to 30 s with a 10 ppm tolerance. Automatic gain control target was set to 6×10^4 with an injection time of 150 ms using the underfill ratio of 1%.

Bioinformatics Analysis

Protein identification was performed using the Proteome Discoverer (v1.4.0.288, Thermo Scientific). The acquired MS/MS spectra (*.raw files) were searched using the Sequest HT algorithm against the database provided by Kohrs et al. (2015) extended with Uniprot entries for methanogens and several syntrophic bacteria. Search parameters were set as follows: tryptic cleavage, maximum of two missed cleavage sites, a precursor mass tolerance threshold of 10 ppm and a fragment mass tolerance threshold of 0.02 Da. In addition, carbamidomethylation at cysteine was selected as a static and oxidation of methionine as a variable modification. Only peptides that passed the false discovery rate (FDR) of <1% and peptide rank =1 were considered for protein identification. Label-free quantification was done using peptide spectral matching (PSM). The PROteomics results Pruning & Homology group ANotation Engine (PROPHANE) was used to calculate protein abundances based on the normalized spectral abundance factor (NSAF; von Bergen et al., 2013) and to assign proteins to their taxonomic and functional groups (www.prophane.de). Taxonomic assignment was done by BLASTp v2.2.28+ (*E*-value: ≤0.001). Functional classification was based on TIGRFAM, Pfam-A and cluster of orthologous groups (COG) (*E*-value: ≤0.01).

Data analysis was focused on methanogenesis enzymes of the families *Methanomicrobiaceae* and *Methanosarcinaceae*; any bacterial or other archaeal hits were excluded from further analyses. Transformation, normalization and statistical analysis of protein group intensity data were performed using R (v 2.15.02) and "ggplot2" (v 0.9.3.1) (Wickham, 2009).

RESULTS

Community Composition and Dynamics of Methanogens

The community composition of methanogens was examined by amplicon sequencing of the *mcr*A gene. Number of sequence reads, operational taxonomic units and rarefaction curves are shown in the Supplementary Material (Data Sheet 2, Figure S1). *Methanosarcina* spp. and *Methanoculleus* spp. dominated in both reactors while negligible abundances of *Methanospirillum, Methanobacterium* and other methanogens not classified to the genus level were detected (**Figure 1**). The methanogenic community in the undisturbed reactor R1 underwent minor fluctuations over all sampling times (*Methanosarcina* 71–75%; *Methanoculleus* 23-26%). Reactor R2 showed a comparable community composition in week 74 during sufficient TE supply (*Methanosarcina* 77%; *Methanoculleus* 21%), while the relative abundance of *Methanoculleus* dropped to 13% after initiating TE deprivation in week 77, then recovered until week 80 to 24% and increased further to 33% in week 84. The relative abundance of *Methanosarcina* in R2 behaved inversely, suggesting an adaptation of the methanogenic community to the incremental TE depletion.

Active Methanogenic Pathways

Methanogenic pathways were analyzed by feeding ^{13}C-methyl-labeled acetate to batch cultures set up using reactor content, and recording the formation of ^{13}C-labeled methane. The amount of ^{13}C-labeled methane formed from the ^{13}C-methyl-labeled acetate fed to each batch culture (0.25 mmol) was calculated based on the measured methane volume and the ratio of $^{13}C\text{-}CH_4$ to $^{12}C\text{-}CH_4$ as determined by GC-MS. In the first two sampling times (weeks 65 and 77) the batch cultures set up from both reactors produced similar amounts of ^{13}C-labeled methane within 8 h after feeding methyl-labeled acetate. In sampling week 80 samples from reactor R2 produced slightly more labeled

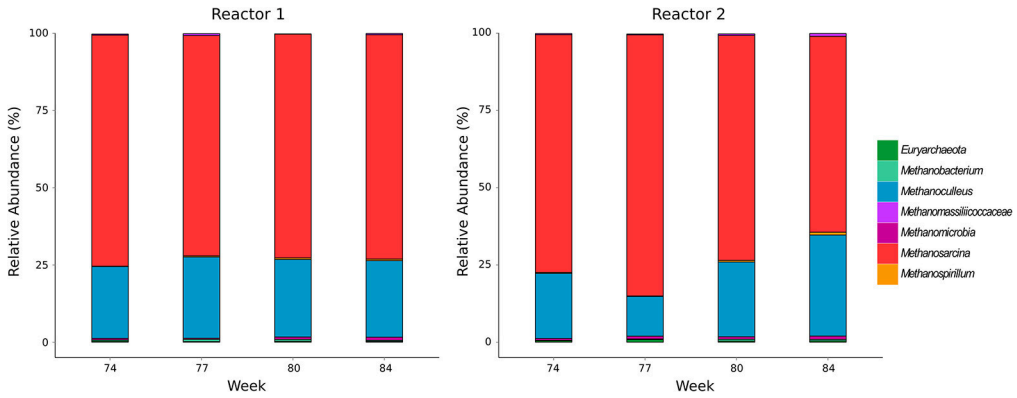


FIGURE 1 | Community composition of methanogens determined by amplicon sequencing of *mcr*A genes.

methane than those from reactor R1, whereas in week 84 a slight drop of the labeled methane was observed in reactor R2 compared to reactor R1 (Figure S2). This data indicates a partial metabolic shift of the active methanogenic pathways from acetoclastic methanogenesis toward syntrophic acetate oxidation (SAO) coupled to hydrogenotrophic methanogenesis.

The metabolic shift was analyzed in detail by examining the enzyme abundances of the different methanogenic pathways. Results of the proteome analysis are provided in the Supplementary Material (Data Sheet 3). **Table 1** lists all detected enzymes, their reactions and enzyme classification. Similar to the community composition as detected by *mcrA* amplicon sequencing, the enzyme abundances in the control reactor R1 underwent minor fluctuations over the four sampling times (Figure S3). In contrast, reactor R2 showed remarkable trends linked to TE deprivation as illustrated in **Figure 2**. In the beginning (week 77) the declining TE concentrations caused lower abundances of several enzymes involved in hydrogenotrophic methanogenesis of the *Methanomicrobiaceae*, such as methenyl-H_4MPT cyclohydrolase (Mhc), methylene-H_4MPT reductase (Mer) and methyl-CoM reductase (Mcr). However, Mcr abundance increased again in week 80. Other enzymes of the *Methanomicrobiaceae* became more abundant in week 77 and decreased later in abundance, such as formylmethanofuran dehydrogenase (Fmd) and methylene-H_4MPT dehydrogenase (Mtd). Abundance of the cobalt-dependent formylmethanofuran:H_4MPT formyltransferase (Ftr) decreased as well. Only two enzymes of *Methanomicrobiaceae* were more abundant at lower TE concentrations over all sampling times: coenzyme F_{420}-reducing hydrogenase (Frh) and methyl-H_4MPT:CoM methyltransferase (Mtr). For the *Methanosarcinaceae*, abundance shifts were more pronounced. Abundances of cobalt-dependent enzymes involved in methylotrophic methanogenesis declined as well as acetate kinase (Ack) (involved in acetoclastic methanogenesis) and [NiFe] hydrogenase. Surprisingly, the nickel-dependent acetyl-CoA decarboxylase/synthase complex (ACDS) slightly increased in its abundance with declining TE concentrations. Only formylmethanofuran dehydrogenase (Fmd) and methyl-H_4MPT:CoM methyltransferase (Mtr) showed increasing abundances, the latter as observed for the *Methanomicrobiaceae*.

For a better overview, **Figure 3** depicts the hydrogenotrophic pathways of both methanogens in a metabolic scheme highlighting the abundance shifts of involved enzymes at the last sampling time (week 84) for R2. A corresponding scheme for R1 is presented in Figure S4 (Supplementary Material).

DISCUSSION

Our previous study on the effects of TE deprivation in AD suggested that limiting concentrations of Co, Mn, Mo, Ni and W cause activity shifts within the methanogenic communities as detected by T-RFLP profiling of *mcrA* transcripts (Wintsche et al., 2016). The major genera affected in the lab-scale biogas reactors investigated were *Methanosarcina* and *Methanoculleus*. The present study confirms these two genera as dominant methanogens by amplicon sequencing of *mcrA* genes, which is a more precise method than T-RFLP analysis. The TE decline in reactor R2 caused only minor shifts in the methanogenic community composition. This observation confirms our previous results that the effects of TE deprivation were more pronounced on the RNA level than on the DNA level (Wintsche et al., 2016).

Methanosarcina is a versatile, multipotent methanogen able to degrade diverse substrates via acetoclastic, methylotrophic or hydrogenotrophic methanogenesis (Conklin et al., 2006; De Vrieze et al., 2012). *Methanoculleus* is a hydrogenotrophic methanogen that can also act as syntrophic partner of syntrophic acetate oxidizing bacteria (SAOB). Based on the T-RFLP patterns of *mcrA* transcripts, we hypothesized in Wintsche et al. (2016) a relative activity increase of *Methanoculleus* over *Methanosarcina* and a shift from acetoclastic to hydrogenotrophic methanogenesis in *Methanosarcina* as a consequence of TE deprivation. This hypothesis is supported by the results of the tracer experiment with ^{13}C-labeled acetate in the present study. The central AD intermediate acetate is degraded by acetoclastic methanogens or by SAOB. Degradation

TABLE 1 | Reactions of the hydrogenotrophic, acetoclastic and methylotrophic methanogenesis and all involved enzymes detected in the proteome analysis.

Reaction	Enzyme	Abbr.	Equation	Enzyme class	
1	Formylmethanofuran dehydrogenase	Fmd	$CO_2 + methanofuran + reduced\ acceptor \rightleftharpoons formylmethanofuran + H_2O + oxidized\ acceptor$	1.2.99.5	Hydrogenotrophic
2	Formylmethanofuran:H_4SPT formyltransferase	Ftr	$formylmethanofuran + H_4MPT + H^+ \rightleftharpoons 5 - formyl - H_4MPT + methanofuran$	2.3.1.101	
3	Methenyl-H_4SPT cyclohydrolase	Mch	$5 - formyl - H_4MPT + H^+ \rightleftharpoons 5,10 - methenyl - H_4MPT + H_2O$	3.5.4.27	
4	F_{420}-dependent methylene-H_4SPT dehydrogenase	Mtd	a) $5,10 - methenyl - H_4MPT + red.\ F_{420} + H^+ + \rightleftharpoons 5,10 - methylene - H_4MPT + ox.\ F_{420}$ b) $5,10 - methenyl - H_4MPT + H_2 + \rightleftharpoons 5,10 - methylene - H_4MPT + H^+$	1.5.98.1/ 1.12.98.2	
5	F_{420}-dependent methylene-H_4SPT reductase	Mer	$5,10 - methylene - H_4MPT + reduced\ coenzyme\ F_{420} \rightleftharpoons 5 - methyl - H_4MPT + coenzyme\ F_{420}$	1.5.98.2	
6	Methyl-H_4MPT:coenzyme M methyltransferase	Mtr	a) $5 - methyl - H_4MPT + coenzyme\ M \rightleftharpoons methyl - coenzyme\ M + H_4MPT$ b) $5 - methyl - H_4MPT + cob\ (I)\ alamine \rightleftharpoons methyl - cob\ (III)\ alamine + H_4MPT$ c) $methyl - cob\ (III)\ alamine + coenzyme\ M \rightleftharpoons methyl - coenzyme\ M + cob\ (I)\ alamine$	2.1.1.86	
7	Factor F_{420}-reducing hydrogenase	Frh	$F_{420}\ (oxidized) + H_2 \rightleftharpoons F_{420}H_2$	1.12.98.1	
8	Acetate kinase	Ack	$ATP + acetate \rightleftharpoons ADP + acetyl\ phosphate$	2.7.2.1	Acetoclastic
9	Acetyl-CoA decarbonylase/synthase complex	ACDS	$acetyl - CoA + H_4SPT + H_2O \rightleftharpoons 5 - methyl - H_4SPT + CO_2 + H^+$	2.1.1.-	
10	Trimethylamine methyltransferase	MttB	$trimethylamine + MttC + H^+ \rightleftharpoons methyl - MttC + dimethylamine$	2.1.1.250	Methylotrophic
11	Dimethylamine methyltransferase	MtbB1	$dimethylamine + MtbC + H^+ \rightleftharpoons methyl - MtbC + monomethylamine$	2.1.1.249	
12	Methylamine methyltransferase	MtbA	$methyl - Mt\ (t,b,m)\ C + HS - CoM \rightleftharpoons methyl - S - CoM + Mt\ (t,b,m)\ C$	2.1.1.247	
13	Methyl-CoM reductase	Mcr	$methyl - CoM + CoB \rightleftharpoons CoM - S - S - CoB + CH_4$	2.8.4.1	x
14	Heterodisulfide reductase	Hdr	$CoM - S - S - CoB \rightleftharpoons CoM - SH + CoB - SH$	1.8.98.1	

Enzymes, their abbreviations, reaction equations and the corresponding enzyme class are given. x, common to all methanogenic pathways; H_4MPT, tetrahydromethanopterin; H_4SPT, tetrahydrosarcinapterin.

of ^{13}C-methyl-labeled acetate via acetoclastic methanogenesis maintains the label in the methyl moiety, leading to the formation of ^{13}C-methane, while SAOB convert acetate completely to CO_2. Consequently, a shift from acetoclastic methanogenesis to SAO would be reflected in a decline of the methane labeling ratio. The conditions in our control reactor R1 and the undisturbed reactor R2 during full TE supplementation (low concentrations of H_2S, VFA and low total ammonium nitrogen) were characteristic of AD processes dominated by acetoclastic methanogenesis (Karakashev et al., 2005, 2006; Wintsche et al., 2016). With ongoing TE depletion in reactor R2, the amount of $^{13}CH_4$ decreased, indicating a partial shift from acetoclastic methanogenesis to SAO coupled to hydrogenotrophic methanogenesis. In accordance with this observation, the arising reactor conditions in R2 (increasing concentrations of H_2S, H_2, VFA concentrations and total ammonium nitrogen – Wintsche et al., 2016) are known to favor hydrogenotrophic over acetoclastic methanogenesis (Schnürer et al., 1994; Karakashev et al., 2005, 2006).

To analyze in more detail at the metabolic level how the major methanogens *Methanosarcina* and *Methanoculleus* cope with TE deprivation, protein abundances of the enzymes involved in methanogenic pathways were examined via proteome analysis of the dominant families *Methanosarcinaceae* and *Methanomicrobiaceae*. Abundances of proteins involved in methylotrophic methanogenesis by *Methanosarcinaceae* decreased. The degradation of methylated compounds proceeds via a very specific pathway including individual methyltransferases and corrinoid-binding proteins. These are specific for their respective substrates and exhibit little or no activity with other methylotrophic substrates (van der Meijden et al., 1983; Burke and Krzycki, 1997; Ferguson and Krzycki, 1997; Wassenaar et al., 1998). The respective enzymes MttB, MtbB1, and MtbA (**Table 1**) depend on cobalt

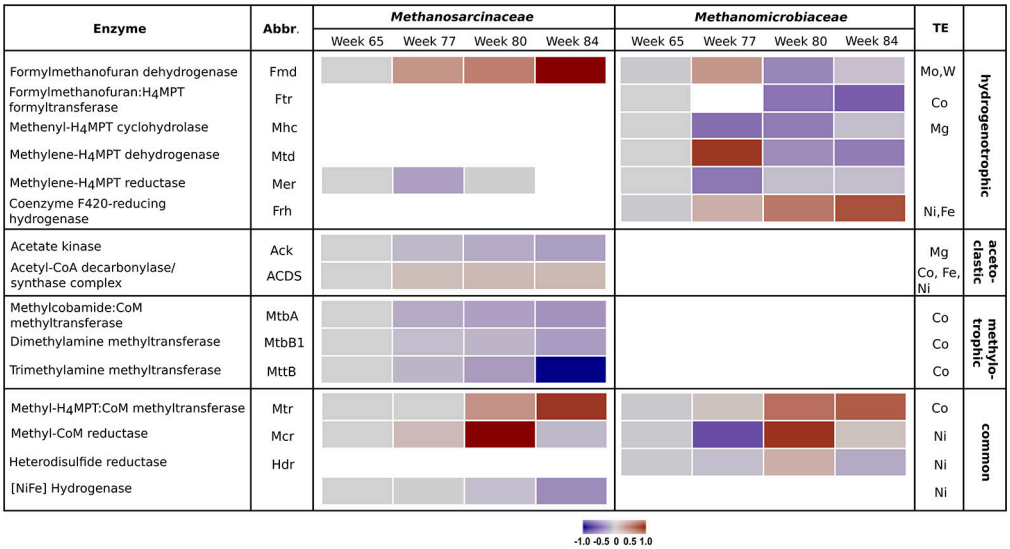


FIGURE 2 | Heatmap of enzyme abundances of the methanogenic pathways employed by *Methanosarcinaceae* and *Methanomicrobiaceae* over the four sampling times in reactor R2. For each enzyme, the specific methanogenic pathway and the required trace elements are given. Gray bars indicate initial conditions, blue bars declining protein abundances, and red bars increasing protein abundances. Missing bars indicate enzymes that were not detected.

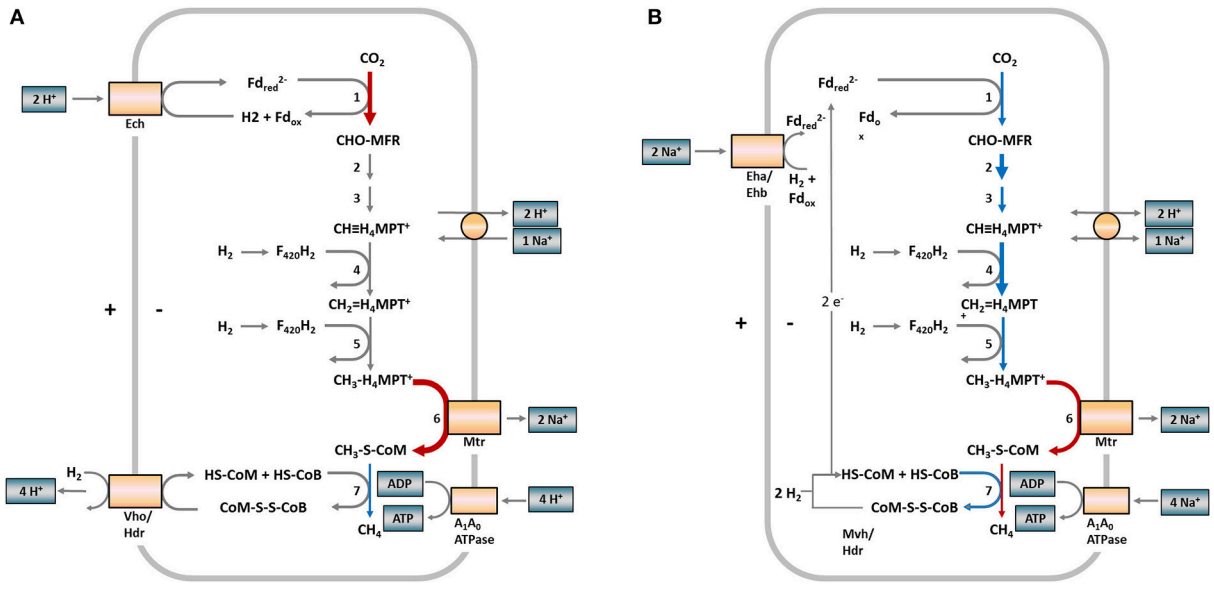


FIGURE 3 | Scheme of hydrogenotrophic methanogenesis in *Methanosarcinaceae* **(A)** and *Methanomicrobiaceae* **(B)** and observed protein abundance shifts in samples from reactor R2 (week 84). The numbers at the reaction arrows correspond to the reaction numbers in **Table 1**. Colored arrows indicate changing protein abundances between week 65 and 84. Red arrows indicate increasing protein abundances, blue arrows decreasing protein abundances, gray arrows show proteins not detected. Arrow thickness indicates the protein abundances. Thick = more abundant, thin = low abundant. **(A)** In *Methanosarcina,* the first and last steps of methanogenesis are chemiosmotically coupled and ATP generation is driven by a proton motive force. **(B)** In *Methanoculleus*, the first and last steps of methanogenesis are coupled by flavin-based electron bifurcation and ATP generation is driven by a sodium motive force. Fd, ferredoxin; MFR, methanofuran; H_4MPT, tetrahydromethanopterin; HS-CoM, coenzyme M; HS-CoB, coenzyme B; Ech, FeS hydrogenase; Vho/Hdr, F_{420} non-reducing hydrogenase/heterodisulfide reductase; Eha/Ehb, energy-converting hydrogenase complex; Mvh/Hdr, methyl-viologen-reducing hydrogenase (modified according to Thauer et al., 2008).

(corrinoid-containing) and seem to be affected heavily by decreasing cobalt concentrations.

Two enzymes involved in acetoclastic methanogenesis were detected for *Methanosarcinaceae* – the Ack required for acetate activation and the acetyl-CoA decarbonylase/synthase complex (ACDS) that cleaves the C-C and C-S bonds in the acetyl moiety of acetyl-CoA, oxidizes the carbonyl group to CO_2 and transfers the methyl group to tetrahydrosarcinapterin. The decreasing

Ack abundance in R2 could be a hint for decreasing activity of acetoclastic methanogenesis. In contrast, ACDS abundance was relatively stable, which could be explained by the carbon assimilation function of this enzyme complex that is also required during autotrophic growth under hydrogenotrophic conditions (Gencic et al., 2010). *Methanosarcinaceae* might upregulate the synthesis of ACDS subunits to ensure assimilation pathways upon switching from acetoclastic to hydrogenotrophic methanogenesis.

The abundance of formylmethanofuran dehydrogenase (Fmd), which is involved in the first step of hydrogenotrophic methanogenesis and requires Mo or W (Vorholt and Thauer, 2002), strongly increased in *Methanosarcinaceae*, confirming our hypothesis of a pathway switch in methanogenesis. Furthermore, the cobalt-dependent Mtr and Mcr increased in their abundances during TE deprivation. Mtr is a membrane-associated, corrinoid-containing enzyme and drives an energy-conserving ion pump (Gottschalk and Thauer, 2001). Mcr requires the prosthetic group F_{430}, which contains Ni as central atom (Whitman and Wolfe, 1980). In week 84, Mtr stayed at an increased level whereas Mcr decreased strongly below the level of week 65 (**Figure 2**). Further, *Methanosarcinaceae* suffered nickel limitation as visible by the decreasing abundance of [NiFe] hydrogenase without the ability to use Ni-free hydrogenases (Thauer et al., 2010).

Methanomicrobiaceae can only perform hydrogenotrophic methanogenesis and are also affected by TE deprivation. Several enzymes of the *Methanomicrobiaceae* increased in abundance and stayed more abundant in week 84 compared to the *Methanosarcinaceae*. This indicates an advantage of *Methanomicrobiaceae* over *Methanosarcinacaea*.

Methanosarcinaceae as well as *Methanomicrobiaceae* seem to stabilize their metabolism by increasing the expression of Mtr and Mcr to preserve the ion motive force for energy conservation with *Methanomicrobiaceae* being more successful. However, protein abundances detected by proteome analysis do not necessarily reflect the presence and activity of functional enzyme complexes. Subunits of TE-dependent enzyme complexes might also be expressed at elevated levels to compensate the increasing number of non-functional enzymes under TE-limiting conditions.

Our study has shown how methanogens react to TE deprivation by adapting their energy metabolism and suggests that *Methanosarcina* and *Methanoculleus* use different strategies to cope with such a limitation. Proteome analysis and tracer experiments revealed that *Methanosarcina* shifted from acetoclastic to hydrogenotrophic methanogenesis while *Methanoculleus* increased the hydrogenotrophic activity to sustain energy conservation. *Methanosarcina* as the versatile and multipotent "heavy duty" methanogen (De Vrieze et al., 2012) is more fastidious with regard to TE supplementation than *Methanoculleus*, which is a sufficient substitute not only as partner for SAOB but also as more robust methanogen stabilizing reactor performance under critical conditions.

AUTHOR CONTRIBUTIONS

BW and SK designed the study and the experiments. BW performed the experiments. BW, NJ, and DP analyzed the data. BW, NJ, DP, HH, and SK interpreted the data. BW drafted the manuscript and NJ, DP, HH, and critically revised it. All authors completed the final version of the manuscript and have approved it.

ACKNOWLEDGMENTS

BW was funded by the German Environmental Foundation (Deutsche Bundesstiftung Umwelt – DBU, grant number 20011/165) and by the Graduate School HIGRADE. Ute Lohse (UFZ Department of Environmental Microbiology) is acknowledged for amplicon sequencing and protein extraction, Kathleen Eismann (UFZ Department of Molecular Systems Biology) for her skilled technical assistance in protein extraction and LC-MS/MS. Dr. Dirk Wissenbach (UFZ Department of Molecular Systems Biology) is acknowledged for support in GC-MS measurements of labeled methane. The authors are grateful for the use of the analytical facilities of the Centre for Chemical Microscopy (ProVIS) at the Helmholtz Centre for Environmental Research, which is supported by European Regional Development Funds (EFRE – Europe funds Saxony) and the Helmholtz Association.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00405/full#supplementary-material

REFERENCES

Ariunbaatar, J., Esposito, G., Yeh, D. H., and Lens, P. N. L. (2016). Enhanced anaerobic digestion of food waste by supplementing trace elements: role of selenium (VI) and iron (II). *Front. Environ. Sci.* 4:8. doi: 10.3389/fenvs.2016.00008

Borrel, G., O'Toole, P. W., Harris, H. M., Peyret, P., Brugère, J.-F., and Gribaldo, S. (2013). Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol. Evol.* 5, 1769–1780. doi: 10.1093/gbe/evt128

Burke, S. A., and Krzycki, J. A. (1997). Reconstitution of monomethylamine:coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. *J. Biol. Chem.* 272, 16570–16577.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Choong, Y. Y., Norli, I., Abdullah, A. Z., and Yhaya, M. F. (2016). Impacts of trace element supplementation on the performance of anaerobic digestion process: a critical review. *Bioresour. Technol.* 209, 369–379. doi: 10.1016/j.biortech.2016.03.028

Conklin, A., Stensel, H. D., and Ferguson, J. (2006). Growth kinetics and competition between *Methanosarcina* and *Methanosaeta* in

mesophilic anaerobic digestion. *Water Environ. Res.* 78, 486–496. doi: 10.2175/106143006X95393

Demirel, B. (2014). Major pathway of methane formation from energy crops in agricultural biogas digesters. *Critical Rev. Environ. Sci. Technol.* 44, 199–222 doi: 10.1080/10643389.2012.710452

Demirel, B., and Scherer, P. (2011). Trace element requirements of agricultural biogas digesters during biological conversion of renewable biomass to methane. *Biomass Bioenergy* 35, 992–998. doi: 10.1016/j.biombioe.2010.12.022

De Vrieze, J., Hennebel, T., Boon, N., and Verstraete, W. (2012). *Methanosarcina*: the rediscovered methanogen for heavy duty biomethanation. *Bioresour. Technol.* 112, 1–9. doi: 10.1016/j.biortech.2012.02.079

Diekert, G., Konheiser, U., Piechulla, K., and Thauer, R. K. (1981). Nickel requirement and factor F_{430} content of methanogenic bacteria. *J. Bacteriol.* 148, 459–464.

Dridi, B., Fardeau, M. L., Ollivier, B., Raoult, D., and Drancourt, M. (2012). Methanomassiliicoccus luminyensis gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 62, 1902–1907. doi: 10.1099/ijs.0.033712-0

Evans, P. N., Parks, D. H., Chadwick, G. L., Robbins, S. J., Orphan, V. J., Golding, S. D., et al. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350, 434–438. doi: 10.1126/science.aac7745

Evranos, B., and Demirel, B. (2015). The impact of Ni, Co and Mo supplementation on methane yield from anaerobic mono-digestion of maize silage. *Environ. Technol.* 36, 1556–1562. doi: 10.1080/09593330.2014.997297

Ferguson, D. J. Jr., and Krzycki, J. A. (1997). Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from *Methanosarcina barkeri*. *J. Bacteriol.* 179, 846–852. doi: 10.1128/jb.179.3.846-852.1997

Fish, J. A., Chai, B., Wang, Q., Sun, Y., Brown, C. T., Tiedje, J. M., et al. (2013). FunGene: the functional gene pipeline and repository. *Front. Microbiol.* 4:291. doi: 10.3389/fmicb.2013.00291

Gencic, S., Duin, E. C., and Grahame, D. A. (2010). Tight coupling of partial reactions in the acetyl-CoA decarbonylase/synthase (ACDS) multienzyme complex from *Methanosarcina thermophila*. *J. Biol. Chem.* 285, 15450–15463. doi: 10.1074/jbc.M109.080994

Glass, J. B., and Orphan, V. J. (2012). Trace metal requirements for microbial enzymes involved in the production and consumption of methane and nitrous oxide. *Front. Microbiol.* 3:61. doi: 10.3389/fmicb.2012.00061

Gottschalk, G., and Thauer, R. K. (2001). The Na^+-translocating methyltransferase complex from methanogenic archaea. *Biochim. Biophys. Acta* 1505, 28–36. doi: 10.1016/S0005-2728(00)00274-7

Karakashev, D., Batstone, D. J., and Angelidaki, I. (2005). Influence of environmental conditions on methanogenic compositions in anaerobic biogas reactors. *Appl. Environ. Microbiol.* 71, 331–338. doi: 10.1128/AEM.71.1.331-338.2005

Karakashev, D., Batstone, D. J., Trably, E., and Angelidaki, I. (2006). Acetate oxidation is the dominant methanogenic pathway from acetate in the absence of Methanosaetaceae. *Appl. Environ. Microbiol.* 72, 5138–5141. doi: 10.1128/AEM.00489-06

Kohrs, F., Wolter, S., Benndorf, D., Heyer, R., Hoffmann, M., Rapp, E., et al. (2015). Fractionation of biogas plant sludge material improves metaproteomic characterization to investigate metabolic activity of microbial communities. *Proteomics* 15, 3585–3589. doi: 10.1002/pmic.201400557

Miller, T. L., and Wolin, M. J. (1985). Methanosphaera stadtmaniae gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch. Microbiol.* 141, 116–122.

Nobu, M. K., Narihiro, T., Kuroda, K., Mei, R., and Liu, W. T. (2016). Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J.* 10, 2478–2487. doi: 10.1038/ismej.2016.33

Park, C., Bega, A., Unlu, C., Chadderton, R. A., McKean, W. R., Kohl, P. M., et al. (2010). Acetoclastic methanogens in an anaerobic digester could be susceptible to trace metal supplementation. *Water Sci. Technol.* 62, 2905–2911. doi: 10.2166/wst.2010.161

Patterson, T., Esteves, S., Dinsdale, R., and Guwy, A. (2011). An evaluation of the policy and techno-economic factors affecting the potential for biogas upgrading for transport fuel use in the UK. *Energy Policy* 39, 1806–1816. doi: 10.1016/j.enpol.2011.01.017

Popp, D., Harms, H., and Sträuber, H. (2016). The alkaloid gramine in the anaerobic digestion process - inhibition and adaptation of the methanogenic community. *Appl. Microbiol. Biotechnol.* 100, 7311–7322. doi: 10.1007/s00253-016-7571-z

Popp, D., Plugge, C. M., Kleinsteuber, S., Harms, H., and Sträuber, H. (2017). Inhibitory effect of coumarin on syntrophic fatty acid oxidizing and methanogenic cultures and biogas reactor microbiomes. *Appl. Environ. Microbiol.* 83, e00438-17. doi: 10.1128/AEM.00438-17

Porsch, K., Wirth, B., Toth, E. M., Schattenberg, F., and Nikolausz, M. (2015). Characterization of wheat straw-degrading anaerobic alkali-tolerant mixed cultures from soda lake sediments by molecular and cultivation techniques. *Microb. Biotechnol.* 8, 801–814. doi: 10.1111/1751-7915.12272

Poulsen, M., Schwab, C., Jensen, B. B., Engberg, R. M., Spang, A., Canibe, N., et al. (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat. Commun.* 4:1428. doi: 10.1038/ncomms2432

Raboni, M., Viotti, P., and Capodaglio, A. G. (2015). A comprehensive analysis of the current and future role of biofuels for transport in the European Union (EU). *Ambiente e Agua* 10, 9–21. doi: 10.4136/ambi-agua.1492

Scherer, P., Lippert, H., and Wolff, G. (1983). Composition of the major elements and trace elements of 10 methanogenic bacteria determined by inductively coupled plasma emission spectrometry. *Biol. Trace Elem. Res.* 5, 149–163. doi: 10.1007/BF02916619

Schmidt, T., Pröter, J., Scholwin, F., and Nelles, M. (2013). Anaerobic digestion of grain stillage at high organic loading rates in three different reactor systems. *Biomass Bioenergy* 55, 285–290. doi: 10.1016/j.biombioe.2013.02.010

Schnürer, A., Houwen, F. P., and Svensson, B. H. (1994). Mesophilic syntrophic acetate oxidation during methane formation by a triculture at high ammonium concentration. *Arch. Microbiol.* 162, 70–74. doi: 10.1007/BF00264375

Speece, R. E., Parkin, G. F., and Gallagher, D. (1983). Nickel stimulation of anaerobic digestion. *Water Res.* 17, 677–683. doi: 10.1016/0043-1354(83)90237-3

Steinberg, L. M., and Regan, J. M. (2008). Phylogenetic comparison of the methanogenic communities from an acidic, oligotrophic fen and an anaerobic digester treating municipal wastewater sludge. *Appl. Environ. Microbiol.* 74, 6663–6671. doi: 10.1128/AEM.00553-08

Sträuber, H., Bühligen, F., Kleinsteuber, S., Nikolausz, M., and Porsch, K. (2015). Improved anaerobic fermentation of wheat straw by alkaline pre-treatment and addition of alkali-tolerant microorganisms. *Bioengineering* 2, 66–93. doi: 10.3390/bioengineering2020066

Surendra, K. C., Takara, D., Hashimoto, A. G., and Khanal, S. K. (2014). Biogas as a sustainable energy source for developing countries: opportunities and challenges. *Renew. Sus. Energy Rev.* 31, 846–859. doi: 10.1016/j.rser.2013.12.015

Takashima, M., Speece, R. E., and Parkin, G. F. (1990). Mineral requirements for methane fermentation. *Crit. Rev. Environ. Control* 19, 465–479. doi: 10.1080/10643389009388378

Thauer, R. K., Kaster, A. K., Goenrich, M., Schick, M., Hiromoto, T., and Shima, S. (2010). Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and H_2 storage. *Annu. Rev. Biochem.* 79, 507–536. doi: 10.1146/annurev.biochem.030508.152103

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* 6, 579–591. doi: 10.1038/nrmicro1931

van der Meijden, P., Heythuysen, H. J., Pouwels, A., Houwen, F., van der Drift, C., and Vogels, G. D. (1983). Methyltransferases involved in methanol conversion by *Methanosarcina barkeri*. *Arch. Microbiol.* 134, 238–242. doi: 10.1007/BF00407765

Vanwonterghem, I., Evans, P. N., Parks, D. H., Jensen, P. D., Woodcroft, B. J., Hugenholtz, P., et al. (2016). Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* 1:16170 doi: 10.1038/nmicrobiol.2016.170

von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., Bastida, F., Herbst, F.-A., et al. (2013). Insights from quantitaive metaproteomics and protein-stable isotope probing into microbial ecology. *ISME J.* 7, 1877–1885. doi: 10.1038/ismej.2013.78

Vorholt, J. A., and Thauer, R. K. (2002). "Molybdenum and tungsten enzymes in enzymes in C1 metabolism," in *Metal Ions in Biological Systems,* eds A. Sigel and H. Sigel (Basel, NY: Marcel Dekker), 571–603.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Wassenaar, R. W., Keltjens, J. T., van der Drift, C., and Vogels, G. D. (1998). Purification and characterization of dimethylamine:5-hydroxybenzimidazolyl-cobamide methyltransferase from *Methanosarcina barkeri* Fusaro. *Eur. J. Biochem.* 253, 692–697. doi: 10.1046/j.1432-1327.1998.2530692.x

Weiland, P. (2010). Biogas production: current state and perspectives. *Appl. Microbiol. Biotechnol.* 85, 849–860. doi: 10.1007/s00253-009-2246-7

Whitman, W. B., Bowen, T. L., and Boone, D. R. (2006). The methanogenic bacteria. *Prokaryotes* 4, 165–207. doi: 10.1007/0-387-30743-5_9

Whitman, W. B., and Wolfe, R. S. (1980). Presence of nickel in factor F_{430} from *Methanobacterium bryantii*. *Biochem. Biophys. Res. Comm.* 92, 1196–1201.

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.

Wintsche, B., Glaser, K., Sträuber, H., Centler, F., Liebetrau, J., Harms, H., et al. (2016). Trace elements induce predominance among methanogenic activity in anaerobic digestion. *Front. Microbiol.* 7:2034. doi: 10.3389/fmicb.2016.02034

Ziganshin, A. M., Liebetrau, J., Pröter, J., and Kleinsteuber, S. (2013). Microbial community structure and dynamics during anaerobic digestion of various agricultural waste materials. *Appl. Microbiol. Biotechnol.* 97, 5161–5174. doi: 10.1007/s00253-013-4867-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wintsche, Jehmlich, Popp, Harms and Kleinsteuber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 February 2018
doi: 10.3389/fmicb.2018.00354

Linking the Effect of Antibiotics on Partial-Nitritation Biofilters: Performance, Microbial Communities and Microbial Activities

Alejandro Gonzalez-Martinez[1], Alejandro Margareto[2,3], Alejandro Rodriguez-Sanchez[4], Chiara Pesciaroli[4], Silvia Diaz-Cruz[2,3], Damia Barcelo[2,3] and Riku Vahala[1]*

[1] Department of Built Environment, School of Engineering, Aalto University, Espoo, Finland, [2] Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research, Barcelona, Spain, [3] Catalan Institute for Water Research, Scientific and Technological Park of the University of Girona, Girona, Spain, [4] Institute of Water Research, University of Granada, Granada, Spain

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Mariusz Cycoń, Medical University of Silesia, Poland
Steve Lindemann, Purdue University, United States

***Correspondence:**
Alejandro Gonzalez-Martinez
alejandro.gonzalezmartinez@aalto.fi; agon@ugr.es

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *07 November 2017*
Accepted: *14 February 2018*
Published: *26 February 2018*

Citation:
Gonzalez-Martinez A, Margareto A, Rodriguez-Sanchez A, Pesciaroli C, Diaz-Cruz S, Barcelo D and Vahala R (2018) Linking the Effect of Antibiotics on Partial-Nitritation Biofilters: Performance, Microbial Communities and Microbial Activities. Front. Microbiol. 9:354. doi: 10.3389/fmicb.2018.00354

The emergence and spread of antibiotics resistance in wastewater treatment systems have been pointed as a major environmental health problem. Nevertheless, research about adaptation and antibiotics resistance gain in wastewater treatment systems subjected to antibiotics has not been successfully developed considering bioreactor performance, microbial community dynamics and microbial activity dynamics at the same time. To observe this in autotrophic nitrogen removal systems, a partial-nitritation biofilter was subjected to a continuous loading of antibiotics mix of azithromycin, norfloxacin, trimethoprim, and sulfamethoxazole. The effect of the antibiotics mix over the performance, bacterial communities and bacterial activity in the system was evaluated. The addition of antibiotics caused a drop of ammonium oxidation efficiency (from 50 to 5%) and of biomass concentration in the bioreactor, which was coupled to the loss of ammonium oxidizing bacteria *Nitrosomonas* in the bacterial community from 40 to 3%. Biomass in the partial nitritation biofilter experienced a sharp decrease of about 80% due to antibiotics loading, but the biomass adapted and experienced a growth by stabilization under antibiotics feeding. During the experiment several bacterial genera appeared, such as *Alcaligenes*, *Paracoccus,* and *Acidovorax*, clearly dominating the bacterial community with >20% relative abundance. The system reached around 30% ammonium oxidation efficiency after adaptation to antibiotics, but no effluent nitrite was found, suggesting that dominant antibiotics-resistant phylotypes could be involved in nitrification–denitrification metabolisms. The activity of ammonium oxidation measured as *amoA* and *hao* gene expression dropped a 98.25% and 99.21%, respectively, comparing the system before and after the addition of antibiotics. On the other hand, denitrifying activity increased as observed by higher expression of *nir* and *nos* genes (83.14% and 252.54%, respectively). In addition, heterotrophic nitrification *cyt c-551* was active only after the antibiotics addition. Resistance to the antibiotics was presumably given by *ermF*, *carA* and *msrA* for azithromycin, mutations of the *gyrA* and

grlB for norfloxacin, and by *sul123* genes for sulfamethoxazole. Joined physicochemical and microbiological characterization of the system were used to investigate the effect of the antibiotics over the bioprocess. Despite the antibiotics resistance, activity of *Bacteria* decreased while the activity of *Archaea* and *Fungi* increased.

Keywords: partial-nitritation, antibiotic resistance, metatranscriptomics, autotrophic nitrogen removal, microbial population, microbial activity

INTRODUCTION

Several human and animal wastes, such as pharmaceutical industry effluents or livestock wastes, are treated through anaerobic digestion processes. In this way, anaerobic digestion offers an efficient treatment of these wastes due to low energy requirements, low sludge production and generation of methane as valuable subproduct, among others (Nasir et al., 2012). Concerning the threat of antibiotics resistance, the concentrations of antibiotics reported in pharmaceutical industry effluents and livestock wastes are in the range of 1–100 mg L^{-1} (Massé et al., 2014; Wang et al., 2015). In addition, the anaerobic digestion systems have shown poor antibiotics removal treatment in the range of 1–10 mg L^{-1} concentrations (Massé et al., 2014; Wang et al., 2015). In this sense, treatment systems handling effluent downstream anaerobic digestion processes should be able to withstand these antibiotics concentrations.

Anaerobic digestion supernatant is a residue obtained after anaerobic digestion process, which is characterized by high ammonium concentrations and low organic matter content. In the last 10 years, autotrophic nitrogen removal technologies have been developed for an efficient, cheap bioremediation of this waste (van der Star et al., 2007). These technologies rely on the unique metabolism of *"Candidatus Brocadiales"* bacteria, named as anaerobic ammonium oxidation, in which ammonium is oxidized using nitrite as terminal electron acceptor, yielding molecular nitrogen as a result (van Teeseling et al., 2016). In practical operation for the treatment of anaerobic digestion supernatant, the oxidation of half of the influent ammonium is required prior to the development of nitrogen removal through anaerobic ammonium oxidation. Thus, a partial-nitritation is a necessary previous step for the successful performance of autotrophic nitrogen removal technologies (van der Star et al., 2007). In order to develop and control separately the partial-nitritation and anaerobic ammonium oxidation, the partial-nitritation/anammox technology has been developed.

The purpose of the partial-nitritation process is to oxidize half of the influent ammonium for a subsequent treatment by *"Candidatus Brocadiales"* bacteria. Traditionally, partial-nitritation processes have been operated in suspended growth configuration. Nevertheless, attached growth partial-nitritation systems have been successfully operated, showing several advantages over suspended growth partial-nitritation processes such as lower hydraulic retention time (HRT) required (Rodriguez-Sanchez et al., 2016a,b). The ability of attached-growth partial-nitritation processes to operate under ciprofloxacin antibiotic pressure has been tested, showing that the system was impacted by the addition of the antimicrobial in terms of partial-nitritation performance and bacterial community structure (Gonzalez-Martinez et al., 2014b).

On the other hand, an emerging environmental and human health issue in the world today is the antibiotics resistant bacteria (Eramo et al., 2017; Zhu et al., 2017). Infectious bacteria resistant to antibiotics increase the costs of human health and increase the mortality of humans (Cao et al., 2016; Miller et al., 2016). Moreover, the bacterial evolution caused by antibiotics exposure could endanger the environmental health worldwide (Botts et al., 2017; Huang et al., 2017). Given that it has been reported that the introduction of antibiotics resistance to the environment is done by human and animal waste disposal, the wastewater treatment plants worldwide stand as a key element in the antibiotics resistance spread (Yu et al., 2016).

The bioprocess engineering science has usually used a monitoring approach for the study of bioreactors based on physicochemical determinations and evaluation of their microbial community structure (Gonzalez-Martinez et al., 2014b, 2015). Nevertheless, in terms of bioprocess functioning, the activity of the microorganisms is fundamental to the performance of biosystems, and thus the understanding of the metabolisms of the microbial communities in wastewater bioreactors (Rodriguez et al., 2015). To date, little work has been made in the investigation of metabolisms regarding the functioning of bioreactors (FitzGerald et al., 2015; Lawson et al., 2017). However, the effect of antibiotics over bioreactor functioning have never been attempted through metatranscriptomics approach.

For these reasons, the ability of an attached biofilm partial-nitritation process to handle high concentrations of an antibiotics mixture has been observed in terms of partial-nitritation performance, bacterial community structure dynamics and microbial community activity through metatranscriptomics. A synthetic wastewater emulating anaerobic digestion leachate from pharmaceutical wastewater treatment plant was used, containing different widely used antibiotics with different action mechanisms, such as macrolide azithromycin (AZT), quinolone norfloxacin(NOR) and sulfonamide trimethoprim (TMP)/sulfamethoxazole (SMZ), in high concentrations reported by several authors (Massé et al., 2014; Wang et al., 2015; Aydin et al., 2015; Meng F. et al., 2015; Meng L.W. et al., 2015). The results obtained showed the effect of the antibiotics addition over performance of the bioreactor and microbial activity.

MATERIALS AND METHODS

Bioreactor Configuration, Start-up, and Operation

A lab-scale partial-nitritation biofilter was set-up for the experimentation in a similar approach based on the previous research (Gonzalez-Martinez et al., 2014a) (Supplementary Figure S1). The system was composed of a 5 L bioreactor filled in its whole volume with BioFlow9 carriers. The influent was introduced by the means of a peristaltic pump to achieve a HRT of 7 h. During the whole experiment, the temperature was controlled at 33 ± 1°C and the pH at 7.5 ± 0.2 with H_2SO_4 0.1 M and NaOH 0.1 M for pH control. The aeration was constant, distributed equally through the bioreactor's volume and maintained at 1.5 ± 0.3 mg-O_2 L^{-1}. The system was started-up using 1 L of activated sludge from the Los Vados WWTP (Granada, Spain) full-scale activated sludge bioreactor. For the start-up phase, synthetic wastewater simulating anaerobic digester supernatant was used following previous research on partial-nitritation biofilters (Gonzalez-Martinez et al., 2014b, 2016a). The composition of the wastewater is shown in **Table 1**. The start-up of the system was prolonged until a stable and efficient partial-nitritation performance under the optimal operational conditions to obtain 50% ammonium and 50% nitrite.

The system was then operated for 30 days before to start the experiment under steady-state conditions. After this step, the experiment started from day 1 to day 60 days under steady-state conditions without antibiotics followed by 60 days (from day 60 to day 120) amended with antibiotics. For this purpose, the partial-nitritation bioreactor was continuously fed with wastewater #1 from day 1 to day 60 but from this day the influent wastewater composition was changed to wastewater #2 with a high antibiotics concentration containing AZT, NOR, SMZ, and TMP, which was maintained until the end of the experiment. Four antibiotics were continuously added to the influent to observe its effect on the partial-nitritation system. The synthetic wastewater composition with the antibiotics is shown in **Table 1**. The operation under antibiotics conditions was prolonged until the system reached steady-state conditions under the antibiotics loading, which occurred in a period of 2 months.

The operation of the bioreactor and the handling of biomass, influent and effluents were done following the Biosafety Level-2 from EU Directives 2000/54/EC on the protection of workers from risks related to exposure to biological agents at work.

Determination of Nitrogenous Inorganic Compounds

The inorganic forms of nitrogen ammonium, nitrite, and nitrate were measured in the influent and effluent of the partial-nitritation biofilter on a daily basis by means of ionic chromatography.

Determination of Biomass Concentration Attached to Carriers

The biomass attached to the BioFlow 9 carriers was evaluated daily. The method for its determination followed the procedure described previously for partial-nitritation biofilters (Gonzalez-Martinez et al., 2014b, 2016a).

Determination of Antibiotics

Standards of AZT, NOR, SMZ, and TMP were supplied by Sigma-Aldrich (Steinheim, Germany). Deuterated standards AZT-d3, enrofloxacin-d5, sulfamethazine-d4, and TMP-d3, were supplied by Toronto Research Chemicals (Toronto, ON, Canada) and were of >99% purity. Methanol (MeOH), HPLC-grade water, and acetone LC-MS-grade solvents were purchased from Merck (Darmstadt, Germany) and acetonitrile (ACN) from Fischer Scientific (Loughborough, United Kingdom). High quality nitrogen (N_2) and argon (Ar) were supplied by Abelló Linde (Barcelona, Spain).

Individual stock standard solutions of antibiotics were gravimetrically prepared in MeOH at 100 μg L^{-1}. A 5 μg mL^{-1} internal standards (IS) solution and a stock standard solution of the mixture of all antibiotics at 5 and 1 μg mL^{-1}, respectively, were prepared. Working mixture standards solutions were freshly prepared by appropriate dilution of the stock standard mixture solution in MeOH. Ten-point calibration curves were built in the range 10 ng L^{-1} to 1,500 ng L^{-1} and their pH was adjusted with HCOOH at 0.1% in volume before analysis. Calibration curves were run in every blanks and samples batch. All solutions were stored in the dark at −20°C and allowed to equilibrate at room temperature before use. Quantification of target antibiotics was performed by means of calibration curves obtained by linear regression analysis using the internal standardization on the basis of the best suited isotopically labeled compound for each analyte. The data were adjusted to linear least square regression curve with 1/x weighting index.

On-line solid phase extraction coupled to high performance liquid chromatography-tandem-mass spectrometry (on-line SPE-HPLC-MS/MS) (Margareto et al., unpublished) was used to determine the concentrations of AZT, NOR, TMP, and SMZ

TABLE 1 | Synthetic wastewater composition.

Chemical	Synthetic wastewater #1	Synthetic wastewater #2	Unit
$(NH_4)_2SO_4$	2.35	2.35	g/L
$NaHCO_3$	3.25	3.25	g/L
$CaCl_2$	0.30	0.30	g/L
KH_2PO_4	0.07	0.07	g/L
$MgSO_4$	0.02	0.02	g/L
$FeSO_4$ $7H_2O$	0.009	0.009	g/L
H_2SO_4	0.005	0.005	g/L
Azithromycin	0	8	mg/L
Norfloxacin	0	2	mg/L
Sulfamethoxazole	0	9	mg/L
Trimethoprim	0	3	mg/L

antibiotics in the influents and effluents of the partial-nitritation biofilter.

The influent and effluent water samples were conveniently diluted to fit into the calibration range, spiked with the IS mixture to reach a concentration of 500 ng L^{-1} and pH adjusted to 2.7. The automated on-line pre-concentration, purification and chromatographic separation of the analytes were performed using an on-line SPE–HPLC instrument Symbiosis^{TM} Pico (Spark Holland; Emmen, Netherlands). The elution of the retained antibiotics from the SPE cartridges (OASIS HLB) and the subsequent chromatographic separation was performed using a mobile phase consisted of HPLC-grade water (A) and ACN (B), both 0.1% in HCOOH. The separation was performed on a Purospher® STAR RP-18 ec (125 mm × 2 mm, 5 μm particle size) LC-column from Merck (Darmstadt, Germany) with a guard column of the same material and setting a flow rate of 0.3 mL min^{-1}. The following gradient was used (all steps linear): 0 min, 85% A; decreasing in 3 min to 20% A, kept constant for 7 min; 10 min 5% A;, kept constant for 2 min, and then returned to initial conditions in 3 min and, finally, 7 additional minutes to allow the column to equilibrate.

Detection was carried out in a 4000 QTRAP mass spectrometer (Applied Biosystems, Foster City, CA, United States) equipped with turbospray electrospray ionization (ESI) source. Data acquisition was performed in the positive ESI mode [ESI (+)] operated in selected reaction monitoring (SRM), allowing us to record two precursor ion-product ion mass transitions per compound. The most intense transition was used for quantification, and the other was used for confirmation, according to the identification and confirmation criteria for the analysis of drugs and other contaminants as defined by Commission Decision 2002/657/EC, implementing the Council Directive 96/23/EC. For data acquisition, peak area integration and quantification calculations the Analyst software v 1.5 (Sciex, Concord, ON, Canada) was used.

The experimental MS/MS parameters are listed in Supplementary Table S1. The method performance is summarized in Supplementary Table S2. Briefly; satisfactory average recovery rates ranging from 48.1 to 108.3% for influent and from 46.0 to 90.3% for effluent waters were achieved and expressed as the mean recovery values obtained at spike levels of 100 ng L^{-1}, 250 ng L^{-1} and 500 ng L^{-1} with 3 replicates each. High sensitivity, expressed in terms of limits of detection (LOD) and quantification (LOQ) were reached, in the ranges 1.3–6.9 ng L^{-1} and 4.3–23.1 ng L^{-1}, respectively, for influent and from 0.4 to 2.3 ng L^{-1} and from 1.5 to 7.8 ng L^{-1}, respectively, for influent.

Biomass Collection, DNA Extraction, and *iTag* High-Throughput Sequencing Procedure

For the purpose of characterization of bacterial community structure in the partial-nitritation biofilter, a total of 90 carriers (about 100 mL volume in total) was collected taking carriers distributed across the whole bioreactor's volume. Biomass detachment from carriers was done according to previous procedures in partial nitritation biofilters (Gonzalez-Martinez et al., 2014b; Rodriguez-Sanchez et al., 2016a,b). Briefly, the carriers were sonicated for 3 min, then the detached biomass was submerged in saline solution (0.9% NaCl) and subjected to centrifugation at 3,500 rpm during 10 min at room temperature. The liquid supernatant was discarded and the collected biomass was kept at −20°C for subsequent DNA extraction.

The DNA extraction was done using the FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, OH, United States) and the FastPrep apparatus following the instructions given by the manufacturer of the DNA extraction kit used and the procedure detailed in Gonzalez-Martinez et al. (2014a). The extracted DNA was then kept at −20°C and sent to Research and Testing Laboratory (Lubbock, TX, United States) for *iTag* high-throughput sequencing process.

The *iTag* high-throughput sequencing process was done using the Illumina MiSeq technology and the Illumina MiSeq Reagents Kit v3 at 2x300. The primers 28F-519R (5′-GAGTTTGATCNTGGCTCAG-3′ and 5′-GTNTTACNGCGGCKGCTG-3′, respectively), which have been used previously to determine the bacterial community structure of partial nitritation biofilters subjected to antibiotics (Gonzalez-Martinez et al., 2014b), were used for the amplification of the V1-V3 hypervariable regions of the 16S rRNA gene of *Bacteria*. The conditions of the PCR developed for the high-throughput sequencing were: 180 s at 94°C; 40 cycles of: 30 s at 94°C, 40 s at 60°C, 60 s at 72°C; 300 s at 72°C.

Metatranscriptomic Analysis

The activities of bacterial communities in the bioreactor were determined by metatranscriptomic analyses at the beginning and the end of the operation under antibiotics pressure. A total volume of 100 mL of carriers samples were collected from the system and submerged into RNA Protect for the preservation of RNAs. The samples were kept at −80°C and sent to Research and Testing Laboratory.

The PowerMicrobiome RNA Isolation Kit (MOBIO, United States) was used for extraction of RNA and removal of genomic DNA from the samples following the instructions given by the manufacturer. The extraction process started with a lysis of cells using glass bead tubes and lysis solution. Then, a binding matrix captured all nucleic acids in the lysate, and DNA was removed by on-column DNase and wash solution, yielding only the RNA, which was preserved in RNase-free water for subsequent real time PCR.

The extracted RNA was then used for construction of libraries using the KAPA Stranded RNA-Seq Library Preparation Kit (KAPA Biosystems, United States) following the manufacturer's instructions. The protocol started with fragmentation of RNA under by the means of heating and presence of Mg^{+2} with insert sizes ranging 200–300. Then, a conversion of first strand to cDNA using random primers was done, followed by a conversion of second strand to transform cDNA:RNA into double-stranded cDNA. dscDNA was then marked with dUTP and ligated to adapters. The adapter-ligated sequences library was then amplified by PCR. The library was sequenced using Illumina MiSeq technology and the Illumina MiSeq Reagents Kit v3 at

2x300. The raw sequences obtained are available in the SRA under the accession number SRP127026.

iTag High-Throughput Sequencing Post-process

The analysis of the *iTag* high-throughput sequencing samples was done with the software mothur v1.34.4 (Schloss et al., 2009). First, paired-end reads were merged into contigs avoiding the generation of ambiguous bases in the overlap region. The contigs generated first passed a quality screening control to eliminate sequences with ambiguous bases and more than eight homopolymers. The remnant sequences were then aligned against the SiLVA SEED 123 release database, and those that failed to align properly were discarded for the analysis. Failure at alignment was regarded as: (i) failed to align at the position of the forward primer, and (ii) ended further than the 95% of the aligned sequences. The remaining sequences were then preclustered into a 2-bases threshold (Huse et al., 2010) and then checked for the presence of chimeric sequences using UCHIME v4.1 (Edgar et al., 2011), which were deleted from the analysis. After chimera deletion, the sequences were taxonomically affiliated and those that failed to classify within the domain *Bacteria* were eliminated.

To develop the bacterial ecology analysis of the *iTag* high-throughput sequencing samples, these were rarified and cut to form subsamples with 15778 sequences each. The sequences in each subsample were separately used to calculate a Phylip distance matrix between them, which was later utilized for the clustering of the sequences into OTUs within a 97% identity threshold. Representative sequences were then chosen for taxonomic classification of each of the OTUs using the SiLVA SEED 123 release database. The taxonomically affiliated OTUs were finally used to form a consensus taxonomy of OTUs within a cutoff of 80%.

Ecological Analysis of the *iTag* Sequencing Subsamples

The *iTag* sequencing subsamples were subjected to an ecological analysis to determine their diversity coverage, α-diversity and β-diversity. For the diversity coverage, the Good's coverage, the redundancy abundance-weighted coverage and the complexity curve of each subsample were calculated. The Good's coverage was calculated using the species richness of each subsample in relation to their number of reads. The redundancy abundance-weighted coverage was calculated through NonPareil software taking a query set size of 1,000 sequences among the unique sequences within each subsample, allowing a minimum overlap of 50% and a 95% of identity between sequences (Rodriguez-R and Konstantinidis, 2014a,b). The complexity curves were calculated using aRarefactWin software. The Shannon–Wiener, Simpson, Chao1, Pielo's evenness and Berger–Parker α-diversity indices were calculated using PAST software. The Morisita–Horn and symmetric β-diversity indices were calculated using the vegan 2.0 and vegetarian packages implemented in statistical software R.

Post-process of Metatranscriptomic Data

The raw data obtained from the high-throughput sequencing of the cDNA were processed to yield information of the bacterial activity within the systems. The procedure mainly involved mapping against publicly available databases for the detection of rRNA, tRNA, and taxonomy and functional annotation of mRNA, as this is a common procedure for metatranscriptomics analyses (Aguiar-Pulido et al., 2016). Nevertheless, the use of closed reference databases restricts the results obtained to the scope of the information contained on these in terms of phylogenetic and functional affiliation of RNA.

For the metatranscriptomics pipeline, first, a quality trimming of sequences was done the paired-end sequences were merged into contigs using the software mothur v1.34.4 (Schloss et al., 2009) avoiding the appearance of ambiguous bases in the overlap region due to different nucleotide quality at the same position. Then, the contigs were trimmed to eliminate those with 1 or more ambiguous bases, 9 or more homopolymers, lower average quality score of 20 and length shorter than 200 bp.

Then, the remaining sequences were screened to eliminate those belonging to rRNA genes. This was done using BLAST software and the SiLVA Ref SSU and LSU databases, and sequences that found a match with an e-value $< 10^{-10}$ were removed from the analysis. Then, an additional screening to eliminate other non-coding RNA sequences was done by BLAST search against non-coding RNA databases offered by RNACentral[1]. After the removal of non-coding RNA, the remaining sequences were thus considered mRNA.

The mRNA sequences were then BLASTed against several reference databases of genomes derived from RefSeq and NCBI for taxonomic affiliation. These databases were obtained from RefSeq[2] and were used for the affiliation of the mRNA sequences to the taxonomic groups of *Archaea*, *Fungi*, *Protozoa*, *Virus,* and *Plasmids*. Also, an additional database of representative prokaryotic genomes derived from the NCBI[3] was also used for classification of the mRNA in order to complete the taxonomic affiliation for the domain *Bacteria*. In all cases, matches were considered positive for bitscore >50 and e-value $< 10^{-5}$ (Leimena et al., 2013; Mobberley et al., 2015).

The mRNA sequences that found a positive match with the RefSeq *Archaea* genomes database were then BLASTed against the corresponding RefSeq *Archaea* protein database for functional annotation. This was also true for the cases of *Fungi*, *Protozoa*, *Virus,* and *Plasmids*. The sequences that found a positive match with the NCBI representative prokaryotic genomes were BLASTed against the SwissProt database. The functional annotation was defined for e-values $< 10^{-5}$, as suggested by previous authors (Leimena et al., 2013; Sayadi et al., 2016).

[1] http://rnacentral.org/

[2] ftp://ftp.ncbi.nlm.nih.gov/refseq/release/

[3] ftp://ftp.ncbi.nlm.nih.gov/blast/db/

Multivariate Redundancy Analyses

The operational parameters of the partial-nitritation biofilter (effluent ammonium, nitrite and nitrate concentrations; total nitrogen removal; biomass concentration; removal efficiencies of antibiotics AZT, NOR, TMP, and SMZ) were linked to the bacterial community structure of the partial-nitritation biofilter and the activity of its microbial communities through two multivariate redundancy analyses. These were done using the software CANOCO 4.5 for Windows and calculated through 499 unconstrained Monte-Carlo simulation under a full permutation model.

RESULTS AND DISCUSSION

Partial-Nitritation Performance of the Partial-Nitritation Biofilter

Once the system reached steady-state conditions, the partial nitritation performance was nearly ideal, with around 50% of ammonium oxidized to nitrite and negligible nitrate concentrations during the 60 days under the synthetic wastewater conditions. The performance of the partial-nitritation biofilter under these conditions was similar to previous experiments (Gonzalez-Martinez et al., 2014b). Nevertheless, the addition of the antibiotics mix to the influent severely impacted the ammonium oxidation capacity (**Figure 1**). The partial-nitritation biofilter showed a reduction in ammonium oxidation efficiency from around 50% to around 5% after 6 days of the addition of the antibiotics. The system reached higher ammonium oxidation efficiencies over time, finishing in a steady-state value of around 30%. In this sense, the addition of antibiotics in high concentrations caused an irreversible loss of performance of the partial-nitritation biofilter. This has also been observed at low concentrations of ciprofloxacin antibiotics (Gonzalez-Martinez et al., 2014a). A sharp decrease in the effluent nitrite was observed coupled to the loss in ammonium oxidation. However, after 80 days of operation (20 days with antibiotics), a higher ammonium oxidation performance was observed. In this way, the increase in ammonium oxidation was not correlated with an increase in effluent nitrite concentration, but with an increase in nitrogen removal from the system. In this sense, it is possible that the addition of antibiotics triggered denitrification metabolisms from nitrite when the antibiotics or other organic matter within the biofilm were used as organic matter. This has also been observed in partial nitritation biofilters subjected to amino acids and antibiotics loading (Gonzalez-Martinez et al., 2014b, 2016a,b).

The addition of the antibiotics mix caused an effect on the attached biofilm in the partial-nitritation biofilter, with a sharp decrease after the addition (**Figure 2**). With ongoing operation under the antibiotics mix the biomass slowly increased, but never could reach the values before the antibiotics addition. In this sense, the performance of the system was also influenced by the loss of biomass. This result resembles the one obtained for biomass growth in partial nitritation biofilters under different ciprofloxacin concentrations, with the exception that the system reached higher biomass concentrations under stable operations with antibiotic loading (Gonzalez-Martinez et al., 2014b). The differences may be driven by the differences in antibiotics concentrations between the two experiments, which were about 10^6-fold, and the presence of a mix of antibiotics against only one compound. In this sense, the higher concentrations and the presence of several antibiotic compounds could exert more pressure over the partial nitritation biomass.

Removal of Antibiotics

A net removal of the antibiotics AZT, NOR, TMP, and SMZ was observed during the experiment time (**Table 2**). SMZ was the antibiotic with lower removal efficiencies, showing a mean of 7.24 ± 4.90% during the operation under antibiotics addition. The higher mean was for AZT with 44.94 ± 14.55%, while for NOR and TMP were of 32.74 ± 9.63% and 32.59 ± 10.31%, respectively. Interestingly, the removal of AZT decreased significantly from day 90 to day 120, which might be explained by exhaustion of the adsorption capacity of the biofilm for AZT antibiotic, since reports showed that only around 10% AZT in WWTPs could be sorbed to biomass (Ivanová et al., 2017). Norfloxacin, SMZ, and trimethoprim were found to be removed by activated sludge process rather than becoming attached to biofilms (Yan et al., 2014a,b), and therefore their removal in this experiment could be attributed to degradation. The removal efficiencies at operational day 75 were the lowest for NOR, TMP, and SMZ. This might be caused by a turnover of microbial species within the bioreactor at that time. Also, the sorption of AZT, NOR, and SMZ could cause a net removal during the operation of the partial-nitritation biofilter.

Ecological Analysis of the *iTag* High-Throughput Sequencing Subsamples

The coverage of the *iTag* sequencing subsamples was sufficient to capture the bacterial diversity of the partial-nitritation biofilter. In this sense, the Good's coverage index showed more than 97.4% coverage, while the redundancy abundance-weighted coverage had a minimum of 92.5% (Supplementary Table S3). Along with results obtained from complexity curves (Supplementary Figure S2), the coverage of the high-throughput sequencing subsamples seemed to be successful. Overall, the partial-nitritation biofilter without antibiotics addition had higher species richness than the system under the antibiotics loading, which found its lower diversity at the end of the antibiotics experiment. In this sense, the complexity curves showed a loss in diversity as the system was operated with antibiotics influent.

The Chao-1 index, which is related closely to species richness, showed its higher value at operational day 60 (S60), then decreased constantly until day 120 and had a slight increase by the end of the experiment (Supplementary Table S4), as showed by the complexity curves. Pielou's evenness and Simpson indices values indicated that the diversity of the system decreased with operation time, showing that the antibiotics addition exerted an efficient selection that allowed few bacterial phylotypes to thrive under the antibiotics addition. In this sense, the evenness was lowest during operational days 75

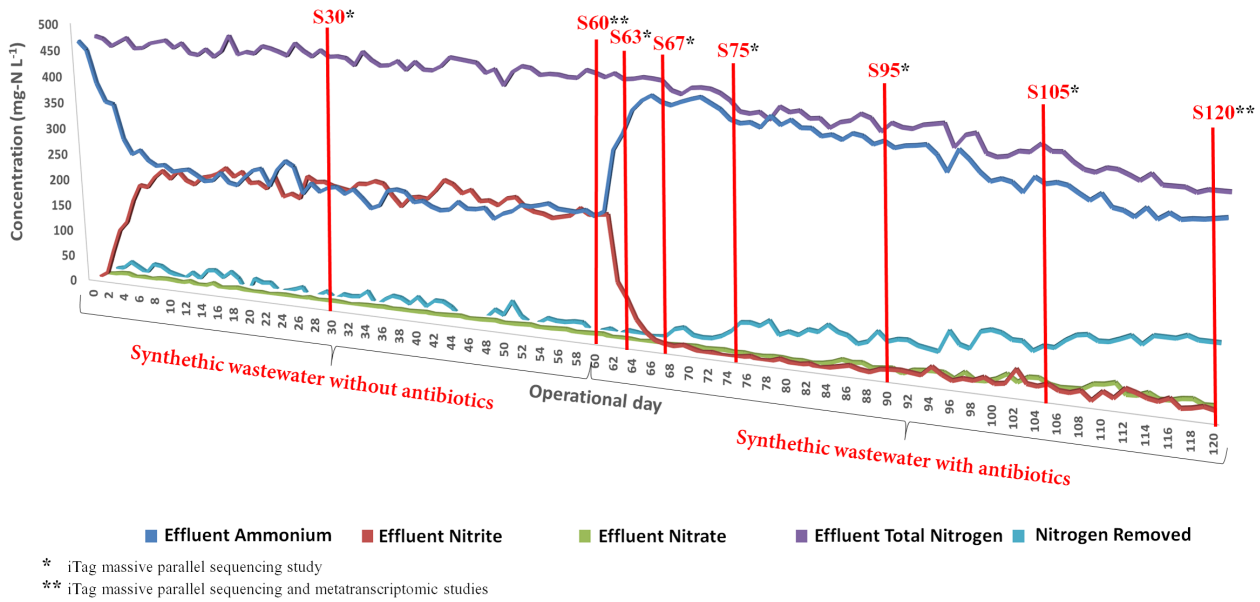


FIGURE 1 | Performance of the partial-nitritation biofilter during the experiment.

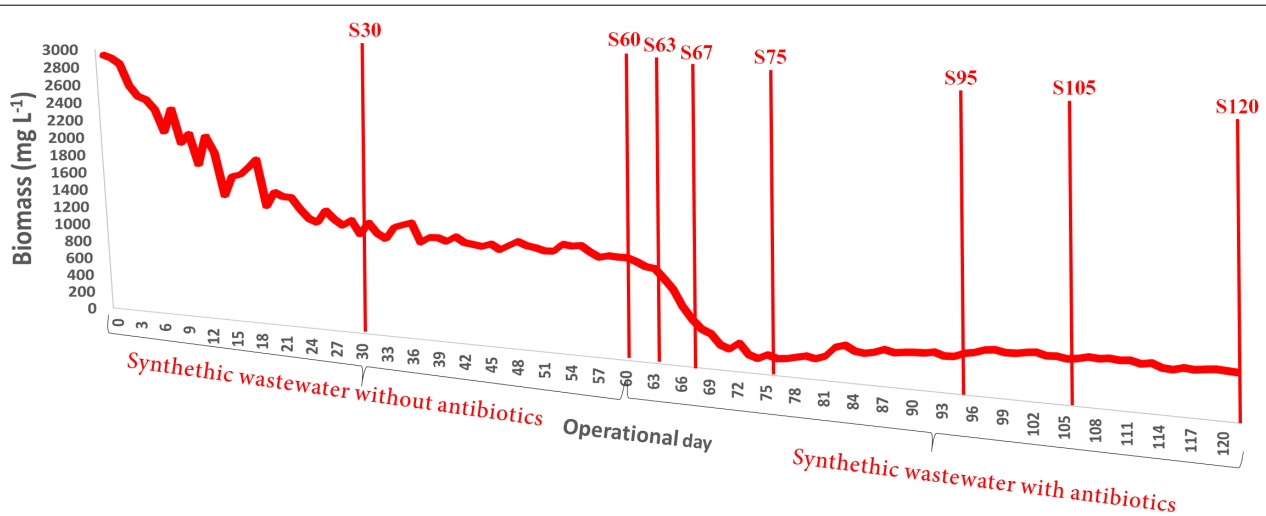


FIGURE 2 | Evolution of the attached biofilm concentration in the partial-nitritation biofilter during the experiment.

(S75) and 90 (S90). The patterns regarding evenness of the bacterial communities in the partial-nitritation biofilter during the antibiotics experiment was also shown by the Berger–Parker index. Thus, the Shannon–Wiener index showed an increase from operational days 30 to 63 due to increase in bacterial diversity and evenness, and a continuous decrease as the antibiotics were added due to loss in bacterial diversity and evenness. Therefore, the values of α-diversity suggested that the adaptation of the system to the antibiotics mix influenced its bacterial community structure in terms of species diversity and evenness. Nevertheless, the lack of replicates in biological samples sets a limitation over the data collected regarding some of these indices, such as Chao-1.

The Morisita–Horn and the symmetric indices for the pair of *iTag* sequencing subsamples of interest are represented in Supplementary Figure S3. It could be found that the dominant bacterial phylotypes at operational days 30 (S30) and 60 (S60) were very different, as shown by high Morisita–Horn index value, while their rare phylotypes were of low similarity due to low symmetric index values. The dominant genera changed after the addition of the antibiotics and never recovered during the operation time, while the rare species persisted in the system. The dominant phylotypes changed drastically from day 60 (S60) to day 63 (S63), followed by a mild change from day 63 (S63) to day 67 (S67), and a period of acclimation from day 67 (S67) up to day 90 (S90). Nevertheless, from day 90 (S90) to day 105 (S105)

TABLE 2 | Antibiotics concentration (mg L^{-1}) in the influent used in the experimental bioreactor and percentage of antibiotics removal during the experiment.

Influent (mg L^{-1})			Azithromycin 8.0	Norfloxacin 2.0	Sulfamethoxazole 9.0	Trimethoprim 3.0
Removal (%)	**Sample name**	**Operational time (days)**	**Azithromycin**	**Norfloxacin**	**Sulfamethoxazole**	**Trimethoprim**
	S30*	30	–	–	–	–
	S60**	60	–	–	–	–
	S63	63	49.66	26.49	2.30	35.23
	S67	67	44.23	35.60	10.89	36.52
	S75	75	51.22	17.99	0.00	17.30
	S90	90	60.02	43.63	11.37	37.69
	S105	105	47.24	41.56	8.17	23.39
	S120	120	17.30	31.15	10.70	45.42

*Without antibiotics in the influent. **First day with addition of antibiotics in the influent.

a substantial change was observed, and then another stabilization was reached at day 120 (S120).

Bacterial Community Dynamics in the Partial-Nitritation Biofilter

The bacterial community dynamics showed that the addition of the antibiotics mix caused a deep change in the bacterial community structure of > 1% phylotypes in the partial-nitritation biofilter (**Figure 3**).

The first two pyrosequencing samples (days 30 and 60) were taken when the physico-chemical performance was stable in the partial nitritation bioreactor. In this way, it could be proven that the microbial population changes were caused by the antibiotics addition. Thus, at operational days 30 and 60, when the biofilter was fed with no-antibiotics influent, the bacterial community structure was clearly dominated by ammonium oxidizing *Nitrosomonas* genus, also with proliferation of strictly anaerobic *Chloroflexi*-belonging *Thermomarinilinea* (Nunora et al., 2013), which could develop anaerobic cell material degradation by utilization of *N*-acetylglucosamine within the biofilm as other members of its phylum (Gonzalez-Martinez et al., 2016a,b); heterotrophic nitrifier-aerobic denitrifier *Comamonas*, which has been found previously in partial-nitritation biofilters (Gonzalez-Martinez et al., 2016a); *Salirhabdus*, another aerobic denitrifier (Albuquerque et al., 2017); and *Chiayiivirga*, an

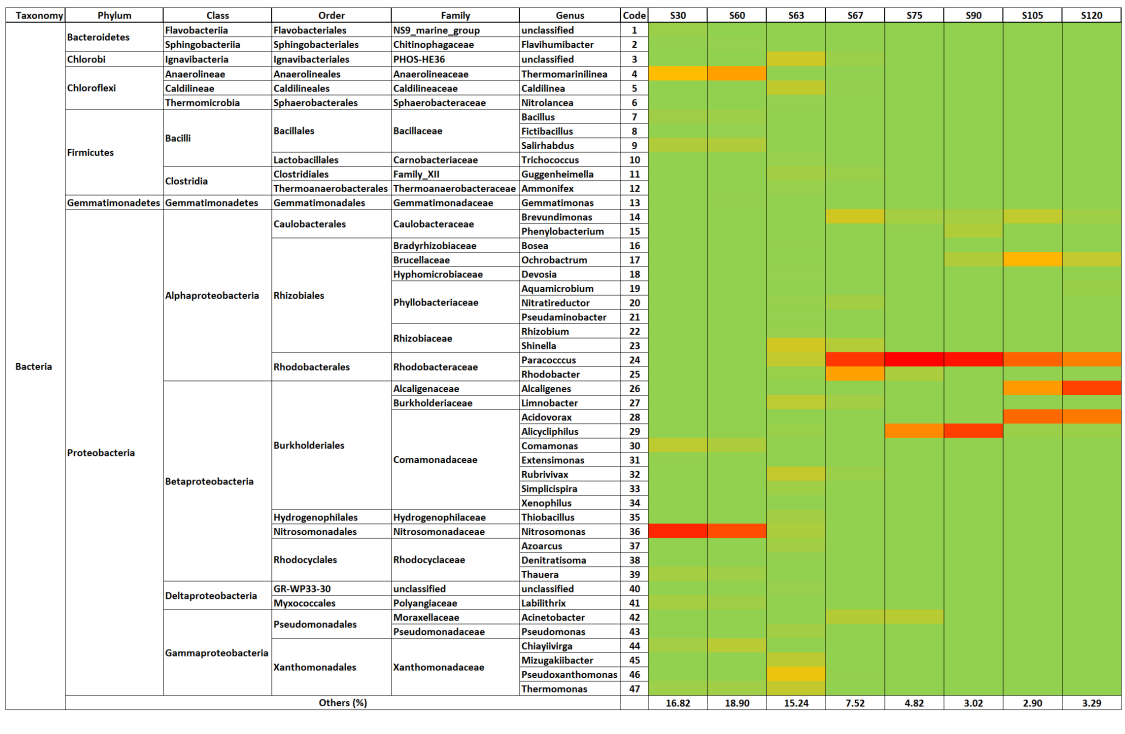


FIGURE 3 | Bacterial community structure of the high-throughput sequencing subsamples.

aerobic, heterotrophic bacteria (Hsu et al., 2013). The bacterial community structure could be related to the performance of the bioreactor, with dominant *Nitrosomonas* linked to ammonium oxidation as observed in previous experimentations on partial-nitritation biofilters (Gonzalez-Martinez et al., 2014a,b, 2015, 2016a; Rodriguez-Sanchez et al., 2016a,b).

The Morisita–Horn index analysis showed that antibiotics caused significant changes in the bacterial community structure of the partial-nitritation biofilter. Accordingly, the analysis of bacterial dynamics showed that the addition of the antibiotics mix to the influent of the partial-nitritation biofilter caused a great decrease of *Nitrosomonas* relative abundance (from around 35–40% to 3%). The impact of the antibiotics caused loss of biomass and the proliferation of *Pseudoxanthomonas*, *Shinella*, *Rubrivivax*, *Thermomonas,* or *Paracoccus*, among others. After 7 days of antibiotics addition *Paracoccus* dominated the system at the same level of abundance as *Nitrosomonas* did with no antibiotics. *Rhodobacter* and *Brevundimonas* were also of importance at this operational day. *Paracoccus* was also the clearly dominant genus at operational day 75 (S75) followed by *Alicycliphilus*, with *Acinetobacter*, *Rhodobacter* and *Brevundimonas* being important but at much lower relative abundance. *Paracoccus* and *Alicycliphilus* still dominated the system by operational day 90 (S90), in which also important populations of *Ochrobactrum*, *Phenylobacterium,* and *Brevundimonas* appeared. By operational day 105 (E45) the genus *Alicycliphilus* had a very low presence in comparison with its abundance at operational day 90 (S90). *Paracoccus* still dominated, but its abundance was evenly matched with *Acidovorax*, with *Alcaligenes* and *Ochrobactrum* falling not far behind. By the last operational day 120 (S120) the domination belonged to *Alcaligenes*, followed by *Acidovorax* and *Paracoccus* which were also significantly and equally abundant. Interestingly, genera *Paracoccus*, *Rhodobacter*, *Brevundimonas*, *Alicycliphilus*, *Acinetobacter*, *Acidovorax* and *Alcaligenes* have been reported for denitrification metabolisms (Oosterkamp et al., 2013; Lu et al., 2014; Ji et al., 2015; Chen et al., 2016a,b; Qu et al., 2016; Yang et al., 2016; Tsubouchi et al., 2017). The presence of *Alcaligenes*, *Paracoccus,* and *Acidovorax* after the addition of an antibiotics mix of AZT, NOR, SMZ, and TMP was also found in a CANON bioreactor (Rodriguez-Sanchez et al., 2017). In this sense, these three genera are able to develop multi-antibiotics resistance and their importance in the proliferation and spread of antibiotics resistance genes in wastewater treatment systems should be explored.

Activity in the Partial-Nitritation Biofilter in Absence and Presence of the Antibiotics Mix

Overview of Microorganisms Groups *Archaea*, *Bacteria*, *Fungi,* and *Protozoa*

The mRNA profile of the partial-nitritation biofilter before the addition of the antibiotics mix and at the end of the experiment were monitored using a metatranscriptomic approach. In this sense, results suggested notable differences in the activity of microorganisms within the biofilter (**Table 3**). As such, before the antibiotics addition, the majority of the mRNA (63.59%) was affiliated to the *Bacteria* domain followed by *Protozoa* group (31.45%), with *Archaea* and *Fungi* having a very low relative abundance (0.69% and 0.63%, respectively) in the general activity within the system. On the other hand, operation under the antibiotics conditions resulted in lower activity of *Bacteria* (54.32%) and *Protozoa* (25.91%) and a 10-fold increase in the activities of *Archaea* (8.26%) and *Fungi* (8.35%) with respect to the no-antibiotics scenario. Therefore, the mRNA profile showed that the addition of antibiotics increased the activities of *Archaea* and *Fungi* in the partial-nitritation biofilter. This could be caused by the lower susceptibility of archaeal and fungal phylotypes to the antimicrobials AZT, NOR, TMP, and SMZ used in the experiment.

The Activity of *Bacteria* Domain

In addition, changes in the global activity of the different microorganisms groups were observed. In the case of *Bacteria* at the no-antibiotics scenario, the dominant genera in terms of mRNA presence was *Nitrosomonas* (68.46%) (**Figure 4A**), which is related to the ammonium oxidation metabolism. This suggested the crucial role of *Nitrosomonas* in the ammonium oxidation in partial-nitritation biofilters, since ammonia monooxygenase activity (7.90% of total mRNA identified in the no-antibiotic scenario) was related to high partial-nitritation performance (about 50% ammonium-nitrite) and high relative

TABLE 3 | Taxonomic classification of mRNA found in the metatranscriptomics samples.

	Taxonomy			
	Before antibiotics (Day 60)		After antibiotics (day 120)	
	Reads	Identified Percent	Reads	Identified Percent
Identified mRNA	38144	100.00	13299	100.00
Prokaryotes	24255	63.59	7224	54.32
Archaea	265	0.69	1099	8.26
Fungi	242	0.63	1110	8.35
Virus	122	0.32	12	0.09
Plasmid	1264	3.31	408	3.07
Protozoa	11996	31.45	3446	25.91

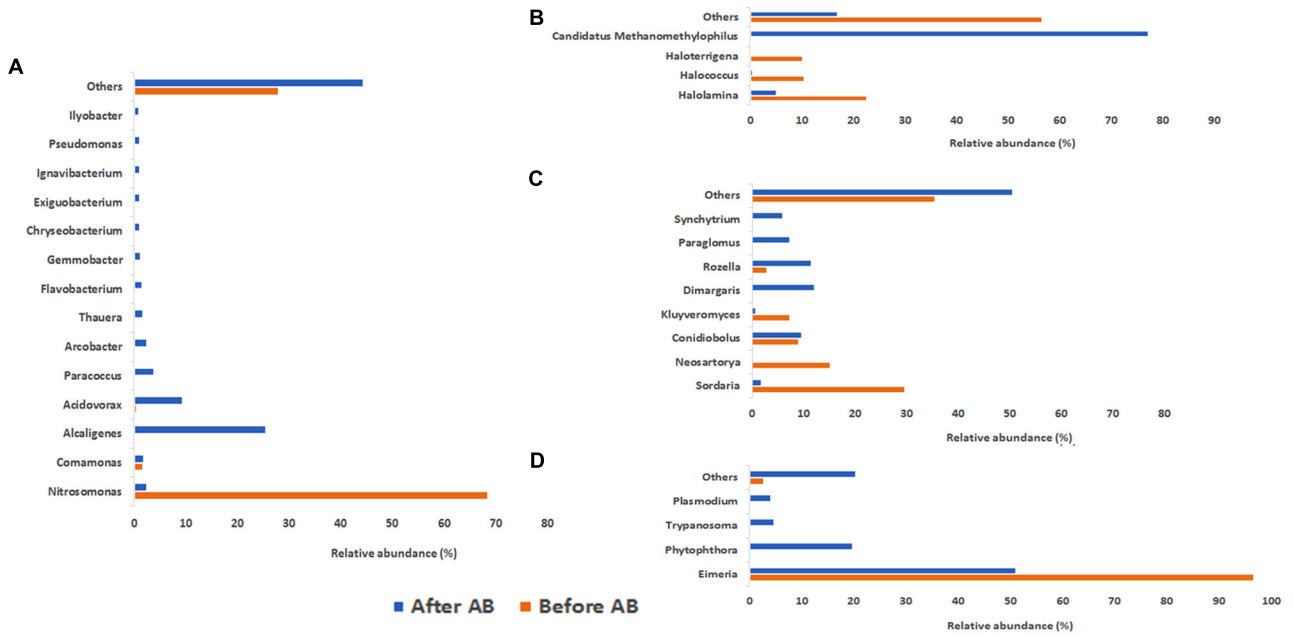


FIGURE 4 | Contribution to mRNA activity at genus level of *Bacteria* **(A)**, *Archaea* **(B)**, *Fungi* **(C)**, and *Protozoa* **(D)** as determined by metatranscriptomic analysis. Before AB: before the antibiotics addition; After AB: after the antibiotics addition.

abundance of *Nitrosomonas* (34.17%). Nevertheless, after the operation under the antibiotics mix, ammonia monooxygenase and hydroxylamine oxidoreductase had a 0.16% and 0.016%, respectively, along with the loss of *Nitrosomonas* relative abundance (0.00%) and partial-nitritation performance (around 70% ammonium-2.7% nitrite). *Nitrosomonas* did not change its dominant activity profile with respect to the antibiotics addition, being ammonia monooxygenase at both scenarios (10.90% and 5.32%, respectively), while a certain decrease in the ammonium oxidation was observed after the addition of the antibiotics (**Figure 5A**).

After the addition of the antibiotics, the most expressed proteins corresponded to *Alcaligenes* (25.56%), *Acidovorax* (9.34%), and *Paracoccus* (3.89%). We observed no other transcripts from *Alcaligenes* besides the DNA-directed RNA polymerase subunit β' prior to the addition of antibiotics. However, after the operation under the antibiotics addition, the activity of this genus had a high diversity with predomination of elongation factor G (1.87%) or acetyl-coenzyme A synthase (1.82%), among others, while DNA-directed RNA polymerase subunit β and DNA-directed RNA polymerase subunit β' were still found at high relative abundances (0.83% and 0.83%, respectively) (**Figure 5B**). The presence of acetyl-coenzyme A synthase indicated the heterotrophic metabolism of this genus.

The Activity of *Archaea* Domain

With respect to the domain *Archaea* under no antibiotics, the most active genera were *Halolamina* (22.64%), *Halococcus* (10.57%), *Haloterrigena* (10.19%), *Halalkalicoccus* (5.66%), and *Methanosarcina* (4.91%), among others. On the other hand, the domination under antibiotics operation belonged to *Candidatus Methanomethylophilus* (77.25%), followed by *Methanobacterium* (6.37%), *Halolamina* (5.10%), and *Methanosarcina* (1.46%) (**Figure 4B**). Interestingly, the clear dominance of *Methanomethylophilus* at the antibiotics scenario showed that this microorganism had a wide diversity of mRNA, with predominance of the elongation factor 1-alpha (30.11%), the peptidase C25 (20.69%) and the *N*-acetylglucosamine-1-phosphate uridyltransferase (8.05%) (**Figure 5C**). The elongation factor 1-alpha has been proposed as an omnipresent mechanism for quality control, elongation and termination of protein synthesis in *Archaea* (Saito et al., 2010). Also, peptidase C25 was found to be released by marine archaeal phylotypes for the degradation of proteins in marine sediments (Lloyd, 2013), and its high presence in the activity profile of *Methanomethylophilus* may signify a heterotrophic metabolism. *N*-acetylglucosamine-1-phosphate uridyltransferase, on the other hand, is related to cell wall material formation (Zhang et al., 2009).

The Activity of *Fungi* Microorganisms

Within the *Fungi* members, the activity profile changed from a domination of *Sordaria* (29.75%), *Neosartorya* (15.29%), *Conidiobolus* (9.09%), and *Kluyveromyces* (7.44%) to a composition mainly formed by *Dimargaris* (12.07%), *Rozella* (11.53%), *Conidiobolus* (9.64%), and *Paraglomus* (7.39%) (**Figure 4C**). The activity of *Sordaria* mainly changed from chlorocatechol degradation by expression of dienelactone hydrolase (10.87%) to glycan degradation expressed by high relative abundance of endo β xylanase precursor (4.76%) (**Figure 5D**).

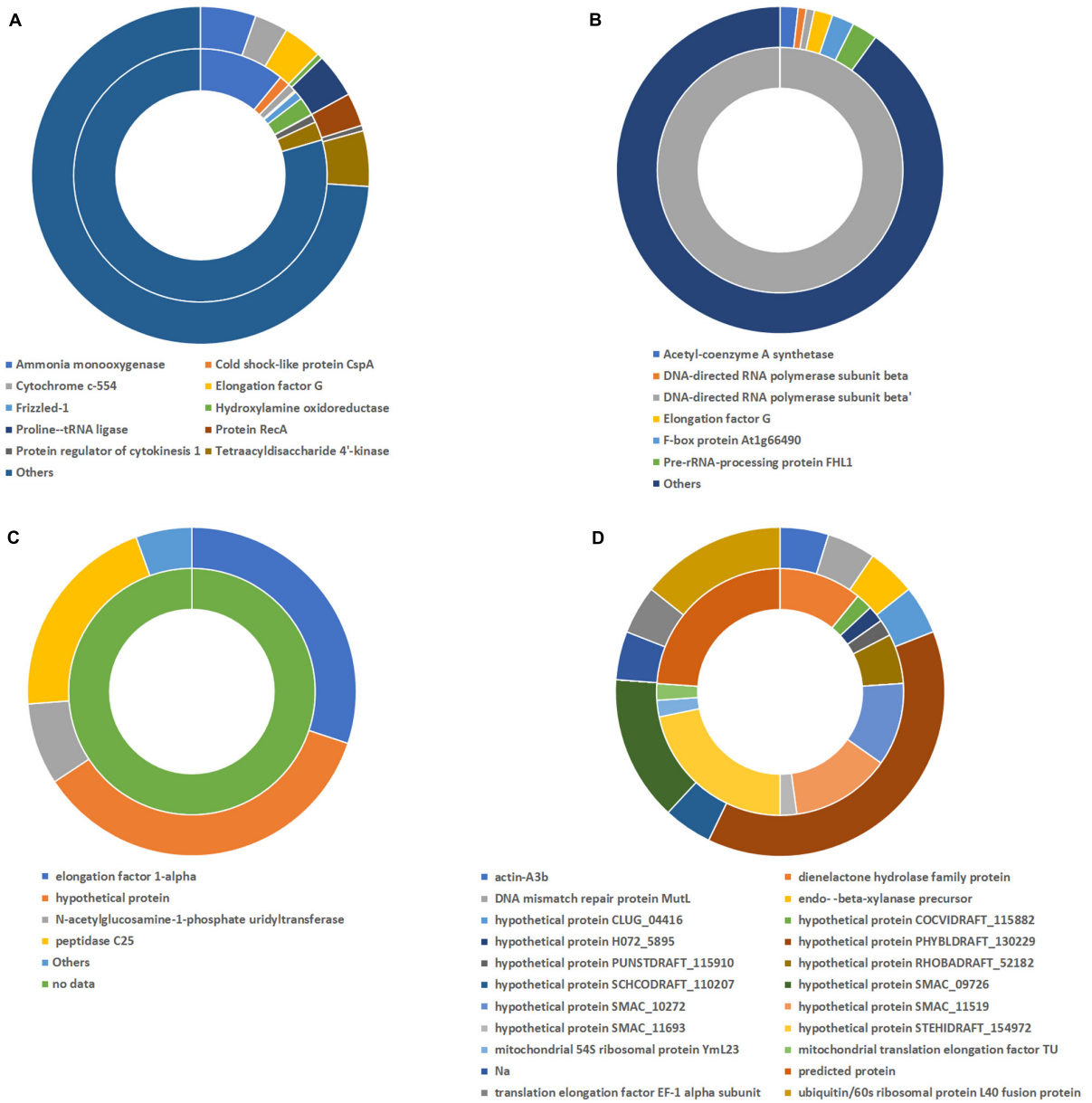


FIGURE 5 | Activity profile of genera *Nitrosomonas* **(A)**, *Alcaligenes* **(B)**, *Candidatus Methanomethylophilus* **(C)**, and *Sordaria* **(D)** as determined by metatranscriptomic analysis. The inner circle represents the activity before the antibiotics addition and the outer circle that after the antibiotics addition.

The Activity of *Protozoa* Microorganisms

At the no-antibiotics scenario, the protozoan mRNA was almost exclusively affiliated to *Eimeria* (96.70%). Nevertheless, after operation under antibiotics, its affiliated mRNA dropped in half (51.16%), yielding an ecological niche for other protozoa, such as *Phytophthora* (19.76%), *Trypanosoma* (4.64%), and *Plasmodium* (4.06%), to grow (**Figure 4D**).

The Activity Concerning the Nitrogen Cycle

The **Figure 6** highlights the activity involving the expression of proteins related to *amo* and *hao* genes, complex *cytochrome c-552* and *cytochrome c-554* genes for nitrification; and *nar*/*nap*, *nor*, *nir* and *nos* genes, *cytochrome b/c1* and *cytochrome c-551* for denitrification. As such, it was found that the antibiotics pressure drastically dropped the activity of ammonium oxidizing bacteria, observed in a decrease of ammonia monooxygenase and hydroxylamine oxidoreductase and the activity of complexes cytochrome c-552 and c-554 (with reductions of 98.25, 99.21, 41.84, and 89.83%, respectively). On the other hand, the addition of antibiotics caused an increase in the activities of nitrate reduction, nitrite reduction and nitrous oxide reduction. Moreover, the activity of complexes cytochromes b/c1 and c-551, which are related to heterotrophic nitrification-aerobic denitrification metabolisms (Guo et al., 2013), was only detected

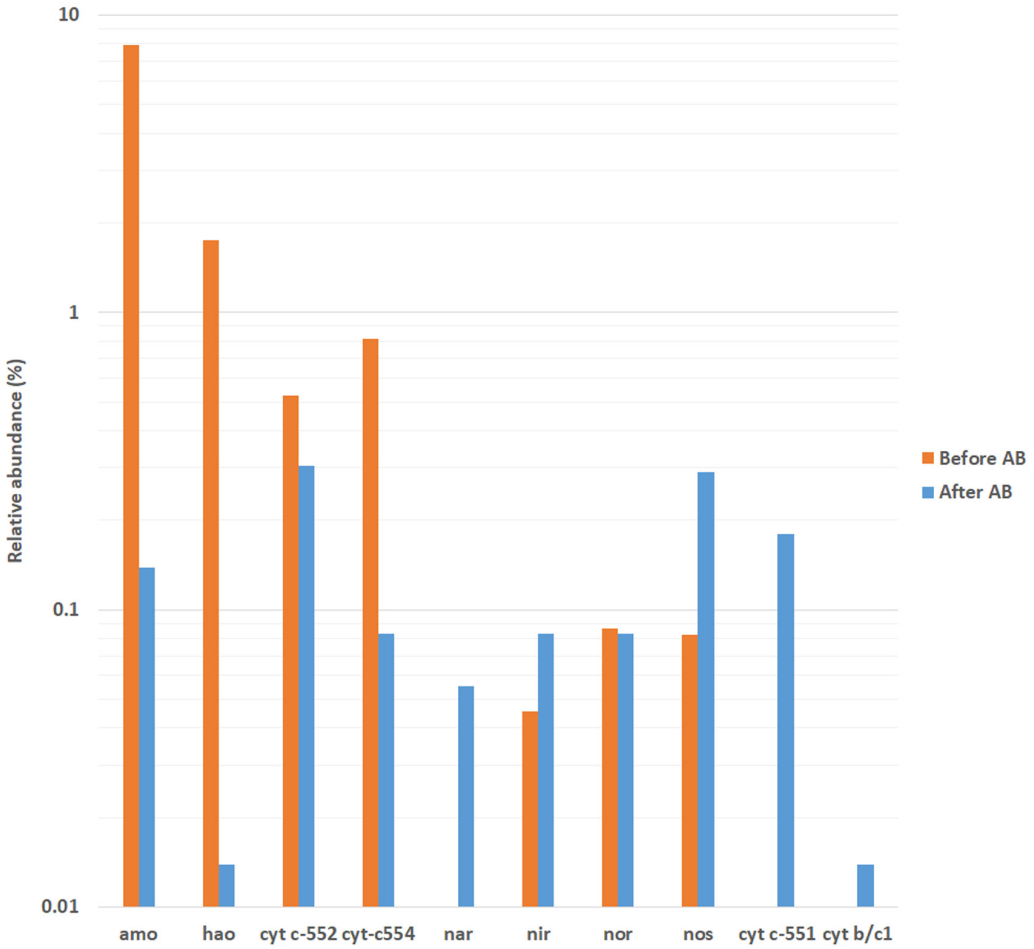


FIGURE 6 | Relative abundance of genes implicated in autotrophic ammonium oxidation and heterotrophic nitrification-denitrification metabolisms as determined by metatranscriptomic analysis. Before AB: before the antibiotics addition; After AB: after the antibiotics addition.

in operation under antibiotics. In this sense, as suggested by the determination of nitrogenous ions, the addition of antibiotics decreased the capacity of ammonium oxidation but increased the capacity of denitrification, which led to higher effluent ammonium and lower effluent total nitrogen concentrations, respectively. Also, as suggested by the analysis of bacterial community composition by 16S rRNA bacterial gene high-throughput sequencing, the antibiotics addition allowed the proliferation of bacteria with heterotrophic nitrification-aerobic denitrification capabilities, similarly to partial-nitritation biofilters subjected to ciprofloxacin (Gonzalez-Martinez et al., 2014b).

The results obtained showed that the addition of antibiotics to the partial-nitritation biofilter not only affected the microbial community structure, but also changed the metabolic pathways related to nitrogen. These results have been also reported by experiments in partial-nitritation biofilters subjected to ciprofloxacin antibiotic (Gonzalez-Martinez et al., 2014b). In this sense, the antibiotic compounds are potentially dangerous for partial-nitritation systems, causing severe losses in performance of ammonium oxidation to nitrite and therefore leading to system failure.

The Activity Related to Antibiotics Resistance

The presence of activities related to antibiotics resistance mechanisms was investigated within the mRNA identified. These activities were related to known resistance mechanisms to: macrolide-lincosamide-steptrogramin group (affiliated with genes *erm*, *car*, *msr*, *ole*, *smr*, *tlr*, *vga*, *vgb*, *lmr*, *sfa*, *mef*, *ere*, *lnu*, *vat* and *mph*), in which AZT is a member (Roberts, 2008); sulfonamides group (affiliated to genes *sul*), which counts SMZ (Wang et al., 2014); and fluoroquinolone group (affiliated with genes *gyr*, *par*, *grl*, *qnr*, *oqx*, *qep*, *pat*, *acr* and *tol*), in which NOR stands (Redgrave et al., 2014). Activities of these genes were only detected within the Bacteria domain. A graphical representation of the relative abundance of these genes before and after the addition of the antibiotics mix is given in **Figure 7**.

No macrolides-resistance genes were found before the antibiotics addition, but three were found at the end of the operation after antibiotics addition: target-site mutation ermF

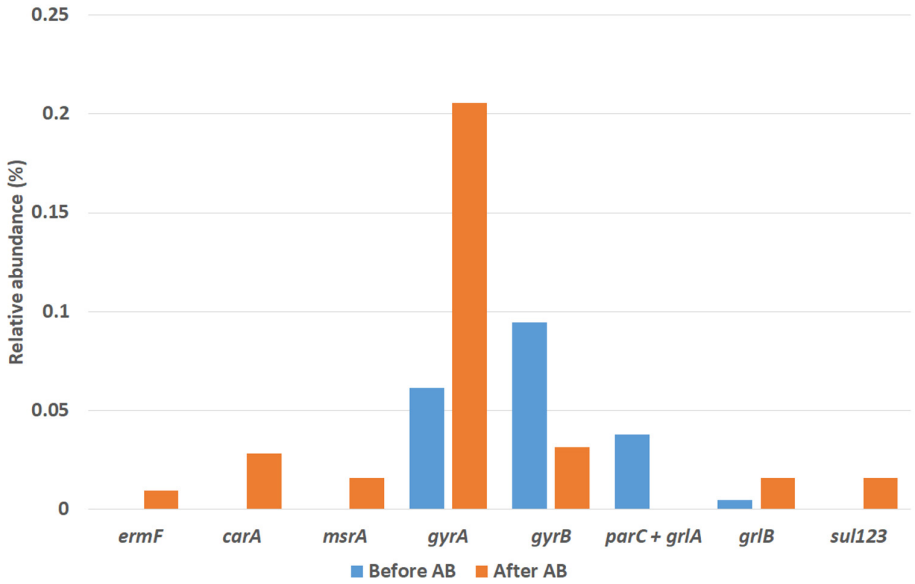


FIGURE 7 | Relative abundance of antibiotics resistance genes found before and after the antibiotics addition as determined by metatranscriptomic analysis.

gene and efflux-related genes *carA* and *msrA*. Similarly to the AZT case, sulfonamide-resistance genes *sul1*, *sul2* and *sul3* were found only after the addition of the antibiotics. On the other hand, the only resistance genes against fluoroquinolone found after the antibiotics addition were *gyrA*, *gyrB*, and *grlB*. On the other hand, the activity of genes *parC*+*grlA* was only present after the antibiotics addition.

In this sense, it is possible that loading of antibiotics could trigger the emergence of resistance to these compounds within the bacterial community structure of the partial nitritation biofilter by the means of mutation leading to formation of dihydropteroate synthase with lower affinity for this antibiotic for sulfonamides; functioning of efflux-related genes against AZT; and by target-site mutation of DNA gyrase subunit A *gyrA* and DNA topoisomerase subunit B *grlB* for fluoroquinolone.

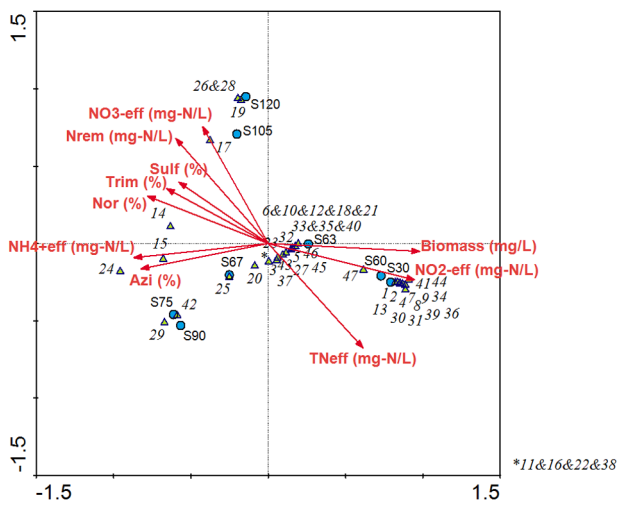


FIGURE 8 | Multivariate redundancy analysis linking the performance of the partial-nitritation biofilter with the bacterial community structure during the operation time. The numbers from 1 to 47 are correlated with the microbial population codes in the **Figure 3**.

Activity Related to Carbon Metabolism

Metatranscriptomics data showed that there was a slight increase in pyruvate kinase enzyme, related to glycolysis, with the addition of antibiotics, which could indicate the preference of glycolysis when the antibiotics were present in the bioreactor (Supplementary Table S5). Also, enzymes alcohol dehydrogenase and lactate dehydrogenase, which are crucial for the fermentation of organic matter, were much higher after the antibiotics addition (2.5-fold increase in alcohol dehydrogenase after antibiotics additions, while no lactate dehydrogenase was found before the antibiotics loading), which seemed to indicate that fermentation of organic matter was more important after the addition of the antibiotics. The relative abundance of Krebs cycle enzymes was higher after the antibiotics addition than before. In this sense, it is possible that the loading of antibiotics caused an increase of the importance in the heterotrophic metabolism. Thus, the growth of the biofilm as the system adapted to the antibiotics loading could be related to the growth of heterotrophs. These microorganisms could grow by degradation of the antibiotics compound or by consumption of biofilm EPS and cell material. In this sense the loading of 377 mg day-1 of antibiotics could support the growth of heterotrophs. Since the removal of AZT coincided temporally with the near-maximum biofilm biomass, it is possible that AZT biodegradation could support the growth of the biofilm. Previous research on partial nitritation biofilters subjected to ciprofloxacin concluded that biofilm growth after antibiotics addition could

be caused by ciprofloxacin degradation of dominant genus *Comamonas* (Gonzalez-Martinez et al., 2014b). The removal of high concentrations of norfloxacin, AZT and trimethoprim was related to an *Alcaligenes* strain (Rodriguez-Sanchez et al., 2017), which was taxonomically related to the dominant genus in the partial nitritation biofilter after the antibiotics addition. The data obtained in this research suggested that microbial community dynamics in bioreactors subjected to antibiotics could be related to heterotrophic metabolisms of antibiotics resistant bacteria, which could biodegrade the antibiotics. More research is necessary in order to fully unravel the mechanisms of biomass adaptation in bioreactors.

Multivariate Redundancy Analysis Linking the Bacterial Community Structure with the Operational Conditions in the Partial-Nitritation Biofilter

The multivariate redundancy analysis developed for the operation under antibiotics showed two groups of variables (**Figure 8**). Mainly, ammonium oxidation performance was negatively correlated with biomass concentration inside the system. In this sense, the observed loss of biomass after the antibiotics addition was strongly correlated with the loss in partial-nitritation performance. As shown by bacterial community dynamics, the loss of *Nitrosomonas* genus caused the decrease in effluent nitrite concentration after antibiotics loading to the biofilter. Genera *Alcaligenes* and *Acidovorax* showed a positive correlation with effluent nitrate concentration and nitrogen removal, which supports the nitrifying and denitrifying activities of both genera found by the metatranscriptomic analysis. *Brevundimonas*, *Paracoccus*, *Ochrobactrum*, and *Alcaligenes*, among others, were correlated with net antimicrobials elimination. These phylotypes were also found to be related to antibiotics removal in a CANON bioreactor subjected to an antibiotics mix of AZT, NOR, TMP, and SMZ (Rodriguez-Sanchez et al., 2017). These results showed that the same antibiotics-resistant genera proliferated in two different technologies with different biomass configurations and thus aim to the fact that the nature of the antibiotics compound was the most important driving factor for the bacterial communities in these bioreactors.

CONCLUSION

A partial-nitritation biofilter was subjected to a high antibiotics concentration influent. The addition of antibiotics reduced the partial-nitritation efficiency. This was caused by the loss of biomass, as showed by multivariate redundancy analysis, and by changes in the bacterial community structure, as shown by high-throughput sequencing of 16S rRNA gene of *Bacteria* domain. The changes in the bacterial communities after the addition of the antibiotics suggested the proliferation of *Alcaligenes*, *Acidovorax*, and *Paracoccus*. Metatranscriptomics analysis demonstrated that dominant bacterial genera after the antibiotics addition expressed proteins that develop heterotrophic nitrification and aerobic denitrification metabolism, which contrasted with the high activity of ammonium oxidation found in the biofilter after the antibiotics addition. The addition of antibiotics could be related to an increase of aerobic and anaerobic heterotrophic metabolisms. The antibiotics caused a decrease in the bacterial activity within the system but an increase in the activities of *Archaea* and *Fungi* phylotypes. In this sense, the presence of the antimicrobial compounds also impacted small players in the performance of the partial-nitritation biofilter. Resistance to AZT, sulfonamides and norfloxacin were related to *ermF*, *carA*, *msrA*, *sul123*, *gyrA*, and *grlB* genes. The coupled analysis using bioreactor's performance, bacterial community structure and metatranscriptomics analysis offered a complete comprehensive understanding of the influence of the antibiotics in the bioreactor analyzed during the experiment. The results showed that the addition of the antibiotics reduced the ammonium oxidation efficiency of the system and enhanced the nitrogen removal capacity, which was correlated with the decrease of autotrophic ammonium oxidizing bacteria, the proliferation of heterotrophs a decrease in ammonium oxidation activity and an increase in the denitrification activity inside the bioreactor. The results linked the performance of the bioreactor, the bacterial community dynamics and the microbial activity in the system in order to reach a complete approach for the evaluation of the antibiotics effect over the system, and will be valuable for the treatment of effluents with high antibiotics concentrations.

AUTHOR CONTRIBUTIONS

AG-M is the main researcher in the bioreactor, physico-chemical and molecular biology techniques. AM is the main researcher in micropollutants studies. AR-S has worked on bioinformatic analysis. CP has worked on molecular biology techniques. SD-C supervised the micropollutant studies and the manuscript preparation. DB has supervised the micropollutant studies and the whole research results. RV has supervised the engineering, molecular biology results, and the whole research results.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support given by the Built Environment Department in the Aalto University and the Institute of Water Research of the University of Granada. They also acknowledge the research funding from the Generalitat de Catalunya (Consolidated Research Group: Water and Soil Quality Unit 2014-SGR-418). Finally, they want to give thanks to B. San Miguel-Conejero for her microbiological support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00354/full#supplementary-material

REFERENCES

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., and Narasihman, G. (2016). Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evol. Bioinform.* 12(Suppl. 1), 5–16. doi: 10.4137/EBO.S36436

Albuquerque, L., Tiago, I., Rainey, F. A., Taborda, M., Nobre, M. F., and Costa, M. S. (2017). *Salirhabdus euzebyi* gen. nov., sp. nov., a Gram-positive, halotolerant bacterium isolated from a sea salt evaporation pond. *Int. J. Syst. Evol. Microbiol.* 57, 1566–1571. doi: 10.1099/ijs.0.64999-0

Aydin, S., Ince, B., and Ince, O. (2015). Development of antibiotic resistance genes in microbial communities during long-term operation of anaerobic reactors in the treatment of pharmaceutical wastewater. *Water Res.* 83, 337–344. doi: 10.1016/j.watres.2015.07.007

Botts, R. T., Apffel, B. A., Walters, C. J., Davidson, K. E., Echols, R. S., Geiger, M. R., et al. (2017). Characterization of four multidrug resistance plasmids captured from the sediments of an urban coastal wetland. *Front. Microbiol.* 8:1922. doi: 10.3389/fmicb.2017.01922

Cao, H., Phe, K., Laine, G. A., Russo, H. R., Putney, K. S., and Tam, V. H. (2016). An institutional review of antimicrobial stewardship interventions. *J. Glob. Antimicrob. Resist.* 6, 75–77. doi: 10.1016/j.jgar.2016.03.006

Chen, J., Gu, S., Hao, H., and Chen, J. (2016a). Characteristics and metabolic pathway of *Alcaligenes* sp. TB for simultaneous heterotrophic nitrification-aerobic denitrification. *Appl. Microbiol. Biotechnol.* 100, 9787–9794. doi: 10.1007/s00253-016-7840-x

Chen, J., Han, Y., Wang, Y., Gong, B., Zhou, J., and Qing, X. (2016b). Start-up and microbial communities of a simultaneous nitrogen removal system for high salinity and high nitrogen organic wastewater via heterotrophic nitrification. *Bioresour. Technol.* 2016, 196–202. doi: 10.1016/j.biortech.2016.05.064

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Eramo, A., Delos Reyes, H., and Fahrenfeld, N. H. (2017). Partitioning of antibiotic resistance genes and fecal indicators varies intra and inter-storm during combined sewer overflows. *Front. Microbiol.* 8:2024. doi: 10.3389/fmicb.2017.02024

FitzGerald, J. A., Allen, E., Wall, D. M., Jackson, S. A., Jerry, D. M., and Dobson, A. D. W. (2015). *Methanosarcina* play an important role in anaerobic co-digestion of the seaweed *Ulva lactuca*: taxonomy and predicted metabolism of functional microbial communities. *PLoS One* 10:e0142603. doi: 10.1371/journal.pone.0142603

Gonzalez-Martinez, A., Pesciaroli, C., Martinez-Toledo, M. V., Hontoria, E., Gonzalez-Lopez, J., and Osorio, F. (2014a). Study of nitrifying microbial communities in a partial-nitritation bioreactor. *Ecol. Eng.* 64, 443–450. doi: 10.1016/j.ecoleng.2014.01.009

Gonzalez-Martinez, A., Rodriguez-Sanchez, A., Garcia-Ruiz, M. J., Osorio, F., and Gonzalez-Lopez, J. (2016a). Impact of methionine on a partial-nitritation biofilter. *Environ. Sci. Pollut. Res.* 23, 6651–6660. doi: 10.1007/s11356-015-5889-1

Gonzalez-Martinez, A., Rodriguez-Sanchez, A., Lotti, T., Garcia-Ruiz, M. J., Osorio, F., Gonzalez-Lopez, J., et al. (2016b). Comparison of bacterial communities of conventional and A-stage activated sludge systems. *Sci. Rep.* 6:18786. doi: 10.1038/srep18786

Gonzalez-Martinez, A., Rodriguez-Sanchez, A., Martinez-Toledo, M. V., Garcia-Ruiz, M. J., Hontoria, E., Osorio-Robles, F., et al. (2014b). Effect of ciprofloxacin antibiotic on the partial-nitritation process and bacterial community structure of a submerged biofilter. *Sci. Total Environ.* 47, 276–287. doi: 10.1016/j.scitotenv.2014.01.012

Gonzalez-Martinez, A., Rodriguez-Sanchez, A., Rodelas, B., Abbas, B. A., Martinez-Toledo, M. V., van Loosdrecht, M. C. M., et al. (2015). 454-pyrosequencing analysis of bacterial communities from autotrophic nitrogen removal bioreactors utilizing universal primers: effect of annealing temperature. *BioMed Res. Int.* 892013, 1–12. doi: 10.1155/2015/892013

Guo, J., Peng, Y., Wang, S., Ma, B., Ge, S., Wang, Z., et al. (2013). Pathways and organisms involved in ammonia oxidation and nitrous oxide emission. *Crit. Rev. Environ. Sci. Technol.* 43, 37–41. doi: 10.1080/10643389.2012.672072

Hsu, Y. H., Lai, W. A., Lin, S. Y., Hameed, A., Shahina, M., Shen, F. T., et al. (2013). *Chiayiivirga flava* gen. nov., sp. nov., a novel bacterium of the family *Xanthomonadaceae* isolated from an agricultural soil, and emended description of the genus *Dokdonella*. *Int. J. Syst. Evol. Microbiol.* 63, 3293–3300. doi: 10.1099/ijs.0.048579-0

Huang, X., Zhen, J., Liu, C., Liu, L., Liu, Y., and Fan, H. (2017). Removal of antibiotics and resistance genes from swine wastewater using vertical flow constructed wetlands: effect of hydraulic flow direction and substrate type. *Chem. Eng. J.* 308, 692–699. doi: 10.1016/j.cej.2016.09.110

Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OUT clustering. *Environ. Microbiol.* 12, 1889–1898. doi: 10.1111/j.1462-2920.2010.02193.x

Ivanová, L., Fáberová, M., Macku'ak, T., Grabic, R., and Bodík, I. (2017). Estimation of amount of selected pharmaceuticals sorbed onto digested sludge from wastewater treatment plant Bratislava-Petržalka. *Environ. Res.* 2017, 31–35. doi: 10.1016/j.envres.2017.02.001

Ji, B., Yang, K., Zhu, L., Jiang, Y., Wang, H., Zhou, J., et al. (2015). Aerobic denitrification: a review of important advances in the last 30 years. *Biotechnol. Bioprocess Eng.* 20, 643–651. doi: 10.1007/s12257-015-0009-0

Lawson, C. E., Wu, S., Bhattacharjee, A. S., Hamilton, J. J., McMahon, K. D., Goel, R., et al. (2017). Metabolic network analysis reveals microbial community interactions in anammox granules. *Nat. Commun.* 8, 1–12. doi: 10.1038/ncomms15416

Leimena, M. M., Ramiro-Garcia, J., Davids, M., van den Bogert, B., Smidt, H., Smid, E. J., et al. (2013). A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genomics* 14:530. doi: 10.1186/1471-2164-14-530

Lloyd, K. G. (2013). Predominant archaea in marine sediments degrade detrital proteins. *Nature* 496, 215–218. doi: 10.1038/nature12033

Lu, H., Chandran, K., and Stensel, D. (2014). Microbial ecology of denitrification in biological wastewater treatment. *Water Res.* 64, 237–254. doi: 10.1016/j.watres.2014.06.042

Massé, D. I., Saady, N. M. C., and Gilbert, Y. (2014). Potential of biological processes to eliminate antibiotics in livestock manure: an overview. *Animals* 4, 146–163. doi: 10.3390/ani4020146

Meng, F., Gao, G., Yang, T. T., Chen, X., Chao, Y., Na, G., et al. (2015). Effects of fluoroquinolone antibiotics on reactor performance and microbial community structure of a membrane bioreactor. *Chem. Eng. J.* 280, 448–458. doi: 10.1016/j.cej.2015b.06.025

Meng, L. W., Li, X. K., Wang, K., Ma, K. L., and Zhang, J. (2015). Influence of the amoxicillin concentration on organics removal and microbial community structure in an anaerobic EGSB reactor treating with antibiotic wastewater. *Chem. Eng. J.* 274, 94–101. doi: 10.1016/j.cej.2015.03.065

Miller, J. H., Novak, J. T., Knocke, W. R., and Pruden, A. (2016). Survival of antibiotic resistant bacteria and horizontal gene transfer control antibiotic resistance gene content in anaerobic digesters. *Front. Microbiol.* 7:263. doi: 10.3389/fmicb.2016.00263

Mobberley, J. M., Khodadad, C. L. M., Visscher, P. T., Reid, R. P., Hagan, P., and Foster, J. S. (2015). Inner workings of thrombolites: spatial gradients of metabolic activity as revealed by metatranscriptome profiling. *Nat. Sci. Rep.* 5:12601. doi: 10.1038/srep12601

Nasir, I. M., Ghazi, T. I. M., and Omar, R. (2012). Anaerobic digestion technology in livestock manure treatment for biogas production: a review. *Eng. Life Sci.* 12, 258–269. doi: 10.1002/elsc.201100150

Nunora, T., Hirai, M., Miyazaki, M., Kazama, H., Makita, H., Hirayama, H., et al. (2013). Isolation and characterization of a thermophilic, obligately anaerobic and heterotrophic marine *Chloroflexi* bacterium from a *Chloroflexi*-dominated microbial community associated with a Japanese shallow hydrothermal system, and proposal for *Thermomarinilinea lacunofontalis* gen. nov., sp. Nov. *Microbes Environ.* 28, 228–235. doi: 10.1264/jsme2.ME12193

Oosterkamp, M. J., Veuskens, T., Saia, F. T., Weelink, S. A. B., Goodwin, L. A., Daligault, H. E., et al. (2013). Genome analysis and physiological comparison of *Alicycliphilus denitrificans* strains BC and K601^T. *PLoS One* 8:0066971. doi: 10.1371/journal.pone.0066971

Qu, Z., Bakken, L. R., Molstad, L., Frostegård, Å, and Bergaust, L. L. (2016). Transcriptional and metabolic regulation of denitrification in *Paracoccus denitrificans* allows low but significant activity of nitrous oxide reductase under oxic conditions. *Environ. Microbiol.* 18, 2951–2963. doi: 10.1111/1462-2920.13128

Redgrave, L. S., Sutton, S. B., Webber, M. A., and Piddock, L. J. V. (2014). Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success. *Trends Microbiol.* 22, 438–445. doi: 10.1016/j.tim.2014.04.007

Roberts, C. M. (2008). Update on macrolide-lincosamide-streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS Microbiol. Lett.* 282, 147–159. doi: 10.1111/j.1574-6968.2008.01145.x

Rodriguez, E., Garcia-Encina, P. A., Stams, A. J. M., Maphosa, F., and Sousa, D. Z. (2015). Meta-omics approaches to understand and improve wastewater treatment systems. *Rev. Environ. Sci. Biotechnol.* 14, 385–406. doi: 10.1007/s11157-015-9370-x

Rodriguez-R, L. M., and Konstantinidis, K. T. (2014a). Estimating coverage in metagenomic data sets and why it matters. *ISME J.* 8, 2349–2351. doi: 10.1038/ismej.2014.76

Rodriguez-R, L. M., and Konstantinidis, K. T. (2014b). Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* 30, 629–635. doi: 10.1093/bioinformatics/btt584

Rodriguez-Sanchez, A., Margareto, A., Robledo-Mahon, T., Aranda, E., Diaz-Cruz, S., Gonzalez-Lopez, J., et al. (2017). Performance and bacterial community structure of a granular autotrophic nitrogen removal bioreactor amended with high antibiotic concentrations. *Chem. Eng. J.* 325, 257–269. doi: 10.1016/j.cej.2017.05.078

Rodriguez-Sanchez, A., Mikola, A., Munoz-Palazon, B., Vahala, R., Mu, B., and Gonzalez-Martinez, A. (2016a). Performance and bacterial community structure of a submerged biofilter subjected to high ammonium and high organic carbon concentrations. *Int. Biodeterior. Biodegradation* 115, 224–233. doi: 10.1016/j.ibiod.2016.09.001

Rodriguez-Sanchez, A., Munoz-Palazon, B., Maza-Marquez, P., Gonzalez-Lopez, J., and Gonzalez-Martinez, A. (2016b). Process performance and bacterial community dynamics of partial-nitritation biofilters subjected to different concentrations of cysteine amino acid. *Biotechnol. Prog.* 32, 1254–1263. doi: 10.1002/btpr.2331

Saito, K., Kobayashi, K., Wada, M., Kikuno, I., Takusagawa, A., Mochizuki, M., et al. (2010). Omnipotent role of archaeal elongation factor 1 alpha (EF1 α) in translational elongation and termination, and quality control of protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1–6. doi: 10.1073/pnas.1009599107

Sayadi, A., Immonen, E., Bayram, H., and Arnqvist, G. (2016). The de novo transcriptome and its functional annotation in the seed beetle *Callosobruchus maculatus*. *PLoS One* 11:e0158565. doi: 10.1371/journal.pone.0158565

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Tsubouchi, T., Koyama, S., Mori, K., Shimane, Y., Usui, K., Tokuda, M., et al. (2017). *Brevundimonas denitrificans* sp. nov., a denitrifying bacterium isolated from deep subseafloor sediment. *Int. J. Syst. Evol. Microbiol.* 64(pt 11), 3709–3716. doi: 10.1099/ijs.0.067199-0

van der Star, W. R. L., Abma, W. R., Blommers, D., Mulder, J. W., Tokutomi, T., Strous, M., et al. (2007). Startup of reactors for anoxic ammonium oxidation: experiences from the first full-scale anammox reactor in Rotterdam. *Water Res.* 41, 4149–4163. doi: 10.1016/j.watres.2007.03.044

van Teeseling, M. C. F., Maresch, D., Rath, C. B., Figl, R., Altmann, F., Jetten, M. S. M., et al. (2016). The S-layer protein of the anammox bacterium *Kuenenia stuttgartiensis* is heavily *O-glycosylated*. *Front. Microbiol.* 7:1721. doi: 10.3389/fmicb.2016.01721

Wang, J., Mao, D., Mu, Q., and Luo, Y. (2015). Science of the total environment fate and proliferation of typical antibiotic resistance genes in five full-scale pharmaceutical wastewater treatment plants. *Sci. Total Environ.* 526, 366–373. doi: 10.1016/j.scitotenv.2015.05.046

Wang, N., Yang, X., Jiao, S., Zhang, J., Ye, B., and Gao, S. (2014). Sulfonamide-resistant bacteria and their resistance genes in soils fertilized with manures from Jiangsu Province. Southeastern China. *PLoS One* 9:e112626. doi: 10.1371/journal.pone.0112626

Yan, Q., Gao, X., Chen, Y. P., Peng, X. Y., Zhang, Y. X., Gan, X. M., et al. (2014a). Occurrence, fate and ecotoxicological assessment of pharmaceutically active compounds in wastewater and sludge from wastewater treatment plants in Chongqing, the Three Gorges Reservoir Area. *Sci. Total Environ.* 47, 618–630. doi: 10.1016/j.scitotenv.2013.09.032

Yan, Q., Gao, X., Huang, L., Gan, X. M., Zhang, Y. X., Chen, Y. P., et al. (2014b). Occurrence and fate of pharmaceutically active compounds in the largest municipal wastewater treatment plant in Southwest China: mass balance analysis and consumption back-calculated model. *Chemosphere* 99, 160–170. doi: 10.1016/j.chemosphere.2013.10.062

Yang, L., Ren, Y. X., Zhao, S. Q., Liang, X., and Wang, J. P. (2016). Isolation and characterization of three heterotrophic nitrifying-aerobic denitrifying bacteria from a sequencing batch reactor. *Ann. Microbiol.* 66, 737–747. doi: 10.1007/s13213-015-1161-7

Yu, W. C., Zhan, S. H., Shen, Z. Q., Zhou, Q. X., and Yang, D. (2016). Efficient removal mechanism for antibiotic resistance genes from aquatic environments by graphene oxide nanosheet. *Chem. Eng. J.* 313, 836–846. doi: 10.1016/j.cej.2016.10.107

Zhang, Z., Esther, M. M., Bunker, R. D., Baker, E. N., and Squire, C. J. (2009). Structure and function of GlmU from *Mycobacterium tuberculosis*. *Acta Crystallogr. D Biol. Crytalogr.* D65(Pt 3), 275–283. doi: 10.1107/S0907444909001036

Zhu, Y., Wang, Y., Jiang, X., Zhou, S., Wu, M., Pan, M., et al. (2017). Microbial community analysis for membrane bioreactor treating antibiotics containing wastewater. *Chem. Eng. J.* 325, 300–309. doi: 10.1016/j.cej.2017.05.073

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gonzalez-Martinez, Margareto, Rodriguez-Sanchez, Pesciaroli, Diaz-Cruz, Barcelo and Vahala. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 October 2017
doi: 10.3389/fmicb.2017.02052

Characterization of Bacterial and Fungal Community Dynamics by High-Throughput Sequencing (HTS) Metabarcoding during Flax Dew-Retting

*Christophe Djemiel†, Sébastien Grec† and Simon Hawkins**

Univ. Lille, Centre National de la Recherche Scientifique, UMR 8576 - Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Alessio Mengoni, University of Florence, Italy
Ulas Karaoz, University of California, Berkeley, United States

***Correspondence:**
Simon Hawkins
simon.hawkins@univ-lille1.fr

†Christophe Djemiel
orcid.org/0000-0002-5659-7876
†Sébastien Grec
orcid.org/0000-0003-4143-4035

Specialty section:
This article was submitted to Terrestrial Microbiology, a section of the journal Frontiers in Microbiology

Received: *10 July 2017*
Accepted: *06 October 2017*
Published: *20 October 2017*

Citation:
Djemiel C, Grec S and Hawkins S (2017) Characterization of Bacterial and Fungal Community Dynamics by High-Throughput Sequencing (HTS) Metabarcoding during Flax Dew-Retting. Front. Microbiol. 8:2052. doi: 10.3389/fmicb.2017.02052

Flax dew-retting is a key step in the industrial extraction of fibers from flax stems and is dependent upon the production of a battery of hydrolytic enzymes produced by micro-organisms during this process. To explore the diversity and dynamics of bacterial and fungal communities involved in this process we applied a high-throughput sequencing (HTS) DNA metabarcoding approach (16S rRNA/ITS region, Illumina Miseq) on plant and soil samples obtained over a period of 7 weeks in July and August 2014. Twenty-three bacterial and six fungal phyla were identified in soil samples and 11 bacterial and four fungal phyla in plant samples. Dominant phyla were Proteobacteria, Bacteroidetes, Actinobacteria, and Firmicutes (bacteria) and Ascomycota, Basidiomycota, and Zygomycota (fungi) all of which have been previously associated with flax dew-retting except for Bacteroidetes and Basidiomycota that were identified for the first time. Rare phyla also identified for the first time in this process included Acidobacteria, CKC4, Chlorobi, Fibrobacteres, Gemmatimonadetes, Nitrospirae and TM6 (bacteria), and Chytridiomycota (fungi). No differences in microbial communities and colonization dynamics were observed between early and standard flax harvests. In contrast, the common agricultural practice of swath turning affects both bacterial and fungal community membership and structure in straw samples and may contribute to a more uniform retting. Prediction of community function using PICRUSt indicated the presence of a large collection of potential bacterial enzymes capable of hydrolyzing backbones and side-chains of cell wall polysaccharides. Assignment of functional guild (functional group) using FUNGuild software highlighted a change from parasitic to saprophytic trophic modes in fungi during retting. This work provides the first exhaustive description of the microbial communities involved in flax dew-retting and will provide a valuable benchmark in future studies aiming to evaluate the effects of other parameters (e.g., year-to year and site variability etc.) on this complex process.

Keywords: flax dew-retting, bacterial and fungal microbiota dynamics, 16S rRNA and ITS amplicons, metabarcoding, high-throughput sequencing, HTS, CAZyme predictions, trophic modes

BACKGROUND

Land plants fix ∼123 billion tons of carbon per year (Beer et al., 2010) of which an important part becomes channeled into the production of lignocellulosic biomass in plant cell walls (Kuhad and Singh, 1993; Boerjan et al., 2003; Zhou et al., 2011). Soil microflora function as key decomposers in various ecosystems (Soliveres et al., 2016) and are able to degrade this biomass, consisting of lignin and polysaccharide polymers such as cellulose, hemicelluloses, and pectins, by producing a set of synergistically acting hydrolytic enzymes (Warren, 1996; Lynd et al., 2002; Kubicek et al., 2014; Cragg et al., 2015).

During this process, monosaccharides are released and used by microorganisms for energy production thereby contributing to maintenance of the carbon cycle. The microbial diversity associated with this biomass degradation can vary depending on plant cell wall structure and the stage of the decomposition (Akin, 2008; Ventorino et al., 2015; Montella et al., 2017). Microbial dynamics can also vary depending on site location, soil composition, plant species, and biomass architecture (Schneider et al., 2012; Voříšková and Baldrian, 2013; Cardenas et al., 2015; Ventorino et al., 2015). Several investigations have reported that fungal communities change during leaf (e.g., beech, oak, maize) litter decay with an initial predominance of species assigned to the Ascomycota phylum, replaced gradually by Basidyomycota (Schneider et al., 2012; Kuramae et al., 2013; Voříšková and Baldrian, 2013). Bacterial dynamics generally involve changes in the relative proportions of species assigned to Proteobacteria, Actinobacteria, and Bacteroidetes depending on sampling location and wood species to be degraded (Ventorino et al., 2015). The composition and succession of different microbial communities is presumably related to their capacity to degrade and utilize biomass present at a given moment. In this context it is interesting to note that particular microbial CAZymes such as endo- and exo-cellulases, xylanases, pectinases, and peroxidases have been associated with specific ecological groups during plant cell wall decomposition (Eastwood et al., 2011; Zhao et al., 2014; Ventorino et al., 2015).

In this study we investigate the microflora associated with a particularly interesting, and ancient example of human exploitation of microbial lignocellulose degradation known as retting that is believed to date back to the Upper Paleolithic and/or Neolithic (Gübitz and Cavaco-Paulo, 2001; Kvavadze et al., 2009). This process is still used today and constitutes the first step in the industrial separation of long bast fibers from the stems of different fiber species such as flax, hemp, jute, and kenaf (Md. Tahir et al., 2011) used for textiles and composites (Campilho, 2015; Pil et al., 2016). During retting, bast fiber bundles become progressively separated from the surrounding stem tissues, and inter-fiber cohesion is reduced via the action of hydrolytic enzymes produced by straw-colonizing microorganisms (Rosemberg, 1965; Zhang et al., 2005; Md. Tahir et al., 2011; Akin, 2013; Preisner et al., 2014). Retting is performed by either leaving plants on the soil (dew- or field-retting), or by placing them in ponds, rivers, or water tanks (water-retting). Although good quality fibers are produced by water retting this process is more labor intensive and associated with extensive water pollution. Currently the majority of the world's flax fiber is produced by dew-retting (Akin, 2013; Preisner et al., 2014).

The main challenge during retting is to facilitate fiber decohesion without degrading cellulosic fibers by over-retting (Brown and Sharma, 1984; Akin et al., 1998; Henriksson et al., 1999). Since this process relies on enzymes produced by colonizing microorganisms, a better knowledge of the different groups/species involved should enable a greater understanding and control of this complex process. Although various bacteria and fungi have been identified in a number of different studies by using isolation and culturing approaches (Sharma, 1986a; Henriksson et al., 1997), such a strategy is not powerful enough to obtain a complete inventory of the microorganisms present as only a small percentage of taxa can be successfully cultured under laboratory conditions (Staley and Konopka, 1985; Amann et al., 1995). More recently, molecular tools such as 16S rRNA gene amplification were used to identify new bacteria during bamboo, hemp, and flax retting (Tamburini et al., 2003; Fu et al., 2011; Ribeiro et al., 2015) and 18S rRNA gene amplification was used to identify fungi during hemp retting (Ribeiro et al., 2015). Nevertheless, these approaches are unable to generate an exhaustive inventory of the retting microbiome.

Over the last decade, microbial ecology studies have greatly benefited from the use of high throughput sequencing (HTS) technologies that can produce an exhaustive inventory of bacteria and fungi from complex samples such as soil, litter compost, rumen, and the midgut of cellulosic-feeding insects via targeted-metagenomics (Hirsch et al., 2010; Ihrmark et al., 2012; Suenaga, 2012). These approaches were also used to study plant-microbe interactions (Knief, 2014; Peršoh, 2015) in rhizospheres or endospheres (Lundberg et al., 2012; Bodenhausen et al., 2013; Beckers et al., 2017). Only two studies have reported the use of HTS technologies (Ion Torrent PGM system) on kenaf retting (Visi et al., 2013) and more recently during water-retting of flax (Zhao et al., 2016). In both studies, the microbial community analysis was limited to bacterial domain, despite the importance of fungal taxa in the production of extracellular hydrolytic enzymes (Schneider et al., 2012). In this work, we report the first exhaustive HTS microbial inventory focusing on both bacterial and fungal communities using rRNA amplicon sequencing during dew-retting of flax.

METHODS

Experimental Design–Study Site–Sampling

Flax plants (*Linum usitatissimum* L., Cultivar Lorea) were sown on 14 March 2014 near Martainneville (F-27210 Region Hauts-de-France) in the north of France (50°00′03″N and 1°42′27″E). Plants were cultivated and retted on a typical silt loam soil with a neutral/slightly acid pH (INRA Soil Analysis Laboratory, LAS, Arras, France, http://www.lille.inra.fr/las) (Supplementary Table 1). Climatic data during the retting period was obtained from the Abbeville meteorological station at 10 km from Martainneville (infoclimat: http://www.infoclimat.fr/observations-meteo/temps-reel/abbeville/07005.html) (**Supplementary Figure 1**). "Early" and "standard" flax cultures were pulled (up-rooted) on the 16.07.2014 and 24.07.2014, respectively and dew-retted

in the field until the 25.08.2014 (early cultures) and 05.09.2014 (standard cultures). Replicate straw (plant) and soil samples were collected at regular intervals (R0–R6) during retting from five different locations in the retting field chosen according to a non-systematic W pattern as previously described (Plassart et al., 2012) and shown in **Supplementary Figure 2**. For straw samples, the middle region (30 cm long × total swath height) of the swath was collected; for soil samples, cores (20 cm deep × 8 cm diameter) were used. Stem samples were directly stored at −20°C and soil samples were sieved (pore size <2.0 mm), homogenized and freeze-dried before storage at −80°C.

DNA Extraction

DNA was extracted from 1 g sample using the GnS-GII (Plassart et al., 2012; Terrat et al., 2012, 2015). Briefly, samples were ground in 15 ml Falcon tubes containing a bead mix (ceramic, silica, and glass) and lysis buffer (100 mM Tris-HCl, pH 8; 100 mM EDTA, pH 8; 100 mM NaCl, 2% w/v and sodium dodecyl sulfate, 2% w/v) in a FastPrep ®-24 (MP-Biomedicals, NY, USA) (3 × 30 s at 4, 000 s^{-1} shaking). Proteins were precipitated by adding 100 µl of KAc (3 M) and nucleic acids recovered by isopropanol precipitation and washed with 70% ethanol, before drying and re-suspension in 100 µl water.

DNA Purification, Quantification, and Normalization

DNA extracts were filtered through PVPP (PolyVinylPolyPyrrolidone) Micro Bio-Spin® Columns with Bio-Gel® P-6 (Bio-Rad) by a 4 min at 1,000 g, 10°C centrifugation. Collected samples were then purified using the Geneclean Turbo kit (MP-Biomedicals, NY, USA) following the manufacturer's instructions. DNA was quantified on a LightCycler 480 System (Roche) using the Quant-iT™ PicoGreen® dsDNA Assay kit (Invitrogen). Samples were normalized to a concentration of 5 ng/µl and the DNA from the five replicates pooled using the epMotion® 5075 TMX (eppendorf). Altogether, 16 soil samples and 14 stem samples were recovered for further analysis.

Primers, PCR Amplification, and Sequencing

Bacterial 16S rDNA were amplified using the forward primer S-D-Bact-0341-a-S-17 described by Klindworth (Klindworth et al., 2013) coupled with a customized reverse primer S-D-Bact-0787-a-A-19, based on the 786r primer (Gołębiewski et al., 2014). Fungal ITS regions were amplified using the fITS7 forward primer 5.8S (Ihrmark et al., 2012) and the reverse primer ITS4_KYO1 (Toju et al., 2012; Bokulich and Mills, 2013). All primer sequences are given in Supplementary Table 2.

Amplifications were carried out in a total volume of 40 µl using 5 ng of DNA, 4 µl of 5x HOT FIREPol ® Blend Master Mix with 7.5 mM $MgCl_2$ (Solis Biodyne, Tartu, Estonia), 0.8 µl (0.2 µM) of each primer. PCR1 conditions were: 15 min at 95°C, followed by 30 cycles of 20 s at 95°C, 30 s at 53°C, and 20 s at 72°C, and final elongation for 5 min at 72°C. Single multiplexing was performed using home-made 6 bp indexes that were added to reverse primer during a second PCR2 of 12 cycles using indexed primers. The resulting PCR2 products were purified by HighPrep™ PCR (Magbio) clean-up system as described by the manufacturer, pooled and loaded onto the Illumina MiSeq cartridge according to the manufacturer instructions for a 2 × 250 bp paired-end sequencing on the GeT-PlaGe Genotoul Platform (INRA Castanet Tolosan, France). The quality of the run was checked internally using PhiX, and then each pair-end sequence was assigned to its sample with the help of the previously integrated index.

Sequence Processing

A bioinformatic pipeline based on mothur v.1.37.4 (https://github.com/mothur/mothur/releases) (Schloss et al., 2009) was configured to process the bacterial 16S rRNA gene sequences. This pipeline uses the standard Schloss lab operating procedure (http://www.mothur.org/wiki/MiSeq_SOP). Pair-End (PE) FASTQ files were overlapped to form contiguous reads in a single FASTA file with zero differences to the primer sequence and a quality score threshold of 30. Sequences with the following characteristics were removed: ambiguous bases and mismatches, <300/> 500 bp, homopolymers >8 bp, overlap <30 bp. Bacterial sequences were aligned against both SILVA (SSU SILVA 123) and Greengenes (August 2013 release, for input PICRUSt) reference databases. A pre-clustering was done to reduce noise as recommended (Pruesse et al., 2007; Huse et al., 2010) allowing for up to four differences between sequences. Chimeras were detected and removed *de novo* with the UCHIME (version 4.2) algorithm (Edgar et al., 2011). The clustering of the non-chimeric sequences to Operational Taxonomic Units (OTUs) was done by *de novo* clustering at 0.03 cut-off of dissimilarity using neighbor based on genomic distance matrix. Finally, a general count sequence table for each OTU of all samples was generated to obtain the consensus taxonomy based on the Ribosomal Database Project's naïve bayesian classifier method (Wang et al., 2007) and for the future OTU-based analysis.

For processing ITS2 from fungal ribosomal ITS sequences, the recently described PIPITS v.1.3.3 pipeline (https://github.com/hsgweon/pipits/releases) was used (Gweon et al., 2015). Raw reads were prepared for ITS extraction and the chosen sub-regions extracted with the ITSx software tool (Bengtsson-Palme et al., 2013) before clustering and taxonomic assignation using the UNITE database (version 31.01.2016) (Abarenkov et al., 2010).

All parameters, algorithms and tools for the bioinformatic steps used in the two pipelines are given in Supplementary Table 3.

The microbial DNA sequencing data sets supporting the results in this article are available at the EBI ENA with accession number PRJEB20299.

Statistical Analysis

All estimators used to measure the α-diversity and β-diversity were calculated applying mothur procedures following recommendations and parameters suggested by tutorials (Kozich et al., 2013).

Alpha-diversity was estimated with the chao1 non-parametric estimator (Chao, 1984) and evenness was measured with Heip's

estimator [$E_{heip} = (e^{H'}-1)/(S-1)$ with H' being Shannon's diversity index and S the number of species] (Heip, 1974). Community diversity was estimated with Shannon's diversity index (Ludwig and Reynolds, 1988) and the inverse Simpson's index (Simpson, 1949). Microbial community coverage was tested by calculating the Good's non-parametric coverage estimator (Good, 1953; Esty, 1986) and verified by rarefaction curves. Differences in alpha diversities were evaluated using the Mann-Whitney-Wilcoxon test.

Beta-diversity was assessed using the Yue and Clayton theta similarity coefficient for community structure and the Jaccard index for community membership (Yue and Clayton, 2005; Barwell et al., 2015).

The non-parametric analysis of molecular variance (AMOVA) (Excoffier et al., 1992) was used to examine the significance of differences between and within different groups (Early vs. Standard and Before vs. After turning swaths) with a p-value ≤ 0.05 being considered as statistically significant.

The diversity indices are computed from a standardized file containing the count of OTUs for each sample.

Spearman rank correlation coefficients were calculated from generated dissimilarity matrices to look for any significant correlations between climatic conditions (temperature and rainfall) and bacterial and fungal community structure.

For population level analyses, several tools were used: Metastats (White et al., 2009) (White et al., 2009), LEfSe (Linear discriminant analysis Effect Size (Segata et al., 2011), and Indicator (from Mothur software).

The PICRUSt v.1.0.0 (https://github.com/picrust/picrust/releases) pipeline (http://picrust.github.io/picrust/) (Langille et al., 2013) was used to predict the functional composition of bacterial enzymatic activity abundance using 16S rDNA datasets. An OTUs table (input file) in BIOM format was generated using Mothur and then reference picked against the Greengenes database. Accuracy of metagenome predictions was controlled by measuring the weighted Nearest Sequenced Taxon Index (NSTI) scores that reflect the availability of reference genomes closely related to the most abundant microorganisms for each sample. To analyze the Carbohydrate Active enZymes (CAZymes) prediction, a pre-calculated table was used (https://sourceforge.net/projects/picrust/files/precalculated_files/).

The FUNGuild v1.0 database (https://github.com/UMNFuN/FUNGuild) was used to assign ecological functions (trophic modes) to each OTUs (Nguyen et al., 2016).

Graphic representations were produced using handmade scripts and based on Highcharts facilities (http://www.highcharts.com/) and jvenn plug-in (Bardou et al., 2014).

RESULTS

Metabarcoding and Sequencing

16S rDNA (bacterial) and ribosomal ITS (fungal) amplicons were sequenced using the Illumina MiSeq system. Redesigned primers (Supplementary Table 2) were used in order to avoid potential amplification of plant chloroplastic/mitochondrial DNA. Sequencing generated a very large data set ranging between 103,859 and 279,553 (average = 162,390 ± 33,164) bacterial raw sequences and between 187,055 and 483,703 (average = 285,070 ± 62,420) fungal raw sequences (Supplementary Tables 4A,B). OTU tables listing all OTUs detected and their abundance normalized by a subsampling are given in Supplementary Table 5 (bacteria) and Supplementary Table 6 (fungi).

Community Coverage and Diversity

To estimate how representative our samples were of the bacterial and fungal communities Good's coverage estimator was calculated for all samples (Supplementary Tables 6, 7). For bacterial samples Good's coverage values were greater than 99% for all straw samples and between 91 and 92% for soil samples indicating: (i) the high coverage of the sampling community and (ii) that the redesigned reverse primer did not significantly affect V3-V4 bacterial amplification (Supplementary Table 6). For fungal samples, Good's coverage estimators were above 99% for all samples confirming that the population is well-sampled (Supplementary Table 7). These results indicate that the sequencing depth used provides an accurate view of microbial community diversity and were also confirmed by the rarefaction curves (**Supplementary Figures 3**,**4**).

To analyze community diversity (alpha diversity) within our microbial samples we calculated Chao1 (species richness), Heip's (species evenness), and Inverse Simpson index metric estimators (**Supplementary Figures 5**, Supplementary Tables 7–9). For all estimators and all conditions [soil vs. plant (straw), early vs. standard cultures] community diversity was always higher in bacterial samples when compared to fungal samples. For both bacterial and fungal samples all indicators indicated that community diversity was higher in soil samples when compared to plant samples. In contrast, the same indicators revealed no difference in bacterial community diversity (both soil and plant samples) between early vs. standard cultures. For fungal communities, the situation was more complex. While no differences in species richness (Chao1) were observed between early vs. standard plant samples, Heip's estimator values suggested differences in sample evenness. The Inverse Simpson index values also suggested differences in community diversity between these two samples.

Examination of indicator values during the retting period (R0–R6) revealed a range of different profiles suggesting that sample community diversity evolves during this process (**Supplementary Figure 5**). When only the Inverse Simpson index is taken into account as an overall measure of community diversity (Supplementary Tables 7, 8) all profile types show an overall bimodal form with a peak/trough mainly occurring at R2 (5/8 profiles), but also at R3 (2/8 profiles) and R1 (1/8 profiles). Taken together, these results would suggest that sample community diversity changes (increases or decreases) at some point after R2 (and/or R3). The R2/R3 points are close to the moment when the stem swathes were turned and the observed change in community diversity values might be related to this process. When "early" and "standard" culture sample values are pooled (to provide sufficient data points) analyses shows that there is a significant effect (p-value <0.05) of swath turning on bacterial, but not fungal community diversity.

Comparison of soil Chao1, Heip's and Inverse Simpson estimators for the first retting point (R0) with those obtained at the sowing stage (R-1) (**Supplementary Figure 5**) show that both bacterial and fungal community diversity are always lower at R-1. Such an observation suggests that flax plants modify soil diversity through either a rhizosphere effect and/or the input of other organic material (e.g., leaves). Nevertheless, other abiotic effects (e.g., temperature, soil moisture content) may also have an effect and should not be neglected.

Community Membership and Structure

To obtain an idea of the beta diversity between our samples we analyzed bacterial/fungal community membership and structure. Principal Coordinate analysis (PCoA) using Jaccard distances (**Figure 1**) clearly revealed that membership between soil samples and plant samples differed for both bacterial (**Figure 1A**) and fungal (**Figure 1B**) communities. While no differences in community membership could be observed between early vs. standard culture plant samples (AMOVA centroid with p-value ≥ 0.05) the presence of two distinct clusters indicated that community membership clearly differed between early vs. standard soil samples. Statistical analyses also indicated that swath turning had a significant effect (AMOVA centroid with p-value < 0.05) on both bacterial and fungal community membership of plant samples, but not soil samples.

To analyze community structure we then used the Non-parametric MultiDimensional Scaling (NMDS) ordination of Yue & Clayton dissimilarities to determine distance matrices (Theta YC distances) between all samples (**Figure 2**). The results show that stress values for both bacterial (0.089) and fungal (0.085) communities are inferior to 0.1 as recommended by Mothur SOP (Standard Operating Procedure, https://www.mothur.org). Overall, and as observed for community membership data, clear differences in community *structure* occur

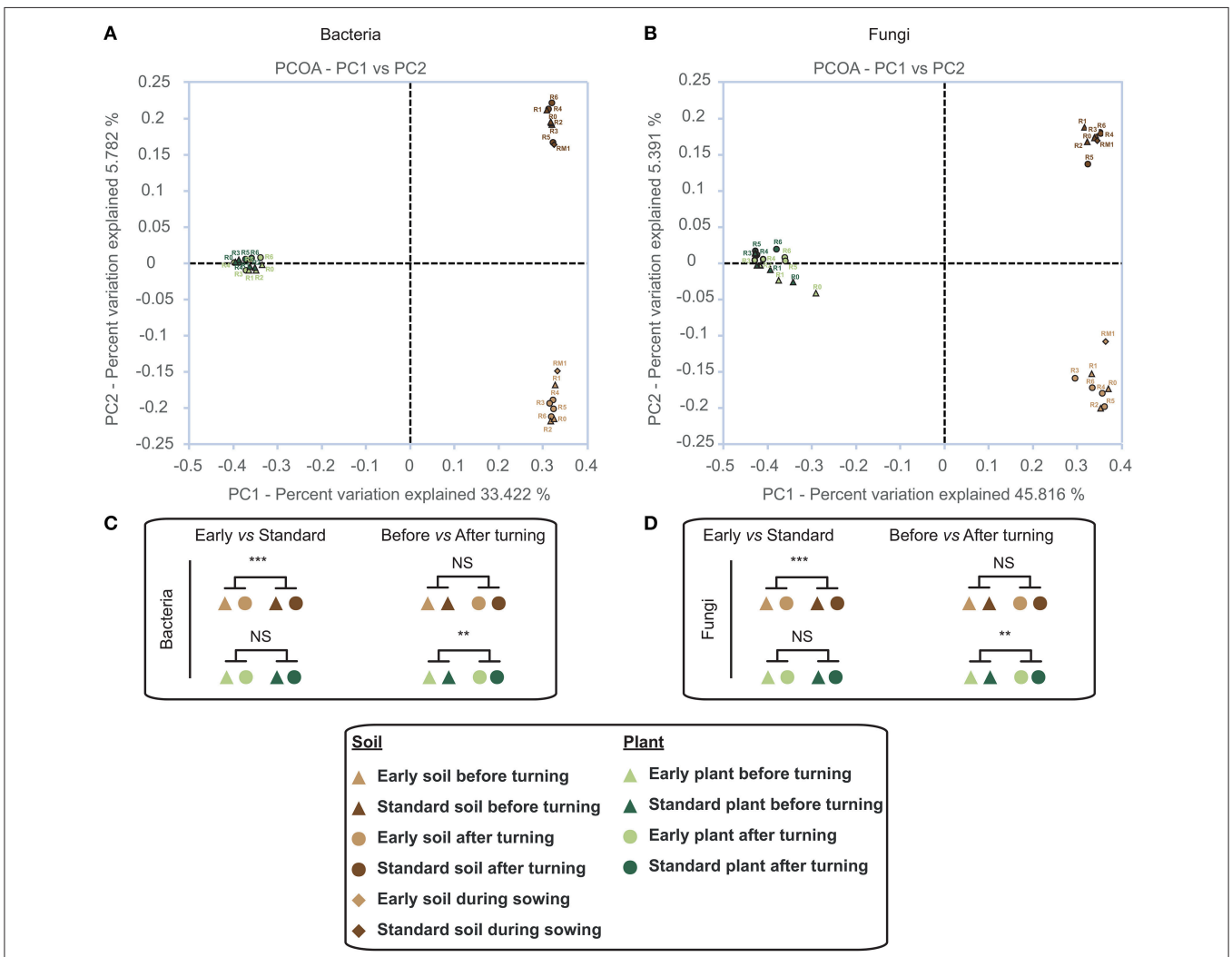


FIGURE 1 | Principal coordinate analysis (PCoA) of bacterial **(A)** and fungal **(B)** community membership based on Jaccard distances. Different colors indicate the source samples (light brown, soil early harvest; dark brown, soil standard harvest; light green, plant early harvest; dark green, plant standard harvest). Triangles, samples before swath turning; circles, samples after swath turning; lozenge, soil sample during sowing. **(C,D)** statistically significant clustering based on AMOVA.

between soil and plant samples (early and standard cultures) for both bacterial (**Figure 2A**) and fungal (**Figure 2B**) samples. However, the fungal R0 (early and standard cultures) plant samples form a separate cluster from the other plant samples whereas bacterial R0 plant samples do not. The community structure of early and standard bacterial/fungal soil samples, but not plant samples, is also significantly different (p-value < 0.001). As observed for community membership, swath turning also appeared to modify community structure, but not necessarily in the same samples. For bacteria, swath turning had a significant effect on community structure in both soil and plant samples (cf. community membership, significant effect only in plant, but not soil, samples). In contrast for fungi, swath turning only had a significant effect on the community structure of plant, but not soil samples. Calculation of Spearman rank correlation coefficients indicated that there was no significant correlation between climatic conditions (temperature, rainfall) and community structure (Supplementary Table 10).

Taxonomic Distribution of Identified Bacteria and Fungi

To evaluate taxonomic distribution of identified bacteria and fungi, OTUs were analyzed to determine consensus taxonomy (**Figure 3**). Overall more phyla (bacteria and fungi) were present in soil samples when compared to plant samples with 23 (excluding unclassified) bacterial and six fungal phyla in soil samples and 11 bacterial and four fungal phyla in plant samples. Of these phyla, 8 (bacteria) and 2 (fungi) were not previously associated with flax dew-retting in the literature thereby underlining the interest of a metabarcoding approach

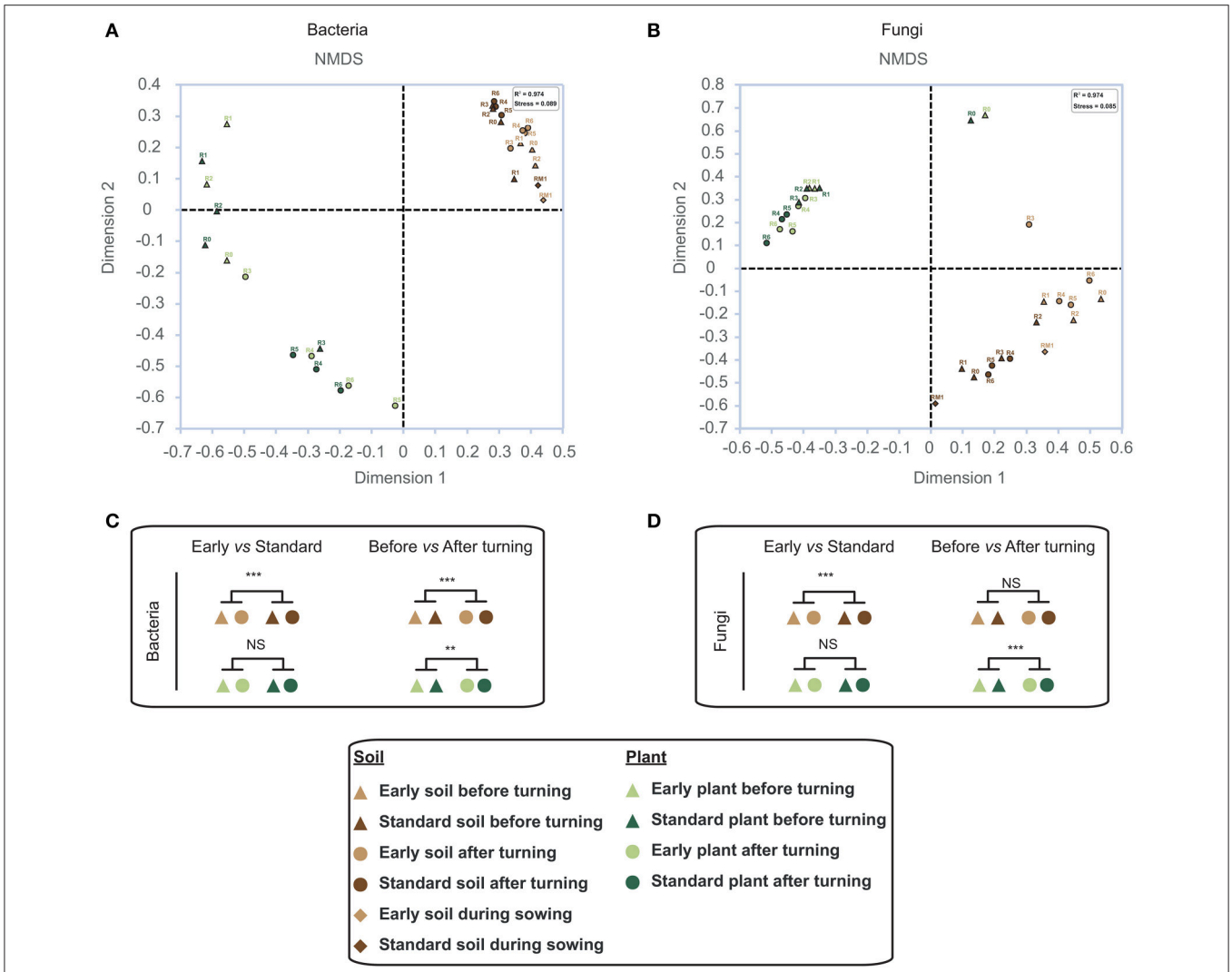


FIGURE 2 | Non-metric multidimensional scaling (NMDS) of bacterial **(A)** and fungal **(B)** community structure based on Yue & Clayton distance matrix. Different colors indicate the source samples (light brown, soil early harvest; dark brown, soil standard harvest; light green, plant early harvest; dark green, plant standard harvest). Triangles, samples before swath turning; circles, samples after swath turning; lozenge, soil sample during sowing. **(C,D)** statistically significant clustering based on AMOVA.

for the identification of new microorganisms. Although the number of phyla identified in soil samples was higher than in plant samples, the most abundant taxa were the same in both cases as might be expected in an analysis at this level (Phyla): Bacteria—Proteobacteria ($\bar{x}$ 60.64% ± 6.64); Fungi—Ascomycota ($\bar{x}$ 76.29% ± 4.047) (Supplementary Tables 11, 12). For both bacteria and fungi, the type of culture (early vs. standard) appeared to have little effect on phyla relative abundance, neither in soil nor in plant samples. The relative abundance in both bacterial and fungal soil samples appeared to remain fairly constant throughout the retting period. In contrast, relative abundance in bacterial *plant* samples was more dynamic being characterized by a relative increase and/or decrease in percentage relative abundances of Proteobacteria and Bacteroidetes at R2 (**Figure 3A**). The relative abundance in fungal plant samples appeared to be more stable throughout retting.

Subsequent analyses of plant samples at class level (**Figure 4**) indicated that the observed increase (**Figure 3**) in the % relative abundance of the Proteobacteria at R2 was mainly related to a substantial increase (>100%) in the relative

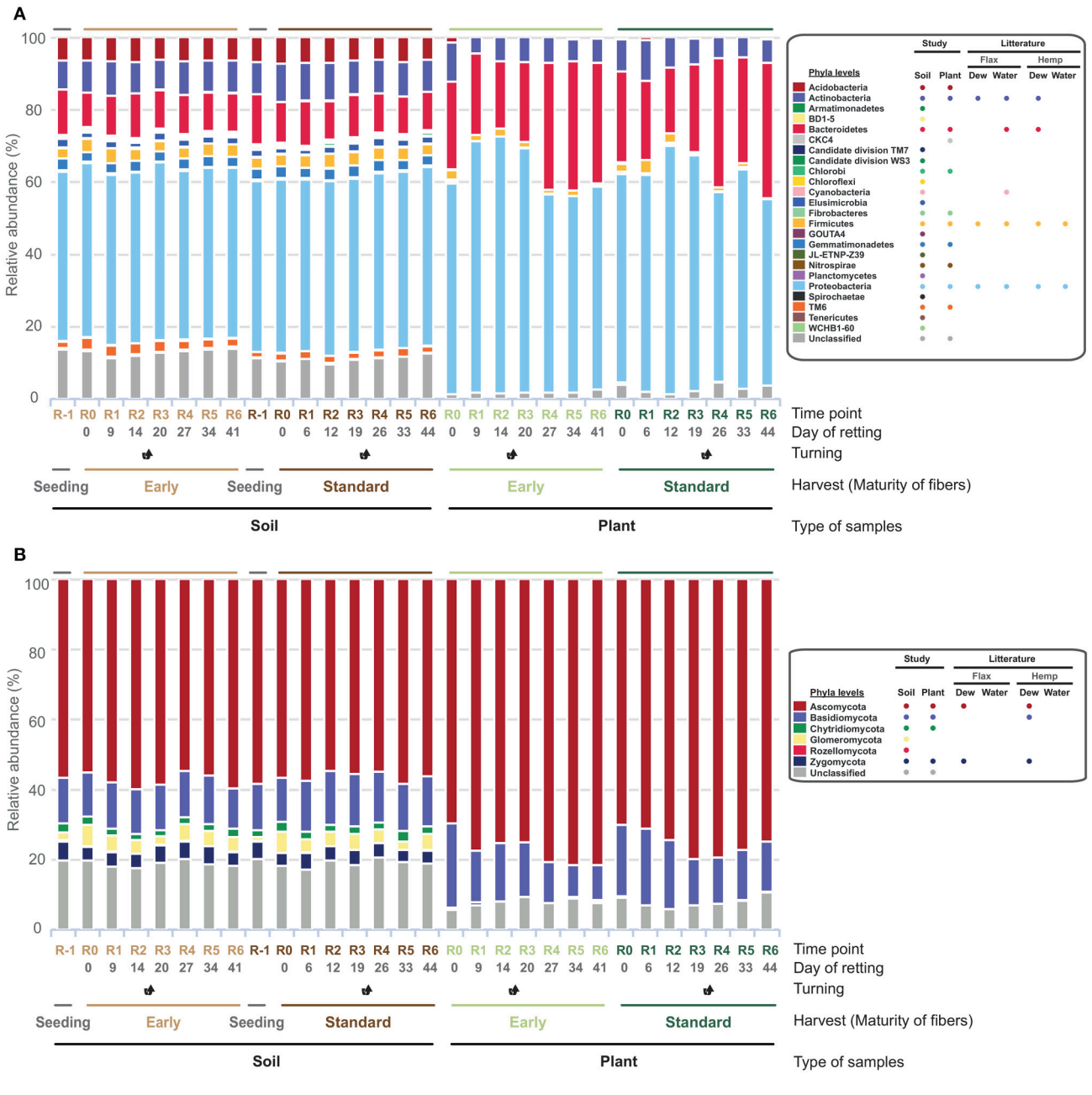


FIGURE 3 | Bacterial **(A)** and fungal **(B)** relative abundance of OTUs at the phyla level (soil samples $n = 16$, and plant samples $n = 14$). The consensus taxonomy for bacterial OTUs was assigned from the SILVA database and for fungal OTUs from the UNITE database.

abundance of the Gammaproteobacteria class in early samples (**Figure 4A**) correlated with a smaller reduction in relative abundances of the Flavobacteria and Sphingobacteria classes. Similarly, in standard samples the previously observed increase in Proteobacteria (**Figure 4**) could be related to the increase in Gammaproteobacteria and Betaproteobacteria, coupled with a decrease in relative abundance of Sphingobacteria at R2 (**Figure 4B**). Additional Proteobacteria peaks were also observed at R3 (Alphaproteobacteria) and R5 (Betaproteobacteria) but had less overall impact on the Proteobacteria/Bacteroidetes ratio in standard cultures due to an increase in relative abundance of Bacteroidetes classes in latter stages of retting.

Examination of relative abundances of fungal classes revealed a different pattern. Although more classes were identified (8/9 classes in the Ascomycota and Basidiomycota, respectively), the class Dothidiomycetes was by far the dominant class in both early (**Figure 4C**) and standard (**Figure 4D**) samples with a relative abundance ranging from a "low" of 40% (R0) and arriving at a

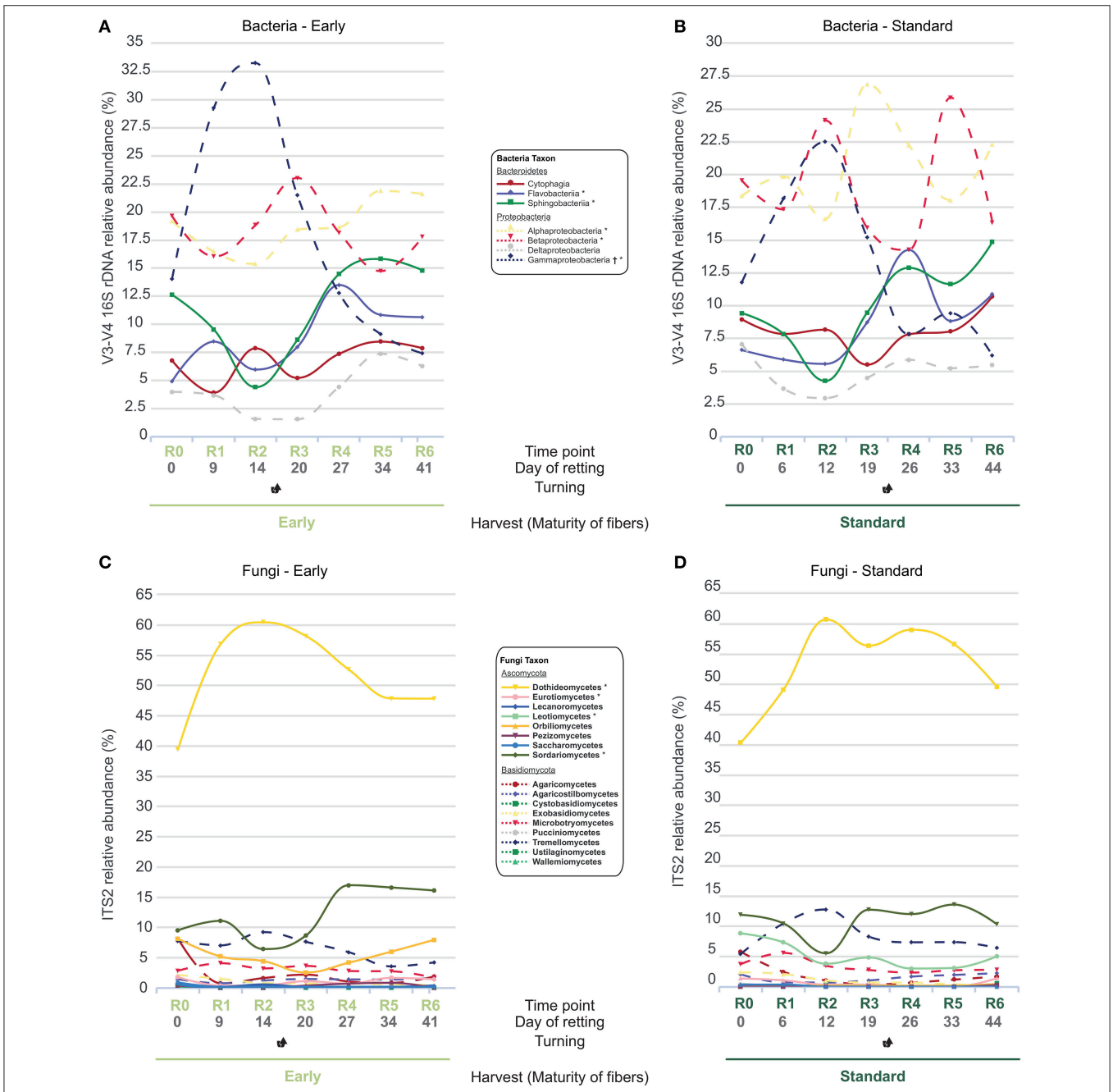


FIGURE 4 | Dynamics of V3-V4 16S rDNA and ITS2 gene relative abundance of different classes in the main bacterial **(A,B)** and fungal phyla **(C,D)** in straw samples. **(A,C)** early harvest, **(B,D)** standard harvest. Classes previously associated with flax dew-retting† and water-retting*.

maximum of 60+ % (R2). The Sordariomycetes were the next most abundant class with a value of between 10 and 15% relative abundance.

The generation of community distance heatmaps (**Figure 5**) for the top ten bacterial and fungal OTUs provided more detailed information on the different taxonomic groups represented in classes identified in plant samples. For bacteria (**Figure 5A**), results underlined the abundance of Sphingomonas and Pseudomonas genera. For fungi (**Figure 5B**), the order Capnodiales represented by *Cladosporium herbarum* was clearly the most abundant group. When analyzed globally, three different profiles could be identified: (i) the OTU is present throughout the retting period (e.g., OTU00001, Sphingomonas sp. and OTU2685, *C. herbarum*); (ii) the OTU is present at the beginning and then decreases (e.g., OTU00002, *Pseudomonas rhizosphaerae* and OTU00006, *Pantoea vagans*); and (iii) the OTU is absent at the beginning and then increases during retting (e.g., OTU00003, Rhizobium genus, OTU00004, Massilia sp. and OTU1918, Altenaria sp.).

Community-Level Analysis

Our results suggested that swath turning during retting had a significant effect on the microbial communities. To identify those OTUs most likely to explain differences highlighted by the diversity analyses between before- and after-turning - and that could therefore represent potential biomarkers of this process, we used three different tests (Metastas, LEfSE, and Indicator). Our results (**Supplementary Figure 6**, Supplementary Tables 13–15) identified 7/8 bacterial and fungal OTUs, respectively "before," and 14/6 bacterial and fungal OTUs "after," swath turning in all three tests. Four and six of these OTUs are present in the top 10 bacterial/fungal OTUs, respectively (**Figure 5**).

Hydrolytic Enzyme Potential and Trophic Mode Prediction

Dew-retting of flax straw occurs via the action of hydrolytic enzymes produced by microorganisms and we therefore used PICRUSt software followed by expert curation to predict the bacterial Carbohydrate Active enZyme (CAZy) families potentially present during dew-retting and playing a role in the degradation of cell wall polymers. For all stem samples the NSTI scores were around the 0.15 level considered as acceptable according to PICRUSt instructions. Our results (**Figure 6**, **Supplementary Figure 7**) show that a wide range of different enzymes targeting both the backbones and side chains of the major polysaccharide cell wall polymers (cellulose, hemicelluloses, pectins) are present. Altogether, 22, 32, and 6 CAZy families targeting pectin, hemicelluloses and cellulose polymers were identified (**Supplementary Figure 7**). Generally,

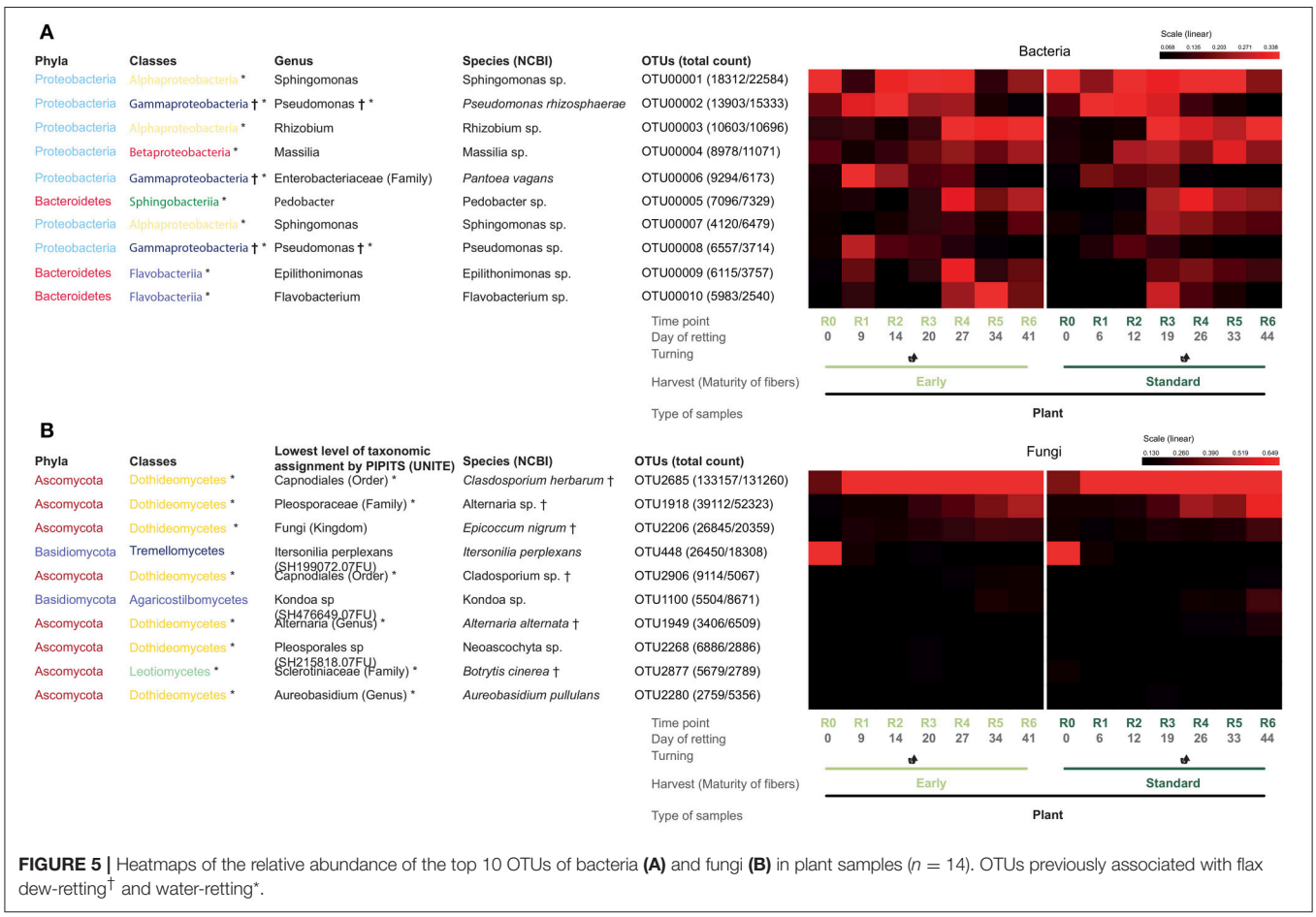


FIGURE 5 | Heatmaps of the relative abundance of the top 10 OTUs of bacteria **(A)** and fungi **(B)** in plant samples ($n = 14$). OTUs previously associated with flax dew-retting† and water-retting*.

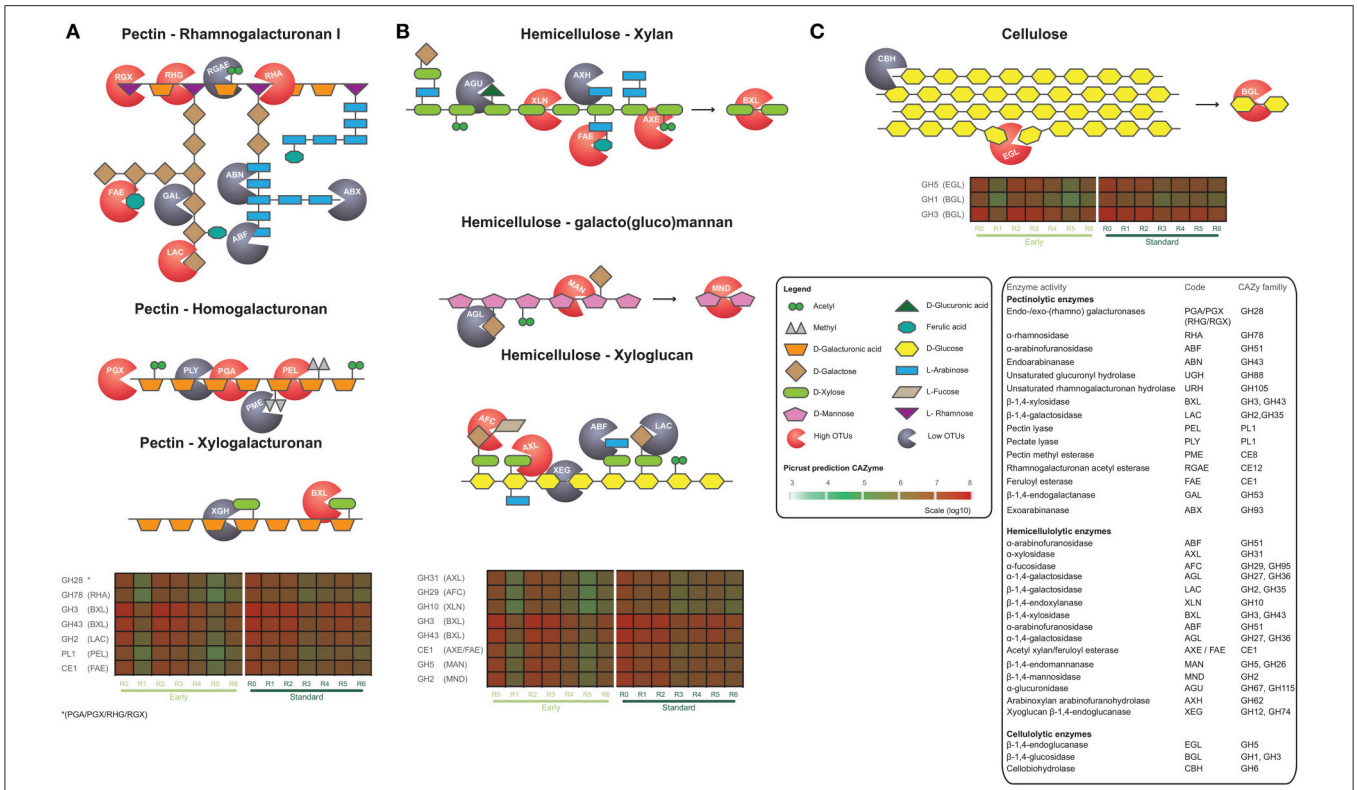


FIGURE 6 | Schematic representation of plant cell wall polysaccharides and predicted bacterial hydrolytic enzymes present in flax straw samples. **(A)** pectins, **(B)** hemicelluloses, and **(C)** cellulose. Heatmaps show the total counts of the most abundant CAZyme families in early and standard harvests obtained from the OTUs table (Greengenes Database used for consensus taxonomy) and generated by PICRUSt software. Red "pacman", high OTUs abundance enzymes; gray "pacman", low OTUs abundance enzymes.

the hydrolytic enzyme potential (all polymers) was greater during the first stages of retting (R0-R2/R3) compared to latter stages (R3/R4–R6) for both "early" and "standard" cultures. The drop in hydrolytic potential observed for R1 and R5 stages in "early" cultures is most likely related to the corresponding decrease in the most abundant bacterial OTU (e.g., Sphingomonas OTU00001, **Figure 5**). PICRUSt prediction does not exist for fungal OTUs and so hydrolytic enzyme potential cannot be directly predicted. Nevertheless, we were able to gain a relative idea of the overall hydrolytic enzyme potential by using the FUNGuild software that describes fungal trophic mode. Our results (**Figure 7**) show a progressive decrease in relative abundance of pathotrophs associated with a steady increase in saprotrophs and saprotrophs-pathotrophs as retting progresses. Pathogenic fungi generally produce a wider range of cell wall degrading enzymes than rot fungi and observed change in trophic mode during retting could suggest a decrease in hydrolytic enzyme diversity (Choi et al., 2013).

DISCUSSION

Microbial Identification and Retting Parameters

Previous studies using culture-based approaches and non-HTS metabarcoding have identified different bacteria and fungi phyla present during retting including Actinobacteria, Firmicutes, Proteobacteria (bacteria), and Ascomycota and Zygomycota (fungi) (Lanigan, 1950; Rosemberg, 1965; Brown, 1984; Sharma, 1986a,b; Donaghy et al., 1990; Henriksson et al., 1997). Our results obtained using metabarcoding coupled with HTS not only identified these phyla, but also allowed the identification of new phyla not previously associated with dew-retting. Overall we identified 95 bacteria and 215 fungi species in dew-retted flax straw (plant) samples. HTS metabarcoding has been recently used to investigate bacterial (but not fungal) population dynamics in water-retted flax (Zhao et al., 2016). A comparison of relative abundances of the major bacterial phyla identified indicates that water retting is very different from dew-retting, despite the fact that the same lignocellulosic material is being degraded. Major phyla identified during water-retting were Firmicutes (genus Clostridium) and Proteobacteria (genera Azotobacter and Enterobacter). In contrast, Firmicutes were only present in low abundance during dew-retting and Azotobacter were absent. These differences can be most likely related to the anaerobic environment of water-retting compared to the more aerobic environment of dew-retting. Indeed, Clostridium is an obligate anaerobe and is known to be an agent of water-retting (Donaghy et al., 1990; Tamburini et al., 2003).

Phyla, identified in our study and not previously associated with flax dew-retting, included, for the bacteria,

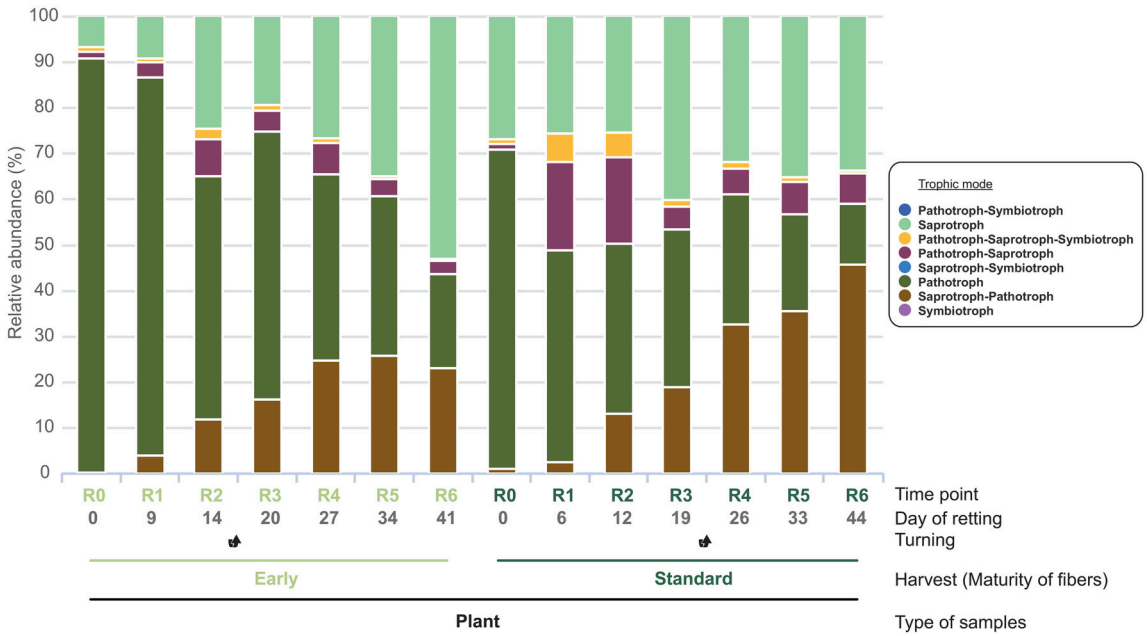


FIGURE 7 | Relative abundance of fungal OTUs classified by trophic mode. The trophic mode was assigned using the FUNGuild database and based on 35% of OTUs (i.e., total = 1048).

Acidobacteria, Bacteroidetes, CKC4, Chlorobi, Fibrobacteres, Gemmatimonadetes, Nitrospirae and TM6; and for the fungi, Basidiomycota and Chytridiomycota. The Bacteroidetes phylum has been associated with cellulose degradation in agricultural soils (Schellenberger et al., 2010) and was previously detected in hemp dew-retting (Ribeiro et al., 2015) and flax water-retting (Zhao et al., 2016). Our observation of this phylum could indicate that it is also involved in flax dew-retting. Basidiomycota are linked to plant cell wall degradation in different ecosystems (Baldrian et al., 2008; Schneider et al., 2012; Kuramae et al., 2013; Voříšková and Baldrian, 2013; Rytioja et al., 2014) and were also detected in hemp dew-retting (Ribeiro et al., 2015).

Although the observation that new bacterial phyla (except for the Bacteroidetes) and fungal phyla represent less than 2% of the whole microbiota might suggest that they are not involved in the retting process, some of these phyla are related to microorganisms characterized as biomass degraders in previous studies (Zhao et al., 2014). This observation, together with the fact that low abundance OTUs can still contribute to the decomposition of plant matter (Baldrian et al., 2012) indicates that these phyla should not be ignored during the study of dew-retting.

A number of parameters potentially affecting microbial population structure during retting were examined. It is commonly admitted by farmers that the maturity of flax plants has a direct impact on the retting time and influences the choice for the pulling (up-rooting) date. Generally, straw from younger plants (flowering/green capsule stage) rets more quickly than that of more mature plants (yellow/brown capsule stage). This is thought to be related to differences in cell wall composition (e.g., pectin/lignin modifications and/or deposition) and water content (Meijer et al., 1995; Day et al., 2005; Akin, 2013). Our results showing that there was no significant difference in microbial communities and colonization dynamics between the early vs. standard cultures would suggest that differences in retting time may indeed be related to differences in cell wall structure and not to population differences.

Compared to litter decay that normally proceeds undisturbed, dew-retting is a semi-controlled process during which the straw swaths are turned by farmers to obtain a more uniform fiber separation. Our analyses revealed that this practice had a significant effect on both bacterial and fungal community membership and structure of the flax straw microbiome confirming a real microbiological effect of swath turning that probably contributes to a more uniform retting.

Although our results indicated no significant correlation between measured climatic conditions (temperature and rainfall) and community structures during the retting period it is important to remember that our study was conducted within a single year. It is possible that significant variations in community structures may occur between different seasons and further work is necessary to clarify this point.

Microbial Dynamics

During dew-retting the relative abundance of the Bacteroidetes phylum increases while that of the Protobacteria decreases. A similar dynamic also occurs during biodegradation of field biomass from different angiosperm species (e.g., *Arundo donax*, *Eucalyptus camaldulensis,* and *Populus nigra*) suggesting, as might be expected, that similarities exist between the temporary

dew-retting ecosystem and degradation of lignocellulose in the field (Ventorino et al., 2015). Interestingly, the bacterial dynamics of flax dew-retting appear to be closer to that of field lignocellulose degradation than to that observed during flax water retting where Protobacteria increased during retting (Zhao et al., 2016). In this latter case, the phylum Proteobacteria was mainly represented by the genera Azotobacter that increased during retting and (to a much lesser extent) Enterobacter that remained constant. For fungal phyla we observed an increase in the relative abundance of Ascomycota at the expense of Basidiomycota in contrast to the situation generally observed during both field and forest litter decomposition (Schneider et al., 2012; Kuramae et al., 2013; Voříšková and Baldrian, 2013). The observed increase of Ascomycota was due to the saprophytic Altenaria species (Dang et al., 2015) that has previously been linked to later stages of dew-retting (Brown et al., 1986). In contrast, Altenaria species are more abundant during initial stages of litter decay (Snajdr et al., 2011). Our results also indicated that *C. herbarum* and *Epicoccum nigrum* contributed to the increase in Ascomycota relative abundance. During this stage less recalcitrant components of the biomass (pectins, and hemicelluloses) are progressively degraded (Dilly et al., 2001). Contrary to litter decay, dew-retting is a semi-controlled process and the challenge is to limit degradation of major quality related polymers such as crystalline cellulose. In this context, changes in the relative abundance of Ascomycota vs. Basidiomycota could represent an interesting bioindicator of retting progress.

More detailed information on population dynamics at different time points during retting was provided by analyzing the relative abundance of OTUs at different taxonomic rank (e.g., phyla, classes, or genus/species level). The most abundant bacterial OTU corresponded to Sphingomonas sp. that was present throughout most of the retting period in both early and standard cultures. Although Sphingomonas species have been previously identified during bamboo and hemp retting, as well as in forest litter microbiome, this is the first time they have been found in flax retting (Fu et al., 2011; Ribeiro et al., 2015) (Urbanová et al., 2015). These species are able to hydrolyze terminal non-reducing alpha-L-rhamnose residues in alpha-L-rhamnosides giving them the ability to degrade pectin (rhamnogalacturonan I and rhamnogalacturonan II) in the middle lamella (Hashimoto and Murata, 1998). Another *Sphingomonas* species, *S. paucimobilis* is also able to degrade lignin (Masai et al., 1999; de Gonzalo et al., 2016). The second most abundant OTU corresponded to *P. rhizosphaerae*, present during the early and medium retting stages but decreasing in latter stages. A number of Pseudomonas species have previously been associated with retting of different fiber plants (e.g., flax, hemp, jute, ramie) (Rosemberg, 1965; Munshi and Chattoo, 2008; Duan et al., 2012; Ribeiro et al., 2015). Pseudomonas sp. is considered as one of the most efficient lignin degradation bacterium (Shui Yang et al., 2007) and the genomes of both *Pseudomonas putida* and *Pseudomonas aeruginosa* contain genes encoding endoglucanases (Talia et al., 2012). Other abundant OTUs corresponded to Rhizobium, Pedobacter, and Flavobacterium that are known to show pectinase, cellulose, and hemicellulose activities (Mateos et al., 1992; McBride et al., 2009; López-Mondéjar et al., 2016). In addition, Pedobacter has also been identified during bamboo and hemp retting (Fu et al., 2011; Ribeiro et al., 2015) or forest litter degradation (Urbanová et al., 2015). In contrast to Sphingomonas and Pseudomonas, these organisms become more abundant toward the end of the retting period and could be associated with "over-retting" when the structural integrity of the fiber starts to be degraded.

In contrast to the more evenly distributed abundance of the bacterial OTUs, fungal OTUs were dominated by one major species—*C. herbarum*—that rapidly increased during early retting. This species, as well as the third most abundant OTU (*E. nigrum*) are known to be common dew-retting agents and are believed to degrade cellulose (Brown, 1984). Of the other fungal OTUs, all have previously been associated with dew-/water-retting except for *Itersonilia perplexans*. Interestingly, our results also indicated that *Alternaria alternata* is present at the start of retting. Traditionally, the appearance of this species is used as a signal that retting is starting to go too far and that the swaths should be collected (Brown et al., 1986).

Hydrolytic Enzyme Potential

Prediction of hydrolytic enzymes potentially present during retting was performed by using PICRUSt (Langille et al., 2013). This software successfully predicts bacterial enzymatic activities represented in different databases (e.g., KEGG Ortholog, COGs, or CAZy). Overall, a large collection of enzyme activities targeting both the main backbones and side chains of the major polysaccharide polymers were identified. Based on OTU counts, ∼38, 43, and 19 percent of the total hydrolytic enzyme potential targeted pectins, hemicelluloses, and cellulose, respectively. Despite the clear dynamics and significant changes in the straw microbiome these values remained constant throughout the retting period. Similar software does not exist for predicting fungal enzyme potential. This represents an important hurdle for obtaining a complete overview of the dew-retting process as fungi are major producers of extracellular hydrolytic enzymes (Schneider et al., 2012). Nevertheless, FUNGuild analysis showed that pathogenic taxa, present at the beginning of retting are progressively replaced by saprophytic fungi, more able to degrade lignocellulose. This change is most likely related to the fact that flax plants are still living when up-rooted.

In conclusion, we have shown that HTS metabarcoding is a powerful technique for analyzing complex bacterial and fungal community dynamics during flax dew-retting that can be used to identify different factors affecting the microbiota and—potentially—fiber isolation and quality. However, these results were obtained on samples retted in 1 year and it will be necessary to validate these data over several seasons. The use of PICRUSt data allows a predictive study of potential bacterial hydrolytic activity but should be coupled in future studies with alternative meta-omics methods such as metatranscriptomic or metaproteomic coupled with metagenomics to facilitate the assembling with appropriate reference genomes (Schneider et al., 2012; Dai et al., 2015; Hesse et al., 2015; Kuske et al., 2015; Wu et al., 2015). Such an approach would not only allow confirmation of bacterial enzyme dynamics but would also enable identification of fungal enzymes involved in this process.

AVAILABILITY OF DATA AND MATERIALS

The microbial DNA sequencing data sets supporting the results in this article are available at the EBI ENA with accession number PRJEB20299.

AUTHOR CONTRIBUTIONS

Conceptualization: CD, SG, and SH; Methodology: CD, SG, and SH; Experimentation: CD; Bioinformatic and statiscal analysis: CD; Writing—original draft: CD; Writing—Review and editing: SG and SH; Funding Acquisition: SG and SH.

FUNDING

This work was funded within the framework of the collaborative French "Future project" SINFONI. CD thanks the region of Hauts-de-France and Bpifrance for their financial support.

ACKNOWLEDGMENTS

The authors would like to thank the following people/organizations: The flax farmer C.A.L.I.R.A (Coopérative Agricole LInière de la Région d'Abbeville) for growing and retting flax plants, in particular Vincent DELAPORTE for his valuable assistance. The Genomic and Transcriptomic platform of Genopole Occitanie-Toulouse (INRA, GeT-Plage, http://get.genotoul.fr/) where sequencing was performed and more particularly Catherine ZHANCHETTA and Olivier BOUCHEZ. The Plateau d'Ecologie moléculaire et biochimie Evo-Eco-Paléo (CNRS—UMR 8198, Evolution, Ecologie et Paléontologie) for use of the LightCycler 480 and epMotion and more particularly Cécile GODÉ and Anne-Catherine HOLL for their technical support. Antoine PORTELETTE, Julien LE ROY, Sandrine ARRIBAT, and Brigitte CHABBERT for their valuable assistance during the sampling. Alexandrine THORE for her invaluable help in the Supplementary Figure 2 graphical design.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02052/full#supplementary-material

Supplementary Figure 1 | Rainfall, daily amplitudes of temperature and moisture, and retting time points at Martainneville during the 2014 dew-retting period. Raw data from the Abbeville station available on the infoclimat website (https://www.infoclimat.fr).

Supplementary Figure 2 | Schematic representation of the experimental site at Martainneville during the 2014 dew-retting campaign.

Supplementary Figure 3 | Rarefaction curves of observed OTUs in bacterial (16S) soil **(A)** and stem **(B)** samples, defined at a 97% sequence similarity cut-off.

Supplementary Figure 4 | Rarefaction curves of observed OTUs in fungal (ITS) soil **(A)** and plant **(B)** samples, defined at a 97% sequence similarity cut-off.

Supplementary Figure 5 | Alpha diversity of bacterial **(A–C)** and fungal **(D–F)** communities in soil and plant samples during flax dew-retting. **(A,D)** Richness (Chao1), **(B,E)** Evenness (Heip's), and **(C,F)** Diversity (Inverse Simpson). Different colors indicate the source samples (light brown, soil early harvest; dark brown, soil standard harvest; light green, plant early harvest; dark green, plant standard harvest). Plots show the distribution of results of 20,548 bacterial and 42,436 fungal sequences subsampled from each sample 1,000 times and calculated for the average (error bars for Chao1 and Inverse Simpson correspond to lower and higher bound 95% confidence intervals and for Heip's represent standard deviation).

Supplementary Figure 6 | Before- **(A,C)** and after- **(B,D)** swath turning biomarkers for bacteria **(A,B)** and fungi **(C,D)**. Biomarkers were identified by analyzing differential OTUs abundance using Metastats, Indicator and LEfSe.

Supplementary Figure 7 | Heatmaps showing the total counts of all CAZyme families in early and standard harvests obtained from the OTUs table (Greengenes Database used for consensus taxonomy) and generated by PICRUSt software.

REFERENCES

Abarenkov, K., Nilsson, R. H., Larsson, K. H., Alexander, I. J., Eberhardt, U., Erland, S., et al. (2010). The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytol.* 186, 281–285. doi: 10.1111/j.1469-8137.2009.03160.x

Akin, D. E. (2008). Plant cell wall aromatics: influence on degradation of biomass. *Biofuels Bioprod. Biorefining* 2, 288–303. doi: 10.1002/bbb.76

Akin, D. E. (2013). Linen most useful: perspectives on structure, chemistry, and enzymes for retting flax. *ISRN Biotechnol.* 2013, 23. doi: 10.5402/2013/186534

Akin, D. E., Rigsby, L. L., Henriksson, G., and Eriksson, K.-E. L. (1998). Structural effects on flax stems of three potential retting fungi. *Text. Res. J.* 68, 515–519. doi: 10.1177/004051759806800708

Amann, R. I., Ludwig, W., and Schleifer, K.-H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.

Baldrian, P., Kolařík, M., Stursová, M., Kopecký, J., Valášková, V., Vřtrovský, T., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* 6, 248–258. doi: 10.1038/ismej.2011.95

Baldrian, P., Valásková, V., and Valášková, V. (2008). Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiol. Rev.* 32, 501–521. doi: 10.1111/j.1574-6976.2008.00106.x

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15:293. doi: 10.1186/1471-2105-15-293

Barwell, L. J., Isaac, N. J., and Kunin, W. E. (2015). Measuring β - diversity with species abundance data. *J. Anim. Ecol.* 84, 1112–1122. doi: 10.1111/1365-2656.12362

Beckers, B., Op De Beeck, M., Weyens, N., Boerjan, W., and Vangronsveld, J. (2017). Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome* 5, 25. doi: 10.1186/s40168-017-0241-2

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al. (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* 329, 834–838. doi: 10.1126/science.1184984

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* 4, 914–919. doi: 10.1111/2041-210X.12073

Bodenhausen, N., Horton, M. W., and Bergelson, J. (2013). Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS ONE* 8:e56329. doi: 10.1371/journal.pone.0056329

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi: 10.1146/annurev.arplant.54.031902.134938

Bokulich, N. A., and Mills, D. A. (2013). Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl. Environ. Microbiol.* 79, 2519–2526. doi: 10.1128/AEM.03870-12

Brown, A. E. (1984). *Epicoccum nigrum*, a primary saprophyte involved in the retting of flax. *Trans. Br. Mycol. Soc.* 83, 29–35. doi: 10.1016/S0007-1536(84)80242-9

Brown, A. E., and Sharma, H. S. S. (1984). Production of polysaccharide-degrading enzymes by saprophytic fungi from glyphosate-treated flax and their involvement in retting. *Ann. Appl. Biol.* 105, 65–74. doi: 10.1111/j.1744-7348.1984.tb02803.x

Brown, A. E., Sharma, H. S. S., and Black, D. L. R. (1986). Relationship between pectin content of stems of flax cultivars, fungal cell wall-degrading enzymes and pre-harvest retting. *Ann. Appl. Biol.* 109, 345–351. doi: 10.1111/j.1744-7348.1986.tb05326.x

Campilho, R. D. S. G. (2015). *Natural Fiber Composites*. Boca Raton, FL: CRC Press.

Cardenas, E., Kranabetter, J. M., Hope, G., Maas, K. R., Hallam, S., and Mohn, W. W. (2015). Forest harvesting reduces the soil metagenomic potential for biomass decomposition. *ISME J.* 9, 1–12. doi: 10.1038/ismej.2015.57

Chao, A. (1984). Nonparametric estimation of the number of classes in a populatin. *Scand. J. Stat.* 11, 265–270. doi: 10.1214/aoms/1177729949

Choi, J., Kim, K.-T., Jeon, J., and Lee, Y.-H. (2013). Fungal plant cell wall-degrading enzyme database: a platform for comparative and evolutionary genomics in fungi and Oomycetes. *BMC Genomics* 14(Suppl. 5):S7. doi: 10.1186/1471-2164-14-S5-S7

Cragg, S. M., Beckham, G. T., Bruce, N. C., Bugg, T. D., Distel, D. L., Dupree, P., et al. (2015). Lignocellulose degradation mechanisms across the Tree of Life. *Curr. Opin. Chem. Biol.* 29, 108–119. doi: 10.1016/j.cbpa.2015.10.018

Dai, X., Tian, Y., Li, J., Luo, Y., Liu, D., Zheng, H., et al. (2015). Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Appl. Environ. Microbiol.* 81, 1375–1386. doi: 10.1128/AEM.03682-14

Dang, H. X., Pryor, B., Peever, T., and Lawrence, C. B. (2015). The Alternaria genomes database: a comprehensive resource for a fungal genus comprised of saprophytes, plant pathogens, and allergenic species. *BMC Genomics* 16:239. doi: 10.1186/s12864-015-1430-7

Day, A., Ruel, K., Neutelings, G., Crônier, D., David, H., Hawkins, S., et al. (2005). Lignification in the flax stem: evidence for an unusual lignin in bast fibers. *Planta* 222, 234–245. doi: 10.1007/s00425-005-1537-1

de Gonzalo, G., Colpa, D. I., Habib, M. H., and Fraaije, M. W. (2016). Bacterial enzymes involved in lignin degradation. *J. Biotechnol.* 236, 110–119. doi: 10.1016/j.jbiotec.2016.08.011

Dilly, O., Bartsch, S., Rosenbrock, P., Buscot, F., and Munch, J. C. (2001). Shifts in physiological capabilities of the microbiota during the decomposition of leaf litter in a black alder (*Alnus glutinosa* (Gaertn.) L.) forest. *Soil Biol. Biochem.* 33, 921–930. doi: 10.1016/S0038-0717(00)00239-X

Donaghy, J. A., Levett, P. N., and Haylock, R. W. (1990). Changes in microbial populations during anaerobic flax retting. *J. Appl. Bacteriol.* 69, 634–641. doi: 10.1111/j.1365-2672.1990.tb01556.x

Duan, S., Liu, Z., Feng, X., and Zheng, K. (2012). Diversity and characterization of ramie-degumming strains. *Sci. Agric.* 69, 119–125. doi: 10.1590/S0103-90162012000200006

Eastwood, D. C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., et al. (2011). The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. *Science* 333, 762–765. doi: 10.1126/science.1205411

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Esty, W. W. (1986). The efficiency of good's nonparametric coverage estimator. *Ann. Stat.* 14, 1257–1260. doi: 10.1214/aos/1176350066

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491. doi: 10.1007/s00424-009-0730-7

Fu, J., Mueller, H., de Castro, J. V., Yu, C., Cavaco-Paulo, A., Guebitz, G. M., et al. (2011). Changes in the bacterial community structure and diversity during bamboo retting. *Biotechnol. J.* 6, 1262–1271. doi: 10.1002/biot.201100105

Gołębiewski, M., Deja-Sikora, E., Cichosz, M., Tretyn, A., and Wróbel, B. (2014). 16S rDNA pyrosequencing analysis of bacterial community in heavy metals polluted soils. *Microb. Ecol.* 67, 635–647. doi: 10.1007/s00248-013-0344-7

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* 40, 237–264. doi: 10.1093/biomet/40.3-4.237

Gübitz, G. M., and Cavaco-Paulo, A. (2001). Biotechnology in the textile industry–perspectives for the new millennium. *J. Biotechnol.* 89, 89–90. doi: 10.1016/S0168-1656(01)00317-0

Gweon, H. S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D. S., et al. (2015). PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods Ecol. Evol.* 6, 973–980. doi: 10.1111/2041-210X.12399

Hashimoto, W., and Murata, K. (1998). alpha-L-rhamnosidase of Sphingomonas sp. R1 producing an unusual exopolysaccharide of sphingan. *Biosci. Biotechnol. Biochem.* 62, 1068–1074. doi: 10.1271/bbb.62.1068

Heip, C. (1974). A new index measuring evenness. *J. Mar. Biol. Assoc. U. K.* 54, 555–557. doi: 10.1017/S0025315400022736

Henriksson, G., Akin, D. E., Rigsby, L. L., Patel, N., and Eriksson, K.-E. L. (1997). Influence of chelating agents and mechanical pretreatment on enzymatic retting of flax. *Text. Res. J.* 67, 829–836. doi: 10.1177/004051759706701107

Henriksson, G., Akin, D. E., Slomczynski, D., and Eriksson, K. E. L. (1999). Production of highly efficient enzymes for flax retting by *Rhizomucor pusillus*. *J. Biotechnol.* 68, 115–123. doi: 10.1016/S0168-1656(98)00192-8

Hesse, C. N., Mueller, R. C., Vuyisich, M., Gallegos-Graves, L. V., Gleasner, C. D., Zak, D. R., et al. (2015). Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. *Front. Microbiol.* 6:337. doi: 10.3389/fmicb.2015.00337

Hirsch, P. R., Mauchline, T. H., and Clark, I. M. (2010). Culture-independent molecular techniques for soil microbial ecology. *Soil Biol. Biochem.* 42, 878–887. doi: 10.1016/j.soilbio.2010.02.019

Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* 12, 1889–1898. doi: 10.1111/j.1462-2920.2010.02193.x

Ihrmark, K., Bödeker, I. T., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., et al. (2012). New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* 82, 666–677. doi: 10.1111/j.1574-6941.2012.01437.x

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1. doi: 10.1093/nar/gks808

Knief, C. (2014). Analysis of plant microbe interactions in the era of next generation sequencing technologies. *Front. Plant Sci.* 5:216. doi: 10.3389/fpls.2014.00216

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.01043-13

Kubicek, C. P., Starr, T. L., and Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu. Rev. Phytopathol.* 52, 427–451. doi: 10.1146/annurev-phyto-102313-045831

Kuhad, R. C., and Singh, A. (1993). Lignocellulose biotechnology: current and future prospects. *Crit. Rev. Biotechnol.* 13, 151–172. doi: 10.3109/07388559309040630

Kuramae, E. E., Hillekens, R. H., De Hollander, M., Van Der Heijden, M. G., Van Den Berg, M., Van Straalen, N. M., et al. (2013). Structural and functional variation in soil fungal communities associated with lit ter bags containing maize leaf. *FEMS Microbiol. Ecol.* 84, 519–531. doi: 10.1111/1574-6941.12080

Kuske, C. R., Hesse, C. N., Challacombe, J. F., Cullen, D., Herr, J. R., Mueller, R. C., et al. (2015). Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecol.* 14, 133–137. doi: 10.1016/j.funeco.2014.12.005

Kvavadze, E., Bar-Yosef, O., Belfer-Cohen, A., Boaretto, E., Jakeli, N., Matskevich, Z., et al. (2009). 30,000-year-old wild flax fibers. *Science* 325, 1359–1359. doi: 10.1126/science.1175404

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Lanigan, G., w. (1950). Relationship of fiber quality of bacterial flora in flax retting. *Nature* 165, 516–517. doi: 10.1038/165516a0

López-Mondéjar, R., Zühlke, D., Becher, D., Riedel, K., and Baldrian, P. (2016). Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. *Sci. Rep.* 6:25279. doi: 10.1038/srep25279

Ludwig, J. A., and Reynolds, J. F. (1988). *Statistical Ecology: a Primer on Methods and Computing*. Available online at: http://dl.acm.org/citation.cfm?id=61827 (Accessed May 8, 2016).

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488, 86–90. doi: 10.1038/nature11237

Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66, 506–577. doi: 10.1128/MMBR.66.3.506-577.2002

Masai, E., Katayama, Y., Nishikawa, S., and Fukuda, M. (1999). Characterization of *Sphingomonas paucimobilis* SYK-6 genes involved in degradation of lignin-related compounds. *J. Ind. Microbiol. Biotechnol.* 23, 364–373. doi: 10.1038/sj.jim.2900747

Mateos, P. F., Jimenez-Zurdo, J. I., Chen, J., Squartini, A. S., Haack, S. K., Martinez- Molina, E., et al. (1992). Cell-associated pectinolytic and cellulolytic enzymes in *Rhizobium leguminosarum* biovar trifolii. *Appl. Environ. Microbiol.* 58, 1816–1822.

McBride, M. J., Xie, G., Martens, E. C., Lapidus, A., Henrissat, B., Rhodes, R. G., et al. (2009). Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl. Environ. Microbiol.* 75, 6864–6875. doi: 10.1128/AEM.01495-09

Md. Tahir, P., Ahmed, A. B., SaifulAzry, S. O. A., and Ahmed, Z. (2011). Retting process of some bast plant fibres and its effect on fibre quality: a review. *BioResources* 6, 5260–5281. doi: 10.15376/biores.6.4.5260-5281

Meijer, W. J. M., Vertregt, N., Rutgers, B., and van de Waart, M. (1995). The pectin content as a measure of the retting and rettability of flax. *Ind. Crops Prod.* 4, 273–284. doi: 10.1016/0926-6690(95)00041-0

Montella, S., Ventorino, V., Lombard, V., Henrissat, B., Pepe, O., and Faraco, V. (2017). Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses. *Sci. Rep.* 7:42623. doi: 10.1038/srep42623

Munshi, T. K., and Chattoo, B. B. (2008). Bacterial population structure of the jute-retting environment. *Microb. Ecol.* 56, 270–282. doi: 10.1007/s00248-007-9345-8

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 20, 241–248. doi: 10.1016/j.funeco.2015.06.006

Peršoh, D. (2015). Plant-associated fungal communities in the light of meta'omics. *Fungal Divers.* 75, 1–25. doi: 10.1007/s13225-015-0334-9

Pil, L., Bensadoun, F., Pariset, J., and Verpoest, I. (2016). Why are designers fascinated by flax and hemp fibre composites? *Compos. Part A Appl. Sci. Manuf.* 83, 193–205. doi: 10.1016/j.compositesa.2015.11.004

Plassart, P., Terrat, S., Thomson, B., Griffiths, R., Dequiedt, S., Lelievre, M., et al. (2012). Evaluation of the ISO standard 11063 DNA extraction procedure for assessing soil microbial abundance and community structure. *PLoS ONE* 7:e44279. doi: 10.1371/journal.pone.0044279

Preisner, M., Wojtasik, W., Kulma, A., Żuk, M., and Szopa, J. (2014). "Flax Fiber," in *Kirk-Othmer Encyclopedia of Chemical Technology* (Hoboken, NJ: John Wiley & Sons, Inc.), 1–32. doi: 10.1002/0471238961.0612012401110914.a01.pub2

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864

Ribeiro, A., Pochart, P., Day, A., Mennuni, S., Bono, P., Baret, J.-L., et al. (2015). Microbial diversity observed during hemp retting. *Appl. Microbiol. Biotechnol.* 99, 4471–4484. doi: 10.1007/s00253-014-6356-5

Rosemberg, J. A. (1965). Bacteria responsible for the retting of Brazilian flax. *Appl. Microbiol.* 13, 991–992.

Rytioja, J., Hildén, K., Yuzon, J., Hatakka, A., de Vries, R. P., and Mäkelä, M. R. (2014). Plant-polysaccharide-degrading enzymes from basidiomycetes. *Microbiol. Mol. Biol. Rev.* 78, 614–649. doi: 10.1128/MMBR.00035-14

Schellenberger, S., Kolb, S., and Drake, H. L. (2010). Metabolic responses of novel cellulolytic and saccharolytic agricultural soil Bacteria to oxygen. *Environ. Microbiol.* 12, 845–861. doi: 10.1111/j.1462-2920.2009.02128.x

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schneider, T., Keiblinger, K. M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., et al. (2012). Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J.* 6, 1749–1762. doi: 10.1038/ismej.2012.11

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60

Sharma, H. (1986a). Enzymatic degradation of residual non-cellulosic polysaccharides present on dew-retted flax fibres. *Appl. Microbiol. Biotechnol.* 1987, 358–362.

Sharma, H. (1986b). The role of bacteria in retting of desiccated flax during damp weather. *Appl. Microbiol. Biotechnol.* 24, 463–467.

Shui Yang, J., Ren Ni, J., Li Yuan, H., and Wang, E. (2007). Biodegradation of three different wood chips by Pseudomonas sp. PKE117. *Int. Biodeterior. Biodegrad.* 60, 90–95. doi: 10.1016/j.ibiod.2006.12.006

Simpson, E. H. (1949). Measurement of diversity. *Nature* 163, 688–688. doi: 10.1038/163688a0

Snajdr, J., Cajthaml, T., Valášková, V., Merhautová, V., Petránková, M., Spetz, P., et al. (2011). Transformation of *Quercus petraea* litter: successive changes in litter chemistry are reflected in differential enzyme activity and changes in the microbial community composition. *FEMS Microbiol. Ecol.* 75, 291–303. doi: 10.1111/j.1574-6941.2010.00999.x

Soliveres, S., van der Plas, F., Manning, P., Prati, D., Gossner, M. M., Renner, S. C., et al. (2016). Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature* 536, 456–459. doi: 10.1038/nature19092

Staley, J. T., and Konopka, A. (1985). Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* 39, 321–346. doi: 10.1146/annurev.mi.39.100185.001541

Suenaga, H. (2012). Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ. Microbiol.* 14, 13–22. doi: 10.1111/j.1462-2920.2011.02438.x

Talia, P., Sede, S. M., Campos, E., Rorig, M., Principi, D., Tosto, D., et al. (2012). Biodiversity characterization of cellulolytic bacteria present on native Chaco soil by comparison of ribosomal RNA genes. *Res. Microbiol.* 163, 221–232. doi: 10.1016/j.resmic.2011.12.001

Tamburini, E., León, A. G., Perito, B., and Mastromei, G. (2003). Characterization of bacterial pectinolytic strains involved in the water retting process. *Environ. Microbiol.* 5, 730–736. doi: 10.1046/j.1462-2920.2003.00462.x

Terrat, S., Christen, R., Dequiedt, S., Lelièvre, M., Nowak, V., Regnier, T., et al. (2012). Molecular biomass and MetaTaxogenomic assessment of soil microbial communities as influenced by soil DNA extraction procedure. *Microb. Biotechnol.* 5, 135–141. doi: 10.1111/j.1751-7915.2011.00307.x

Terrat, S., Plassart, P., Bourgeois, E., Ferreira, S., Dequiedt, S., Adele-Dit-De-Renseville, N., et al. (2015). Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition. *Microb. Biotechnol.* 8, 131–142. doi: 10.1111/1751-7915.12162

Toju, H., Tanabe, A. S., Yamamoto, S., and Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* 7:e40863. doi: 10.1371/journal.pone.0040863

Urbanová, M., Šnajdr, J., and Baldrian, P. (2015). Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. *Soil Biol. Biochem.* 84, 53–64. doi: 10.1016/j.soilbio.2015.02.011

Ventorino, V., Aliberti, A., Faraco, V., Robertiello, A., Giacobbe, S., Ercolini, D., et al. (2015). Exploring the microbiota dynamics related to vegetable biomasses degradation and study of lignocellulose-degrading bacteria for industrial biotechnological application. *Sci. Rep.* 5:8161. doi: 10.1038/srep08161

Visi, D. K., D'Souza, N., Ayre, B. G., Webber Iii, C. L., and Allen, M. S. (2013). Investigation of the bacterial retting community of kenaf (*Hibiscus cannabinus*) under different conditions using next-generation semiconductor sequencing. *J. Ind. Microbiol. Biotechnol.* 40, 465–475. doi: 10.1007/s10295-013-1242-1

Voříšková, J., and Baldrian, P. (2013). Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME J.* 7, 477–486. doi: 10.1038/ismej.2012.116

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Warren, R. A. (1996). Microbial hydrolysis of polysaccharides. *Annu. Rev. Microbiol.* 50, 183–212.

White, J. R., Nagarajan, N., and Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* 5:e1000352. doi: 10.1371/journal.pcbi.1000352

Wu, S., Ren, Y., Peng, C., Hao, Y., Xiong, F., Wang, G., et al. (2015). Metatranscriptomic discovery of plant biomass-degrading capacity from grass carp intestinal microbiomes. *FEMS Microbiol. Ecol.* 91:fiv107. doi: 10.1093/femsec/fiv107

Yue, J. C., and Clayton, M. K. (2005). A similarity measure based on species proportions. *Commun. Stat. Theor. Methods* 34, 2123–2131. doi: 10.1080/STA-200066418

Zhang, J., Henriksson, H., Szabo, I. J., Henriksson, G., and Johansson, G. (2005). The active component in the flax-retting system of the zygomycete *Rhizopus oryzae* sb is a family 28 polygalacturonase. *J. Ind. Microbiol. Biotechnol.* 32, 431–438. doi: 10.1007/s10295-005-0014-y

Zhao, D., Liu, P., Pan, C., Du, R., Ping, W., and Ge, J. (2016). Bacterial succession and metabolite changes during flax (*Linum usitatissimum* L.) retting with *Bacillus cereus* HDYM-02. *Sci. Rep.* 6:31812. doi: 10.1038/srep31812

Zhao, Z., Liu, H., Wang, C., and Xu, J. (2014). Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 15:6. doi: 10.1186/1471-2164-15-6

Zhou, C.-H., Xia, X., Lin, C.-X., Tong, D.-S., and Beltramini, J. (2011). Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels. *Chem. Soc. Rev.* 40, 5588–5617. doi: 10.1039/c1cs15124j

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Djemiel, Grec and Hawkins. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

REVIEW
published: 23 May 2018
doi: 10.3389/fmicb.2018.01020

Sequencing of the Cheese Microbiome and Its Relevance to Industry

Bhagya. R. Yeluri Jonnala[1,2], *Paul L. H. McSweeney*[1], *Jeremiah J. Sheehan*[2] *and Paul D. Cotter*[2,3]*

[1] Food and Nutrition Deptartment, University College Cork, Cork, Ireland, [2] Teagasc Food Research Centre, Fermoy, Ireland, [3] APC Microbiome Ireland, Cork, Ireland

The microbiota of cheese plays a key role in determining its organoleptic and other physico-chemical properties. It is essential to understand the various contributions, positive or negative, of these microbial components in order to promote the growth of desirable taxa and, thus, characteristics. The recent application of high throughput DNA sequencing (HTS) facilitates an even more accurate identification of these microbes, and their functional properties, and has the potential to reveal those microbes, and associated pathways, responsible for favorable or unfavorable characteristics. This technology also facilitates a detailed analysis of the composition and functional potential of the microbiota of milk, curd, whey, mixed starters, processing environments, and how these contribute to the final cheese microbiota, and associated characteristics. Ultimately, this information can be harnessed by producers to optimize the quality, safety, and commercial value of their products. In this review we highlight a number of key studies in which HTS was employed to study the cheese microbiota, and pay particular attention to those of greatest relevance to industry.

Keywords: high throughput sequencing, cheese, microbiota, sensory characteristics, metatranscriptomics, industry

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Baltasar Mayo, Consejo Superior de Investigaciones Científicas (CSIC), Spain
Francesco Grieco, Istituto di Scienze delle Produzioni Alimentari (ISPA), Italy
Rodrigo Bibiloni, Arla Foods, Denmark
Fernanda Mozzi, CERELA-CONICET, Argentina

***Correspondence:**
Paul D. Cotter
paul.cotter@teagasc.ie

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *19 December 2017*
Accepted: *30 April 2018*
Published: *23 May 2018*

Citation:
Yeluri Jonnala BR, McSweeney PLH, Sheehan JJ and Cotter PD (2018) Sequencing of the Cheese Microbiome and Its Relevance to Industry. Front. Microbiol. 9:1020. doi: 10.3389/fmicb.2018.01020

INTRODUCTION

Cheese has a diverse microbial community, which indeed can vary within the cheese from the core to the surface that is greatly influenced by manufacturing including ripening conditions. Understanding the composition of this community (microbiota), and its impact on the quality and safety of cheese products, is of critical importance. In addition to, in the majority of cases, consciously added starter and adjunct bacteria (which are added as a supplement), cheese contains a heterogeneous variety of other, non-starter, microorganisms. These various microorganisms can play vital roles in the development of the organoleptic properties of cheese (Fox et al., 2000), nutrient composition, shelf-life, and safety.

Historically, culture-based microbiology techniques were used to gain an understanding of the microbial component of cheese. However, it has become increasingly clear that this approach can be limited in its ability to detect "difficult-to-culture" or sub-dominant microorganisms, thereby potentially providing misleading results. As a result, culture-independent approaches have become

Abbreviations: HTS, High Throughput Sequencing; NGS, Next Generation Sequencing; PDO, Protected designation of Origin; CIP, Cleaning-in-Place; NMC, Natural Milk Starter Cultures; NWC, Natural Whey Cultures; VOC, Volatile Components.

increasingly popular. These approaches involve DNA, and occasionally RNA, based molecular methods, with high throughput sequencing (HTS) gaining particular attention by virtue of its potential to provide an insight into the total microbial community of the cheese. The application of HTS can provide valuable information relating to the influence of geography, manufacturing processes, climatic conditions, seasonal variation, use of raw or pasteurized milk, and a variety of other factors on the cheese microbiota (**Figure 1**). The application of HTS for microbiota analysis can involve any of three primary approaches i.e., (1) Amplicon sequencing—whereby a fragment of highly conserved gene, ideally with variable regions therein, is used for sequencing, with comparison to databases allowing taxonomic assignment. The 16S rRNA gene is most frequently used and provides an insight into the bacterial composition of samples (generally to the genus level), (2) Shotgun metagenomic sequencing—which involves non-targeted sequencing of the DNA in a sample and, again on the basis of comparison with databases, can be used to classify all of the microorganisms, i.e., not just bacteria present (to species, or even strain, level) as well as information regarding the functional potential of the community, (3) Metatranscriptomics (RNASeq)—whereby total mRNA in the sample is sequenced (after first being converted to cDNA) to reveal the extent to which different genes are expressed and, in turn, the relative activity of different components of the community. These three experimental procedures differ with respect to library preparation, sequencing strategy, and the approaches taken for bioinformatic analysis. Some of the most common outputs from these analyses relate to α-diversity, β-diversity, and a determination of the relative abundance of different taxa. α-diversity is a measurement of diversity, such as total number of species, within a sample and measures such as the Simpson and Shannon diversity indices are employed. β-diversity reflects differences in diversity across different samples, with Bray-Curtis and UniFrac being among the distance metrics used. For metagenomics and metatranscriptomics, several tools are available for key steps, i.e., binning, assembly (where relevant), and mapping/assignment of sequences obtained (Di Bella et al., 2013; Walsh et al., 2017). There have been a number of research and review papers dedicated to describing methodologies used for sequencing (Di Bella et al., 2013) work flow, limitations, applications to dairy foods (Ercolini, 2013; Walsh et al., 2017), comparing sequencing platforms, and bioinformatics pipelines as well as highlighting pros and cons associated with all of the above (Kelleher et al., 2015; Clooney et al., 2016; Walsh et al., 2018) Depending on sequencing depth, HTS can facilitate the detection of taxa present at low levels, a feature of great value in the context of testing for spoilage or pathogenic microbes. HTS technology can also be employed to study the microbial diversity of protected designation of origin (PDO) cheeses to establish which components are responsible for the authentic taste, flavors, and textures of these products (Alegría et al., 2012; Fuka et al., 2013; De Filippis et al., 2014; Delcenserie et al., 2014; De Pasquale et al., 2014, 2016; Riquelme et al., 2015; Dalmasso et al., 2016; Parente et al., 2016; Giello et al., 2017; Gonçalves Dos Santos et al., 2017; Li et al., 2017). The same technology could, in the future, be employed to establish authenticity. However,

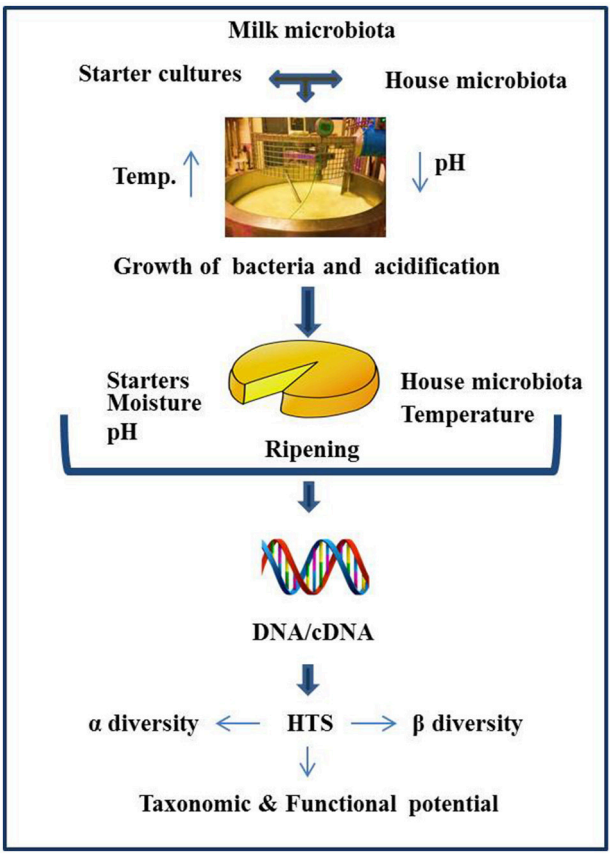


FIGURE 1 | Schematic representation of factors influencing the cheese microbiota, as revealed by HTS.

as these topics, and studies related to microbial dynamics of food ecosystems, processing environments, and production facilities have been reviewed in depth recently (Mayo et al., 2014; Galimberti et al., 2015; Bokulich et al., 2016; De Filippis et al., 2017; Doyle et al., 2017b; Macori and Cotter, 2018), we are not addressing those topics here. In this review, we specifically focus on providing an overview of studies in which HTS has been applied to provide insights into the microbiota of different cheeses that are of value from an industrial perspective as well as the factors that modulate these microbes (**Table 1**) and the identity of adventitious and other microbes (**Table 2**).

IDENTIFICATION OF FACTORS INFLUENCING THE DEVELOPMENT OF THE CHEESE MICROBIOTA

Microbial composition and diversity differs from raw to pasteurized milk, and between curd, whey, and cheese. The raw milk microbiota is influenced by microbes present in the teat canal, the surface of teat skin, hygiene practices, animal handlers, and the indigenous microbiota of equipment and storage containers. The origin of the milk would also seem

TABLE 1 | HTS detection of increase and decrease of microbiota by the action of some influencing factors.

Influencing factor	Cheese type	Increases/decreases	References
Herbs	Irish artisanal	↑Lactobacilli ↓Lactococci	Quigley et al., 2012
Salt	Irish artisanal	↓ *Leuconostoc* ↓ *Pseudomonas*	Quigley et al., 2012
Hay (feed to cows)	Caciocavallo	↑Salty, Sour, Umami	Giello et al., 2017
Silage (feed to cows)	Caciocavallo	↑Tenderness and oiliness	Giello et al., 2017
Wooden surfaces	Cotija	↑*L. mesenteroides* ↑*W. paramesenteroides*	Escobar-Zepeda et al., 2016
Temperature & pH	Mozzarella Grana Padano Parmigiano Reggiano	↑*L. lactis* ↑*L. fermentum* ↑*S. thermophilus* ↑*L. delbrueckii* ↑*L. helveticus*	De Filippis et al., 2014
Moisture	Brie Camembert Cloth bound cheddar, St. Nectaire, Tomme de Savoie	↑*Galactomyces* ↓ *Scopulariopsis* ↓ *Aspergillus* ↓ *Actinobacteria* ↓ *Staphylococcus*	Wolfe et al., 2014
Previous evening milk (Stored at 10°C)	Plaisentif	*Lactococcus, Lactobacillus* and *Streptococcus*	Dalmasso et al., 2016
Smoking	Herve	↓Enterobacteria	Alegría et al., 2012
Scalding temperature	Dutch-type Cheese	↑*Lactobacillus* spp. ↑*Leuconostoc* spp	Porcellato and Skeie, 2016
Washing of rind NMC NWC	Irish Cheese Italian PDO Cheese Italian PDO Cheese	↑Lactococci ↑ *S. thermophilus* ↑ *L. delbrueckii* ↑ *L. lactis*, *Lactobacillus fermentum*, *S. thermophilus*, *L. delbrueckii* and *Lactobacillus helveticus*	Quigley et al., 2012 Parente et al., 2016 De Filippis et al., 2014
Production practices ↑Temperature ↓Humidity	Serpa Caciocavallo Silano	Yeast population ↑NSLAB	Gonçalves Dos Santos et al., 2017 De Filippis et al., 2016

Abundance of bacteria in different cheese samples was detected by HST, ↑ (increase) and ↓ (decrease) indicates levels of certain bacteria as influenced by factors stated

to influence the levels of diversity therein, with cow's milk appearing to be more diverse than that from goats and sheep (Quigley et al., 2012). The type of grazing system employed, i.e., extensive or semi-extensive, affects the dynamics of teat skin microbiota and thus, in turn, the microbes in raw milk cheeses. Indeed, in one study, 27% of bacteria detected in raw milk cheese were also found on the teat surface. These bacteria included species involved in the production of flavor, aroma, and color development, such as *Brevibacterium linens*, *Staphylococcus equorum*, and Lactic Acid Bacteria (LAB) such as *Lactococcus lactis*, *Lactococcus chungangensis/raffinolactis,* and *Lactobacillus casei/paracasei*, which can contribute to protein and fat metabolism (Frétin et al., 2018). The sensory attributes of cheese can also be influenced by the type of feed (e.g., hay, silage) given to cows. Indeed, this was recently demonstrated in a study of Caciocavallo cheese where salty, sour, bitter, and umami flavors were lower in "hay"-cheese compared to "silage"-cheese, and higher levels of tenderness and oiliness were also evident

TABLE 2 | Adventitious, previously overlooked, and spoilage bacteria identified by HTS analysis of different cheese matrix and dairy environment.

Bacteria	Cheese type	Function	References
ADVENTITIOUS			
Pseudoalteromonas spp.	Cheese stored at low temperature	Flavor compounds	Wolfe et al., 2014
Lactic acid bacterial strains	Cotija	Bacteriocins for biopreservation	Escobar-Zepeda et al., 2016
Geotrichum candidum	Camemberti	Aroma, texture	Lessard et al., 2014
Penicillium camemberti			
Brevibacterium linens	Raw milk cheese	Protein and fat metabolism	Frétin et al., 2018
Staphylococcus equorum			
Lactococcus lactis			
Lactococcus chungangensis/raffinolactis			
Lactobacillus casei/paracasei			
Lactococcus	Dutch-type Cheese	Flavor	Porcellato and Skeie, 2016
S. thermophilus	Mozzarella, Grana Padano, Parmigiano Reggiano	*lacS* gene	De Filippis et al., 2014
S. thermophilus	Undefined milk starter	*serB* gene	Parente et al., 2016
Mesophilic lactobacilli	Fiore Sardo	Secondary proteolysis, esters, alcohols, aldehydes, sulfur compounds	De Pasquale et al., 2016
	Pecorino Siciliano		
	Pecorino Toscano		
Thermophilic LAB	Percorino Siciliano	Free amino acids	De Pasquale et al., 2016
Brevibacterium	Percorino Toscano	dimethyl trisulphide, dimethyl disulphide, methional, S-methyl thioesters	De Pasquale et al., 2016
B. linens	Surface ripened Cheese	Sulphur compounds, 2-methyl-1-butanol	Bertuzzi et al., 2018
Geotrichum candidum			
Staphylococcus xylosus			
Debaryomyces hansenii	Surface ripened Cheese	Alcohols	Bertuzzi et al., 2018
Glutamicibacter arilaitensis	Surface ripened Cheese	Ketones, alcohols, acids	Bertuzzi et al., 2018
Acetobacter	Kazak	VOC, amino acids, fatty acids	Zheng et al., 2018
Lactococcus			
Bacillus			
Staphylococcus			
Kurthia, Moraxella.			
Dipodascus,			
Pichia, Penicillium,			
Issatchenkia, Candida			
Lactococcus	Canestrato Pugliese	Degrade carbon sources	De Pasquale et al., 2014
Pseudoalteromonas	Washed rind Cheese	MGL	Wolfe et al., 2014
LAB	Cotija	Bacteriocins	Escobar-Zepeda et al., 2016
G. candidum	Camembert	Sensory characteristics	Lessard et al., 2014
Penicillium camemberti			
S. thermophilus	Reblochon-style French Cheese	Protein folding, Sorting, degradation, signal transduction	Monnet et al., 2016
L. delbrueckii	Reblochon-style French Cheese	Transcription, Translation, nucleotide and lipid metabolism	Monnet et al., 2016
G. candidum	Reblochon-style French Cheese	Lipid and amino acid metabolism	Monnet et al., 2016
D. hansenii	Reblochon-style French Cheese	Other amino acids metabolism	Monnet et al., 2016
L. lactis	Surface ripened Cheese	Lactose metabolism, Lipolysis, Proteolysis, catabolism of amino acids and fatty acids	Dugat-Bony et al., 2015
Kluyveromyces lactis			
D. hansenii			
G. candidum			
Corynebacterium casei			
Hafnia alvei			

(Continued)

TABLE 2 | Continued

Bacteria	Cheese type	Function	References
PREVIOUSLY OVERLOOKED			
Exiguobacterium	Queso Fresco		Lusk et al., 2012
Flavobacterium	Plaisentif		Dalmasso et al., 2016
Brevibacterium			
Salinicoccus			
Vagococcus			
Anaerobacillus			
Sphingobacterium			
Klebsiella			
Dehalobacter	Cotija		Escobar-Zepeda et al., 2016
Desulfohalobium			
Halomonas			
Thermohalobacter			
Haloquadratum			
Prevotella	Irish artisanal cheese		Quigley et al., 2012
Faecalibacterium			
Idiomarina	Brine		Marino et al., 2017
GNO2			
TM7			
Tremellomycetes	Wooden Shelves		Guzzon et al., 2017
SPOILAGE/CONTAMINANTS			
C. tyrobutyricum	Grana Padano	Blowing defect	Bassi et al., 2015
C. butyricum			
Klebsiella	Plaisentif	Poor hygiene, Contaminants of Boilers	Dalmasso et al., 2016
Morganella			
Erwinia			
Acinetobacter			
Brevibacterium	Smear ripened Cheese	Red-brown defect	Guzzon et al., 2017
Corynebacterium			
Microbacterium			
Enterococcus faecalis	Herve	Contaminant	Delcenserie et al., 2014
Thermus thermophilus	Pink defective cheese	Pink color	Quigley et al., 2016
Escherichia spp.	NWC	Contaminants	De Filippis et al., 2014
Enterobacter cowanii			
Enterobacteriaceae			
Agrobacterium spp.			
Alicyclobacillus spp.			
Propionibacterium acnes			
Lactobacillus curvatus	Irish artisanal cheeses	Production of tyramine	
Enterococcus faecium	Reblochon, Morbier,		
Enterococcus faecalis	Tête deMoine,	Production of Histamine	O'Sullivan et al., 2015b
Lactobacillus buchneri	Pecorino Sardo		

in "silage"-cheese relative to "hay"-cheese (Giello et al., 2017). High temperature treatment (when pasteurized milk is used) and low pH also contribute to the selection of specific bacteria in some artisanal cheeses that are made in the absence of starter bacteria (De Filippis et al., 2014). These and other factors, such as salt content, degrees of ripening, addition of ingredients like herbs, and spices also influence the cheese microbiota and, in turn, organoleptic properties of cheese. The influence of such factors was demonstrated in one representative study of Irish artisanal cheeses where, for example, proportions of lactobacilli increased and lactococci decreased due to the addition of herbs, while high salt content suppressed the growth of *Leuconostoc* and *Pseudomonas* (Quigley et al., 2012).

The microbial profile of cheeses can differ between producers, according to the process of production and the inclusion, or absence, of starters. Plaisentif is a traditional Italian cheese

made from fresh full-fat raw cow's milk without an added starter but, rather, milk from the previous evening (kept at <10°C) and bovine liquid rennet are employed. HTS analyses conducted on samples from nine different producers revealed the presence of the genera *Lactococcus, Lactobacillus*, and *Streptococcus* in the core of all samples but also established that these are present in different proportions and are responsible for variations in the niche-specific characteristics of Plaisentif cheese (Dalmasso et al., 2016). Greater homogeneity was apparent in a recent study involving Herve, a PDO cheese from Belgium. In the past, Herve was made from raw milk but, for safety reasons due to the possible presence of *Listeria monocytogenes*, pasteurized milk has begun to be used. The 16S rDNA analysis showed that 95% of the microbial composition was same in both raw and pasteurized cheese; hence the characters of both forms of the cheese are similar. This might be because of consistency in the manufacturing process across artisanal producers (Delcenserie et al., 2014). The impact of temperature was also evident when the scalding temperature during the manufacture of Dutch-type cheeses was increased from 37 to 39°C. This change resulted in an increase in *Lactobacillus* sp. and *Leuconostoc* sp. during ripening. This change also promoted flavor formation by lysing lactococci, which release enzymes like peptidases (Porcellato and Skeie, 2016). The influence of manufacturing processes on bacterial communities was also noted in the case of Oscypek cheese. In that instance, a decrease in the counts of enterobacteria was apparent from the curd to the smoked cheese as a result of the smoking process, thus improving the quality and safety of the traditional product (Alegría et al., 2012). Cotija is another example of a cheese in which fluctuations in composition have been observed. This handmade Mexican cheese is prepared from raw milk with no added starters. Ripening of Cotija is performed in an open environment and, thus, humidity, rain, and temperature are important parameters to be considered. Furthermore, the wooden surfaces used for kneading the curd while salting, vats, and other tables used in the process of cheese-making all act as a source of the microbes found in this cheese. About 80% of the bacterial population of Cotija was consisted of a combination of *Lactobacillus plantarum*, *Leuconostoc mesenteroides*, and *Weissella paramesenteroides*. These species are thought to be responsible for the development of authentic flavor compounds arising from their lipolytic and proteolytic activity during ripening (Escobar-Zepeda et al., 2016). It is thought that the wooden surfaces are most likely the source of the *Leuconostoc* and *Weissella* species present (Settanni et al., 2012). It is important to note, however, that even when cheeses are produced on a larger industrial scale, including the use of starters to minimize variability, differences in the associated microbiota can be seen. This was apparent from a study of different batches of brine-salted continental-type cheeses made during the same day of production. In this study, 16S rRNA sequencing revealed that cheese produced later in the production day had a more diverse bacterial composition than cheese produced earlier, possibly due to the accumulation of bacteria in the system before cleaning-in-place (CIP) processes are employed at the end of the manufacturing day (O'Sullivan et al., 2015a).

Cheese rinds have a very complex microbiota when compared to core samples; this microbiota varies between types of rinds, degrees of ripening, and the environmental conditions. HTS analyzes of 11 artisanal Irish cheeses (soft, hard, and semi hard) revealed the presence of 19 genera, of which *Lactococcus*, *Leuconostoc*, and *Lactobacillus* were in both rind and core samples. *Corynebacterium, Facklamia, Flavobacterium,* and *Cronobacter* were detected in rind samples only. The relative proportions of lactococci were found to be high in naturally ripened rinds relative to smear/washed rinds. This is an ultimate result of the washing of rind in that cheese (Quigley et al., 2012). In one seminal study, 137 different rind samples that included bloomy, natural, and washed types from 10 countries were examined using HTS to understand the mechanisms of how microbial communities form multispecies biofilms. It was noted that rind samples from cheeses made at different geographical regions showed similar patterns of microbial clusters, showing that, in this instance, distance did not have impact on microbial community. Some of the specific patterns noted were as follows; the fungus *Galactomyces*, which is responsible for the formation of a dense white rind on bloomy rind cheeses (e.g., Brie, Camembert), is positively correlated with moisture. Natural rind cheeses (e.g., Clothbound Cheddar, St. Nectaire, and Tomme de Savoie) are abundant in fungi and bacteria, such as *Scopulariopsis, Aspergillus, Actinobacteria,* and *Staphylococcus*, which are negatively correlated with moisture. The microbial composition of washed rind cheese (e.g., Gruyere, Epoisses) was found to be a mix between that of bloomy and natural rinds (Wolfe et al., 2014). Finally with regard to cheese rinds, it is also worth noting that HTS is beginning to be combined with other techniques, such as microscopy, to identify the fungal and bacterial interactions, and the bacterial dispersal on fungal networks (Zhang et al., 2018).

As already noted above, many artisanal cheeses are produced without the addition of starter cultures. Undefined natural milk starter cultures (NMC), produced by heat treatment of raw milk (60–63°C, 20–30 min) followed by incubation at high temperatures (39–42°C), are used to make some Italian PDO cheeses. Such cultures, prepared by backslopping (i.e., inoculating milk with a previous batch of the culture) using raw milk collected from pasta filata cheese plants, were found to be dominated by *Streptococcus thermophilus* and *Lactobacillus delbrueckii* (Parente et al., 2016). Natural whey cultures (NWC), produced by incubating the whey from previous batches of cheese production at high temperature (39–54°C), is also used to make some PDO Italian cheeses. These NWCs were dominated by *L. lactis, Lactobacillus fermentum, S. thermophilus, L. delbrueckii*, and *Lactobacillus helveticus*, though the abundance of these species was found to be significantly different across the three cheese types, Mozzarella, Grana Padano, Parmigiano Reggiano (De Filippis et al., 2014). The combined pressure of temperature and pH selects for these bacteria. In the case of high-moisture Mozzarella cheese, the mode of acidification, i.e., through addition of citric acid, by starters, or by using undefined starters, greatly influences the growth of the dominant microbial components, thereby differentiating cheeses produced by different dairies (Guidone et al., 2016). This study demonstrated that HTS can be used by Mozzarella producers to

identify the dominant bacteria present and, in turn, helps them to choose the mode of acidification that is best suited to obtaining specific desirable characteristics. Pico is an artisanal Azorean cheese, made from raw milk, animal rennet, and salt without any starters. HTS analysis of this cheese by Riquelme et al. (2015) established that Pico manufacture is a *Lactococcus*-driven process with contributions from accompanying *Lactobacillus* and *Gammaproteobacteria* populations. The availability of this information is useful in terms of selecting starter/adjunct cultures specific for Pico cheese, as well as highlighting some hygienic and safety measures that need to be taken by industries to improve the shelf life of these cheeses (Riquelme et al., 2015). From a safety perspective, HTS detection of microbes in Danish raw milk found that *L. helveticus, S. thermophiles,* and *L. lactis* are dominant (Masoud et al., 2011) and, following spiking with *Listeria innocua* and *Staphylococcus aureus* during manufacture of this cheese to detect the fate of pathogenic bacteria during ripening, it was found that these species were absent from the ripened cheese. The control of these microbes was presumed to be due to acidification, environmental conditions or antimicrobials produced by the LAB (Masoud et al., 2012).

Serpa is an artisanal PDO Portuguese cheese, made from raw ewe's milk using aqueous infusion of *Cynara cardunculus* L. (artichoke), without starters. The fungal communities in this cheese are thought to have a key role in developing the associated organoleptic properties. HTS analysis of this fungal community of both PDO and non-PDO forms of Serpa cheeses, highlighted diversity among the yeasts present in a manner suggested that production practices and the associated environment have a considerable influence on the final yeast population. Through these investigations it was possible to alter ripening conditions to favor the dominance of yeast species that are desirable for optimal Serpa production (Gonçalves Dos Santos et al., 2017). Another ewe's milk cheese produced by traditional techniques without starters, in this case from Croatia, has also been the subject of HTS-based analysis. This revealed the presence of pathogenic bacteria in fresh milk and cheese. However, their abundance was low at the end of ripening, highlighting the importance of ripening for 90 days to avoid problems related to the safety of the cheese (Fuka et al., 2013). This approach highlighted the merits of using HTS to amend production processes to alter quality and safety, particularly in instances where neither a starter nor pasteurized milk are used. An alternative, HTS-based approach to studying the microbiota of cheese has been to use single molecular real-time (SMRT) sequencing, a third generation technique that provides longer DNA reads. This technique has been used to sequence a Kazakhstan cheese made using NWC as starters, revealing that *L. lactis, L. helveticus, S. thermophiles,* and *Lactobacillus bulgaricus* were the dominant species in this cheese. Differences between the microbiota of this cheese and artisanal cheeses from Belgium, Italy, and Kalmykia (Li et al., 2017) likely explain the varying characteristics of these cheeses.

The house microbiota that colonize the equipment, brine tanks, vats, wooden surfaces, knives, and other surfaces in production facilities can play a key role in shaping the microbial communities of cheese. This colonization depends on the characteristics of the surfaces, nutrient availability and composition, ability of microbes to form biofilms, ecological factors, and on operators and cleaning processes (Stellato et al., 2015). Studies focusing on environmental microbiota have been conducted on cheese plants in which one company produces two different cheese types (Calasso et al., 2016) and, in another instance, two different facilities that produce the same type of cheese (Bokulich and Mills, 2013). In the former case, it was deduced that *S. thermophilus* colonized, to different degrees, almost all surfaces in the dairy plant and acted as an indirect source of starter inoculation (Calasso et al., 2016). In the latter study, it was revealed that the type of processing facility and selective forces in the environment, such as the temperature in the cheese plants, influenced the growth of specific house microbiota, which were in turn responsible for the chemosensory properties of the artisanal cheese (Bokulich and Mills, 2013).

One of the factors that can contribute to differences across different production facilities is brining. The microbial diversity and composition of brine depends on the type of cheese made, the specific cheese plant and on salinity concentrations. Adventitious bacteria present in brine, such as *Staphylococcus equorum*, which can have a strong antibacterial activity against *L. monocytogenes* on cheese surfaces, can play important roles in the development of the flavor and color properties of smear ripened cheeses. However, contaminated brine can in turn contaminate the cheese core and surface, with Mozzarella cheese being among the cheeses that are susceptible to such spoilage. Indeed, as a consequence, it has been suggested that sanitization of brine should be prioritized and that brine should be checked regularly for the presence of contaminants (Marino et al., 2017). Cleaning is an important factor to eliminate spoilage bacteria present on the surfaces in dairy plant, with the type of surface material having a key influence on bacterial adherence. Surface materials such as plastic, in the case of gaskets used for the molding of some cheese, is more porous and not suitable for thorough cleaning as it is sensitive to hot water, which causes corrosion and increases the possibility of bacterial adherence (Stellato et al., 2015). As a consequence, it is recommended that steel be used where possible to enhance the safety of products. The importance of proper washing during the manufacture of fermented foods was also highlighted by Lee et al. (2017), who reported that pathogens were replaced by harmless bacteria, such as LAB, as a result of washing. It has been speculated that an equilibrium may exist between the dairy food production and processing environments and the resultant products, with microbial transfer occurring in both directions, which can affect processing dynamics and the quality of the final products (Stellato et al., 2015).

IDENTIFICATION OF PREVIOUSLY OVERLOOKED BACTERIA IN CHEESE

Depending on the depth of sequencing, HTS-based taxonomic analysis of cheese can detect microbes present at very low levels (O'Sullivan et al., 2015b; Cotter and Beresford, 2017), which would previously have been overlooked. While, in

many cases, it is not yet clear what roles these microbes play, if any, in the context of cheese microbiology, an initial awareness of their presence can lead to further investigations. Here we are highlighting some such subdominant populations as revealed in a selection of representative studies. In one such study, a subdominant taxon was identified after culture-based enrichment. Enriching a food sample in broth for 24 h followed by selective agar plating is a traditional method to isolate subdominant bacteria, and is an especially relevant technique for the detection of foodborne pathogens, in the sample. However, overnight growth in non-selective broth can result in a bias toward the detection of fast-growing, relative to slow-growing, bacteria. In this instance HTS of overnight enriched Queso Fresco cheese sample did successfully reveal the presence of *Exiguobacterium*, which at that point had not previously been documented in cheese (Lusk et al., 2012). It should be noted, however, that the majority of HTS studies take place without an enrichment step. In one typical study, it was noted that Plaisentif cheese contains some rare genera, representing 0.01–0.0001% of total reads, that correspond with *Flavobacterium, Brevibacterium, Salinicoccus, Vagococcus, Anaerobacillus, Sphingobacterium*, and *Klebsiella*, among others (Dalmasso et al., 2016), with *Klebsiella* in particular being regarded as indicative of poor hygiene conditions during manufacturing. In contrast, sub-dominant genera found in Cotija cheese include the halophilic microbes *Dehalobacter, Desulfohalobium, Halomonas, Thermohalobacter,* and *Haloquadratum* (Escobar-Zepeda et al., 2016). Anaerobic bacteria, such as *Prevotella* and *Faecalibacterium*, that are more typically associated with gastrointestinal environments, were for the first time identified in Irish artisanal cheeses at low levels in 2012 (Quigley et al., 2012). Some other recent "firsts" include the detection of *Idiomarina* and two candidate divisions, namely GNO2 and TM7, in brine samples. Of these, *Idiomarina* is consider to be a contaminant of brine, with salt acting as the source for this bacteria (Marino et al., 2017). *Tremellomycetes*, also known as jelly fungi, were also recently found on wooden shelves used for ripening (Guzzon et al., 2017). While the importance of many of these microbes in the context of cheese microbiology is not clear, the fact that HTS has detected the presence of these and other previously-overlooked taxa means that further investigations of such taxa in cheese can now take place.

BIOTYPE DIVERSITY AND FUNCTIONAL POTENTIAL OF CHEESE

Assessments of biotype diversity among a species enable researchers to detect variation at sub-species level. This can be achieved by HTS analysis of species-specific genes. One approach to determine biotype diversity involved analyzing a species specific gene with sequence heterogeneity across three Italian cheeses, i.e., Mozzarella, Grana Padano, and Parmigiano Reggiano. *S. thermophilus* was the most abundant bacterium in the three cheeses and, to assess species level diversity, the *lacS* gene, encoding a lactose permease, was selected for specific analysis. Ultimately, 28 different sequence types were identified among which 13 were present at relative abundance > 1%. Mutations at 60 positions identified in the promoter region upstream of *lacS* allowed for further differentiation among the 28 sequence types (De Filippis et al., 2014). A similar sub-species differentiation of *S. thermophilus* has been performed, but instead using the phosphoserine phosphatase gene (serB) as a target sequence. In this case, the approach highlighted the presence of six different sequence types in an undefined milk starter culture (Parente et al., 2016).

Shotgun metagenomic sequencing followed by bioinformatic analysis provides the opportunity to study specific pathways encoded within the cheese microbiome. Ultimately, this technology provides an insight into the functional potential of cheese microbes by studying genes that encode enzymes involved in the catabolic and conversion reactions of amino acids relating to the development of flavor, aroma, and a broad range of other features of relevance to cheese physics, chemistry, and biology. With regard to microbial function or functional potential, spatial distribution analysis of the metabolically active microbiota of three Italian PDO cheeses, Fiore Sardo, Pecorino Siciliano, and Pecorino Toscano, showed a correlation between mesophilic lactobacilli (*L. plantarum*) and secondary proteolysis as well as the synthesis of volatile components (VOC), such as esters, alcohols, aldehydes, and sulfur compounds. Thermophilic LAB found in Pecorino Siciliano cheese correlated with total free amino acid (FAA) concentrations and *Brevibacterium* present on the surface of Pecorino Toscano were found to be related to the synthesis of volatile sulfur compounds, such dimethyl trisulphide, dimethyl disulphide, methional, and S-methyl thioesters (De Pasquale et al., 2016). Shotgun sequencing also revealed a strong correlation between bacteria present in commercially available smear cultures and VOC in surface ripened cheeses. In these instances *B. linens, Geotrichum candidum*, and *Staphylococcus xylosus* correlated with levels of sulfur compounds and 2-methyl-1-butanol, *Debaryomyces hansenii* correlated with levels of alcohols and *Glutamicibacter arilaitensis* with ketones, alcohols, and acids (Bertuzzi et al., 2018). Similarly, a study on Kazak, an artisanal cheese made from fermentation of fresh cow milk through a process involving goat skin bags, identified positive and negative correlations between the core microbiota and amino acid, VOC, and fatty acid concentrations. Kazak cheese analysis revealed that the amino acids, glutamic acid (Glu), histidine (His), isoleucine (Ile), and proline (Pro) had higher positive correlation with one or more abundant bacteria, i.e., *Acetobacter, Lactococcus, Bacillus, Staphylococcus, Kurthia,* and *Moraxella*. Among fungi, *Dipodascus* correlated with concentrations of 9-octadecenoic acid. *Pichia* and *Penicillium* correlated with (Z)-9-hexadecenoic acid, *Issatchenkia* and *Candida* were correlated with leucine (Leu) and phenylalanine (Phe) levels. Various correlations with VOCs were also evident (Zheng et al., 2018). Through harnessing of this knowledge it may be possible to enhance flavor and fat content by choosing specific strains for cheese manufacture. Another multi-omic study focussed on Canestrato Pugliese, an Italian PDO cheese made from raw ewe's milk. HTS revealed that *Lactococcus* dominated within 3 days of ripening and showed a high capacity to degrade carbon sources. Over the ripening period mesophilic lactobacilli, responsible for proteolysis, increased. As part of the

study, the authors developed a model system using HTS together with the Biolog Eco-microplate phenotype method that was designed to select adventitious non-starter lactic acid bacteria (NSLAB) to use as adjunct cultures that guarantee high quality standards for traditional cheese (De Pasquale et al., 2014). A shotgun metagenomic sequencing approach was also employed in the pioneering study of washed rind cheese samples from Europe and USA referred to previously. The metabolism of cysteine and methionine produce volatile sulfur compounds such as methanethiol, and the degradation of valine, leucine, and isoleucine give these cheeses a sweaty putrid aroma. Bloomy and washed rind cheeses were found to contain the halotolerant γ-proteobacteria, *Pseudoalteromonas*, a bacterial genus originally associated with marine environments. *Pseudoalteromonas* were found to possess a gene predicted to encode methionine-gamma-lyase (MGL), an enzyme that converts L-methionine to the aforementioned methanethiol. Among cheese microbes, this enzyme had only been found in *Brevibacterium linens* previously. *Pseudoalteromonas* spp. produce cold-adapted enzymes that participate in lipolysis and proteolysis, a feature that is regarded as beneficial in cheese aged and stored at low temperatures as it leads to the development of flavor compounds (Wolfe et al., 2014). Finally, with regard to flavor, pathways involved in the production of flavor and aroma compounds were captured through metagenomic analysis of Cotija cheese (Escobar-Zepeda et al., 2016). The microbes present encode enzymes with the potential to catabolise phenylalanine and branched chain amino acids to yield flavor compounds. The study also detected genes encoding enzymes involved in free fatty acids catabolism, such as carbonyl reductase I and phenol-2 monoxygenase (an enzyme responsible for the production alcohol from xylene) which results in products that impact a characteristic aroma to traditional cheeses. The genetic determinants for antimicrobial compounds such as bacteriocins produced by strains of LAB, which can play an important role in food biopreservation, were also present (Escobar-Zepeda et al., 2016).

Finally, shotgun metagenomic analysis of cheeses has also been employed to determine the microbial basis for cheese colors. Such analyzes of cheeses with a pink discoloration defect revealed the presence of genes that codes metabolic compounds responsible for the production of carotenoids, that were absent from control cheeses. The presence of this pathway was associated with microbes from the genus *Thermus*, which are renowned for their ability to produce carotenoids but had not been thought to be important in the industrial context of the cheese microbiota previously (Quigley et al., 2016). The aforementioned Bertuzzi et al. study also revealed that surface ripened cheese made from commercially available smear cultures have a greater potential for the production of carotenoids responsible for red/orange colors on the surface of cheeses, highlighting the key role of these smear culture-associated microbes in color development (Bertuzzi et al., 2018).

The next logical step beyond shotgun metagenomics (to assess functional potential) is the use of metatranscriptomics to identify genes that are actually expressed. Metatranscriptomic analysis of Camembert cheese, in which *G. candidum* and *Penicillium camemberti* play key roles in ripening, and the *de novo* assembly of the reads was employed to determine at what stage of cheese ripening are pathways associated with the production of flavor, aroma, and texture development expressed. The study found that these fungal species, already known to contribute to the appearance of the cheese, also play a key role in the development of sensory characteristics of Camembert (Lessard et al., 2014). Metatranscriptomics has also been employed to assess the impact of environmental conditions on gene expression in Caciocavallo Silano, a traditional Italian cheese. These investigations arose from observations that increasing the temperature and decreasing the humidity of the ripening room enhanced the growth of NSLAB in the cheese. Cheese ripened at a higher temperature for 20 days had a similar metabolic profile to that of cheese ripened at standard conditions over a long period, highlighting that increase in ripening temperature results in the rapid maturation of a cheese of quality, while helping the manufacturer to reduce operating costs and to increase the turnover of ripening rooms. The explanation for this reduced ripening time became apparent when it was established that higher temperatures boosted the expression of genes involved in proteolysis, lipolysis, amino acid and fatty acid catabolism and, thus, promoted the VOC related to flavor and aroma (De Filippis et al., 2016). Another such study focussed on a Reblochon-style French cheese, rind activity in which is mostly influenced by *S. thermophilus, L. delbrueckii* spp.*bulgaricus*, *G. candidum*, *D. hansenii*, and *Brevibacterium aurantiacum*. Metatranscriptomic analysis of this cheese during ripening showed the up-regulation of *S. thermophilus* genes associated with protein folding, sorting, degradation, and signal transduction. Transcription, translation as well as nucleotide and lipid metabolism was high among *L. delbrueckii*. Among the yeasts (*D. hansenii*, *G. candidum*) for *G. candidum* a large increase in the transcription of genes associated with carbohydrate, lipid, and amino acid metabolism was evident, for *D. hansenii* an increase in genes associated with the metabolism of other amino acids were noted. A decrease in the expression of gene associated with nucleotide metabolism was observed for both yeasts (Monnet et al., 2016). Metatranscriptomics can also be combined with other technologies to provide an even greater insight into the microbiology of cheese. Indeed, a combination of biochemical, metagenomic, and metatranscriptomic analysis has been used to study the microbiology of surface ripened cheese. In this particular study it became clear that dominant bacteria such as *L. lactis, Kluyveromyces lactis, D. hansenii, G. candidum, Corynebacterium casei,* and *Hafnia alvei* showed high levels of expression of transcripts related to primary (lactose metabolism, lipolysis, proteolysis) and secondary (catabolism of free amino acids and fatty acids) biochemical reactions (Dugat-Bony et al., 2015). It is anticipated that metatranscriptomic-based analysis of the cheese ripening process will become increasingly common in the near future.

HTS REVEALS THE PRESENCE OF PATHOGENIC AND SPOILAGE BACTERIA IN CHEESE

HTS can also be employed to detect pathogenic or spoilage microbes in cheeses, milk, or in the production and processing

environments. Numerous sources of contamination exist. These include cow feces and teat surfaces (Doyle et al., 2017a; Frétin et al., 2018) as well as bulk tanks and milking machines can also act as sources, such as in the case of the presence of *Enterococcus faecalis* in raw Herve cheese (Delcenserie et al., 2014).

The consequences of the presence of different microbes are various. The appearance of pink discoloration in cheese manifests through the appearance of pink patches at various locations within the ripened cheese block. As noted above, HTS revealed the presence of spoilage bacteria *Thermus thermophilus*, a species associated with the production of carotenoids, within defective cheeses (Quigley et al., 2016). The genera *Brevibacterium, Corynebacterium,* and *Microbacterium* have been associated with a red-brown defect in smear-ripened cheese, such as Fontina. Deacidification by Debaryomyces facilitates the growth of these bacteria on wooden shelves used for ripening (Guzzon et al., 2017). Analysis of Grana Padano cheeses that exhibited a blowing defect resulted in the detection of seven species of clostridia, i.e., *C. sporogenes, C. butyricum, C. disporicum, C. perfringens, C. difficile, C. sordelii,* and *C. tyrobutyricum*. Among these *C. tyrobutyricum* and *C. butyricum* were more abundant in blowing defect samples (Bassi et al., 2015). In another study, the presence of genera such as *Klebsiella, Morganella, Erwinia,* and *Acinetobacter*, i.e., taxa commonly found in soil and water, in cheese and curd samples of Plaisentif cheese was taken to represent poor hygiene conditions and the contamination of boilers or tools used in the early stages of processing (Dalmasso et al., 2016). Environmental contaminants, including *Escherichia* sp., *Enterobacter cowanii* and other *Enterobacteriaceae*, *Agrobacterium* sp., *Alicyclobacillus* sp., and *Propionibacterium acnes* were also detected in NWC used to make Italian cheeses (De Filippis et al., 2014) and the detection of *Pseudomonas* was observed in the intermediates of Mozzarella cheese production was suggested to be related to handling and hygiene conditions (Ercolini et al., 2012).

Another mechanism via which microbes can influence cheese quality is through the production of biogenic amines. Histamine, tyramine, putrescine, and cadaverine are the most common biogenic amines that occur in food, with their accumulation causing symptoms such as hypertension, headaches, palpitations, and vomiting in sensitive individuals. The presence of species such as *Lactobacillus curvatus, Enterococcus faecium,* and *Enterococcus faecalis* can result in the production of tyramine and *Lactobacillus buchneri* has been associated with the production of histamine. A HTS-based approach that specifically detects the decarboxylase genes that are responsible for biogenic amine production has been developed (O'Sullivan et al., 2015b).

One obvious concern relating to the use of DNA-based approaches to microbial detection is the possibility of false positives arising from the detection of DNA from dead microorganisms. The inclusion of steps involving the use of ethidium monoazide (EMA) IUPAC name: 3-amino-8-azido-5-ethyl-6-phenylphenanthridium bromide or propidium monoazide (PMA) IUPAC name: 3-Amino-8-azido-5-{3-[diethyl(methyl)ammonio]propyl}-6- phenylphenanthridinium to inactivate any DNA which is not sourced from living microbes provides a means of addressing this issue in the context of amplicon based sequencing at least (Rudi et al., 2005; Josefsen et al., 2010; Quigley et al., 2013; Porcellato and Skeie, 2016). Ultimately, while, currently, these investigations have been carried out to identify problematic microorganisms or the source thereof in academic laboratories, it is envisaged that the technology will evolve in a manner that will allow its use in quality assurance laboratories across the food industry.

CONCLUSIONS

HTS is a powerful technique that can be used to provide a detailed insight into the microbiology of dairy related samples, including raw and pasteurized milk, cheese curd, whey, starter cultures, and cheese. The technology also allows a determination of the impact of seasonal variations, feed given to animals, source of milk, and other environmental factors relating to milk production and processing on the milk and cheese microbiota. Once the microbial community in cheese has been formed, HTS can then be employed to determine the role of the microbial population in factors such as the development of organoleptic properties. While there are very many advantages associated with the use of HTS, some barriers, such as the detection of DNA sequences from non-viable bacteria, cost effectiveness, and experience in bioinformatic analysis need to be overcome before these benefits can be applied extensively across industry. Furthermore, there is a need for the construction of a comprehensive dairy microbial gene catalogs to improve the analysis of HTS reads in the dairy environment (Almeida et al., 2014). However, given that solutions to these issues exist or are in development, it is hoped that any delays to the industrial application of this technology will be short.

AUTHOR CONTRIBUTIONS

BY and PC prepared the manuscript and all co-authors contributed to editing and critical reviewing thereof.

FUNDING

This work was funded by the Teagasc Walsh fellowship programme.

REFERENCES

Alegría, Á., Szczesny, P., Mayo, B., Bardowski, J., and Kowalczyk, M. (2012). Biodiversity in Oscypek, a traditional Polish Cheese, determined by culture-dependent and -independent approaches. *Appl. Environ. Microbiol.* 78, 1890–1898. doi: 10.1128/AEM.06081-11

Almeida, M., Hébert, A., Abraham, A. L., Rasmussen, S., Monnet, C., Pons, N., et al. (2014). Construction of a dairy microbial genome catalog opens new perspectives for the metagenomic analysis of dairy fermented products. *BMC Genomics* 15:1101. doi: 10.1186/1471-2164-15-1101

Bassi, D., Puglisi, E., and Cocconcelli, P. S. (2015). Understanding the bacterial communities of hard cheese with blowing defect. *Food Microbiol.* 52, 106–118. doi: 10.1016/j.fm.2015.07.004

Bertuzzi, A. S., Walsh, A. M., Sheehan, J. J., Cotter, P. D., Crispie, F., Mcsweeney, P. L. H., et al. (2018). Omics-based insights into flavor development and microbial succession within surface-ripened cheese. *mSystems* 3:e00211–e00217. doi: 10.1128/mSystems.00211-17

Bokulich, N. A., Lewis, Z. T., Boundy-Mills, K., and Mills, D. A. (2016). A new perspective on microbial landscape within food production. 37, 182–189. doi: 10.1016/j.copbio.2015.12.008

Bokulich, N. A., and Mills, D. A. (2013). Facility-specific "house" microbiome drives microbial landscapes of artisan cheesemaking plants. *Appl. Environ. Microbiol.* 79, 5214–5223. doi: 10.1128/AEM.00934-13

Calasso, M., Ercolini, D., Mancini, L., Stellato, G., Minervini, F., Di Cagno, R., et al. (2016). Relationships among house, rind and core microbiotas during manufacture of traditional Italian cheeses at the same dairy plant. *Food Microbiol.* 54, 115–126. doi: 10.1016/j.fm.2015.10.008

Clooney, A. G., Fouhy, F., Sleator, R. D., O'Driscoll, A., Stanton, C., Cotter, P. D., et al. (2016). Comparing apples and oranges? Next generation sequencing and its impact on microbiome analysis. *PLoS ONE* 11:e0148028. doi: 10.1371/journal.pone.0148028

Cotter, P. D., and Beresford, T. P. (2017). "Chapter 15 – Microbiome changes during ripening," in *Cheese 4th Edn Chemistry, Physics and Microbiology (Reading,. Engl)*, eds P. L. H. McSweeney, P. F. Fox, P. Cotter, and D. W. Everett (Elsevier), 389–409.

Dalmasso, A., de los Dolores Soto del Rio, M., Civera, T., Pattono, D., Cardazzo, B., and Bottero, M. T. (2016). Characterization of microbiota in Plaisentif cheese by high-throughput sequencing. *LWT Food Sci. Technol.* 69, 490–496. doi: 10.1016/j.lwt.2016.02.004

De Filippis, F., Genovese, A., Ferranti, P., Gilbert, J. A., Ercolini, D., and Gallo, P. (2016). Metatranscriptomics reveals temperature-driven functional changes in microbiome impacting cheese maturation rate. *Sci. Rep.* 6:21871. doi: 10.1038/srep21871

De Filippis, F., La Storia, A., Stellato, G., Gatti, M., and Ercolini, D. (2014). A selected core microbiome drives the early stages of three popular italian cheese manufactures. *PLoS ONE* 9:e89680. doi: 10.1371/journal.pone.0089680

De Filippis, F., Parente, E., and Ercolini, D. (2017). Metagenomics insights into food fermentations. *Microb. Biotechnol.* 10, 91–102. doi: 10.1111/1751-7915.12421

Delcenserie, V., Taminiau, B., Delhalle, L., Nezer, C., Doyen, P., Crevecoeur, S., et al. (2014). Microbiota characterization of a Belgian protected designation of origin cheese, Herve cheese, using metagenomic analysis. *J. Dairy Sci.* 97, 6046–6056. doi: 10.3168/jds.2014-8225

De Pasquale, I., Calasso, M., Mancini, L., Ercolini, D., La Storia, A., De Angelis, M., et al. (2014). Causal relationship between microbial ecology dynamics and proteolysis during manufacture and ripening of Protected Designation of Origin (PDO) cheese Canestrato Pugliese. *Appl. Environ. Microbiol.* 80, 4085–4094. doi: 10.1128/AEM.00757-14

De Pasquale, I., Di Cagno, R., Buchin, S., De Angelis, M., and Gobbetti, M. (2016). Spatial distribution of the metabolically active microbiota within Italian PDO ewes' milk cheeses. *PLoS ONE* 11:e0153213. doi: 10.1371/journal.pone.0153213

Di Bella, J. M., Bao, Y., Gloor, G. B., Burton, J. P., and Reid, G. (2013). High throughput sequencing methods and analysis for microbiome research. *J. Microbiol. Methods* 95, 401–414. doi: 10.1016/j.mimet.2013.08.011

Doyle, C. J., Gleeson, D., O'Toole, P. W., and Cottera, P. D. (2017a). Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study. *Appl. Environ. Microbiol.* 83:e02694–16. doi: 10.1128/AEM.02694-16

Doyle, C. J., O'Toole, P. W., and Cotter, P. D. (2017b). Metagenome-based surveillance and diagnostic approaches to studying the microbial ecology of food production and processing environments. *Environ. Microbiol.* 19, 4382–4391. doi: 10.1111/1462-2920.13859

Dugat-Bony, E., Straub, C., Teissandier, A., Onésime, D., Loux, V., Monnet, C., et al. (2015). Overview of a surface-ripened cheese community functioning by Meta-Omics analyses. *PLoS ONE* 10:e0124360. doi: 10.1371/journal.pone.0124360

Ercolini, D. (2013). High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Appl. Environ. Microbiol.* 79, 3148–3155. doi: 10.1128/AEM.00256-13

Ercolini, D., De Filippis, F., La Storia, A., and Iacono, M. (2012). "Remake" by high-throughput sequencing of the microbiota involved in the production of water Buffalo mozzarella cheese. *Appl. Environ. Microbiol.* 78, 8142–8145. doi: 10.1128/AEM.02218-12

Escobar-Zepeda, A., Sanchez-Flores, A., and Quirasco Baruch, M. (2016). Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota. *Food Microbiol.* 57, 116–127. doi: 10.1016/j.fm.2016.02.004

Fox, P. F., Guinne, T. P., Cogan, T. M., and McSweeney, P. L. H. (2000). "Microbiology of cheese ripening," in *Fundamentals of Cheese Science*, eds P. F. Fox, T. P. Guinee, T. M. Cogan, and P. L. H. McSweeney (Boston, MA: Springer US), 206–235.

Frétin, M., Martin, B., Rifa, E., Isabelle, V.-M., Pomiès, D., Ferlay, A., et al. (2018). Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems. *Sci. Rep.* 8:200. doi: 10.1038/s41598-017-18447-y

Fuka, M. M., Wallisch, S., Engel, M., Welzl, G., Havranek, J., and Schloter, M. (2013). Dynamics of bacterial communities during the ripening process of different Croatian cheese types derived from raw ewe's milk cheeses. *PLoS ONE* 8:e80734. doi: 10.1371/journal.pone.0080734

Galimberti, A., Bruno, A., Mezzasalma, V., De Mattia, F., Bruni, I., and Labra, M. (2015). Emerging DNA-based technologies to characterize food ecosystems. *Food Res. Int.* 69, 424–433. doi: 10.1016/j.foodres.2015.01.017

Giello, M., La Storia, A., Masucci, F., Di Francia, A., Ercolini, D., and Villani, F. (2017). Dynamics of bacterial communities during manufacture and ripening of traditional Caciocavallo of Castelfranco cheese in relation to cows' feeding. *Food Microbiol.* 63, 170–177. doi: 10.1016/j.fm.2016.11.016

Gonçalves Dos Santos, M. T. P., Benito, M. J., Córdoba, M., de, G., Alvarenga, N., and Ruiz-Moyano Seco de Herrera, S. (2017). Yeast community in traditional Portuguese Serpa cheese by culture-dependent and -independent DNA approaches. *Int. J. Food Microbiol.* 262, 63–70. doi: 10.1016/j.ijfoodmicro.2017.09.013

Guidone, A., Zotta, T., Matera, A., Ricciardi, A., De Filippis, F., Ercolini, D., et al. (2016). The microbiota of high-moisture mozzarella cheese produced with different acidification methods. *Int. J. Food Microbiol.* 216, 9–17. doi: 10.1016/j.ijfoodmicro.2015.09.002

Guzzon, R., Carafa, I., Tuohy, K., Cervantes, G., Vernetti, L., Barmaz, A., et al. (2017). Exploring the microbiota of the red-brown defect in smear-ripened cheese by 454-pyrosequencing and its prevention using different cleaning systems. *Food Microbiol.* 62, 160–168. doi: 10.1016/j.fm.2016.10.018

Josefsen, M. H., Löfström, C., Hansen, T. B., Christensen, L. S., Olsen, J. E., and Hoorfar, J. (2010). Rapid quantification of viable Campylobacter bacteria on chicken carcasses, using real-time pcr and propidium monoazide treatment, as a tool for quantitative risk assessment. *Appl. Environ. Microbiol.* 76, 5097–5104. doi: 10.1128/AEM.00411-10

Kelleher, P., Murphy, J., Mahony, J., and Van Sinderen, D. (2015). Next-generation sequencing as an approach to dairy starter selection. *Dairy Sci. Technol.* 95, 545–568. doi: 10.1007/s13594-015-0227-4

Lee, H. W., Yoon, S. R., Kim, S. J., Lee, H. M., Lee, J. Y., Lee, J. H., et al. (2017). Identification of microbial communities, with a focus on foodborne pathogens, during kimchi manufacturing process using culture-independent and -dependent analyses. *LWT Food Sci. Technol.* 81, 153–159. doi: 10.1016/j.lwt.2017.04.001

Lessard, M. H., Viel, C., Boyle, B., St-Gelais, D., and Labrie, S. (2014). Metatranscriptome analysis of fungal strains *Penicillium camemberti* and *Geotrichum candidum* reveal cheese matrix breakdown and potential development of sensory properties of ripened Camembert-type cheese. *BMC Genomics* 15:235. doi: 10.1186/1471-2164-15-235

Li, J., Zheng, Y., Xu, H., Xi, X., Hou, Q., Feng, S., et al. (2017). Bacterial microbiota of Kazakhstan cheese revealed by single molecule real time (SMRT) sequencing and its comparison with Belgian, Kalmykian and Italian artisanal cheeses. *BMC Microbiol.* 17:13. doi: 10.1186/s12866-016-0911-4

Lusk, T. S., Ottesen, A. R., White, J. R., Allard, M. W., Brown, E. W., and Kase, J. A. (2012). Characterization of microflora in Latin-style cheeses by next-generation sequencing technology. *BMC Microbiol.* 12:254. doi: 10.1186/1471-2180-12-254

Macori, G., and Cotter, P. D. (2018). Novel insights into the microbiology of fermented dairy foods. *Curr. Opin. Biotechnol.* 49, 172–178. doi: 10.1016/j.copbio.2017.09.002

Marino, M., Innocente, N., Maifreni, M., Mounier, J., Cobo-Díaz, J. F., Coton, E., et al. (2017). Diversity within Italian cheesemaking brine-associated bacterial communities evidenced by massive parallel 16S rRNA gene tag sequencing. *Front. Microbiol.* 8:2119. doi: 10.3389/fmicb.2017.02119

Masoud, W., Takamiya, M., Vogensen, F. K., Lillevang, S., Al-Soud, W. A., Sørensen, S. J., et al. (2011). Characterization of bacterial populations in Danish raw milk cheeses made with different starter cultures by denaturating

gradient gel electrophoresis and pyrosequencing. *Int. Dairy J.* 21, 142–148. doi: 10.1016/j.idairyj.2010.10.007

Masoud, W., Vogensen, F. K., Lillevang, S., Abu Al-Soud, W., Sørensen, S. J., and Jakobsen, M. (2012). The fate of indigenous microbiota, starter cultures, *Escherichia coli, Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. *Int. J. Food Microbiol.* 153, 192–202. doi: 10.1016/j.ijfoodmicro.2011.11.014

Mayo, B., Rachid, C. T., Alegría, Á., Leite, A. M. O., Peixoto, R. S., and Delgado, S. (2014). Impact of next generation sequencing techniques in food microbiology. *Curr. Genomics* 15, 293–309. doi: 10.2174/1389202915666140616233211

Monnet, C., Dugat-Bony, E., Swennen, D., Beckerich, J. M., Irlinger, F., Fraud, S., et al. (2016). Investigation of the activity of the microorganisms in a reblochon-style cheese by metatranscriptomic analysis. *Front. Microbiol.* 7:536. doi: 10.3389/fmicb.2016.00536

O'Sullivan, D. J., Cotter, P. D., O'Sullivan, O., Giblin, L., McSweeney, P. L. H., and Sheehan, J. J. (2015a). Temporal and spatial differences in microbial composition during the manufacture of a continental-type cheese. *Appl. Environ. Microbiol.* 81, 2525–2533. doi: 10.1128/AEM.04054-14

O'Sullivan, D. J., Fallico, V., O'Sullivan, O., McSweeney, P. L. H., Sheehan, J. J., Cotter, P. D., et al. (2015b). High-throughput DNA sequencing to survey bacterial histidine and tyrosine decarboxylases in raw milk cheeses. *BMC Microbiol.* 15:266. doi: 10.1186/s12866-015-0596-0

Parente, E., Guidone, A., Matera, A., De Filippis, F., Mauriello, G., and Ricciardi, A. (2016). Microbial community dynamics in thermophilic undefined milk starter cultures. *Int. J. Food Microbiol.* 217, 59–67. doi: 10.1016/j.ijfoodmicro.2015.10.014

Porcellato, D., and Skeie, S. B. (2016). Bacterial dynamics and functional analysis of microbial metagenomes during ripening of Dutch-type cheese. *Int. Dairy J.* 61, 182–188. doi: 10.1016/j.idairyj.2016.05.005

Quigley, L., O'Sullivan, D. J., Daly, D., O'Sullivan, O., Burdikova, Z., Vana, R., et al. (2016). Thermus and the pink discoloration defect in cheese. *mSystems* 1:e00023–e00016. doi: 10.1128/mSystems.00023-16

Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., and Cotter, P. D. (2012). High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Appl. Environ. Microbiol.* 78, 5717–5723. doi: 10.1128/AEM.00918-12

Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., et al. (2013). The complex microbiota of raw milk. *FEMS Microbiol. Rev.* 37, 664–698. doi: 10.1111/1574-6976.12030

Riquelme, C., Câmara, S., de Enes Dapkevicius, M. N., Vinuesa, P., da Silva, C. C. G., Malcata, F. X., et al. (2015). Characterization of the bacterial biodiversityin Pico cheese (an artisanal Azorean food). *Int. J. Food Microbiol.* 192, 86–94. doi: 10.1016/j.ijfoodmicro.2014.09.031

Rudi, K., Naterstad, K., Drømtorp, S. M., and Holo, H. (2005). Detection of viable and dead Listeria monocytogenes on gouda-like cheeses by real-time PCR. *Lett. Appl. Microbiol.* 40, 301–306. doi: 10.1111/j.1472-765X.2005.01672.x

Settanni, L., Di Grigoli, A., Tornambé, G., Bellina, V., Francesca, N., Moschetti, G., et al. (2012). Persistence of wild *Streptococcus thermophilus* strains on wooden vat and during the manufacture of a traditional Caciocavallo type cheese. *Int. J. Food Microbiol.* 155, 73–81. doi: 10.1016/j.ijfoodmicro.2012.01.022

Stellato, G., Filippis, F. D., Storia, A. L., and Ercolini, D. (2015). Coexistence of lactic acid bacteria and potential spoilage microbiota. *Appl. Environ. Microbiol.* 81, 7893–7904. doi: 10.1128/AEM.02294-15

Walsh, A. M., Crispie, F., Claesson, M. J., and Cotter, P. D. (2017). Translating omics to food microbiology. *Annu. Rev. Food Sci. Technol.* 8, 113–134. doi: 10.1146/annurev-food-030216-025729

Walsh, A. M., Crispie, F., O'Sullivan, O., Finnegan, L., Claesson, M. J., and Cotter, P. D. (2018). Species classifier choice is a key consideration when analysing low-complexity food microbiome data. *Microbiome* 6:50. doi: 10.1186/s40168-018-0437-0

Wolfe, B. E., Button, J. E., Santarelli, M., and Dutton, R. J. (2014). Cheese rind communities provide tractable systems for in situ and *in vitro* studies of microbial diversity. *Cell* 158, 422–433. doi: 10.1016/j.cell.2014.05.041

Zhang, Y., Kastman, E. K., Guasto, J. S., and Wolfe, B. E. (2018). Fungal networks shape dynamics of bacterial dispersal and community assembly in cheese rind microbiomes. *Nat. Commun.* 9: 366. doi: 10.1038/s41467-017-02522-z

Zheng, X., Liu, F., Shi, X., Wang, B., Li, K., Li, B., et al. (2018). Dynamic correlations between microbiota succession and flavor development involved in the ripening of Kazak artisanal cheese. *Food Res. Int.* 105, 733–742. doi: 10.1016/j.foodres.2017.12.007

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Yeluri Jonnala, McSweeney, Sheehan and Cotter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 December 2017
doi: 10.3389/fmicb.2017.02609

Zygosaccharomyces bailii Is a Potential Producer of Various Flavor Compounds in Chinese *Maotai*-Flavor Liquor Fermentation

Yan Xu, Yan Zhi, Qun Wu, Rubing Du and Yan Xu**

State Key Laboratory of Food Science and Technology, Key Laboratory of Industrial Biotechnology, Ministry of Education, Synergetic Innovation Center of Food Safety and Nutrition, School of Biotechnology, Jiangnan University, Wuxi, China

Zygosaccharomyces bailii is a common yeast in various food fermentations. Understanding the metabolic properties and genetic mechanisms of *Z. bailii* is important for its industrial applications. Fermentation characteristics of *Z. bailii* MT15 from Chinese *Maotai*-flavor liquor fermentation were studied. *Z. bailii* MT15 produced various flavor compounds, including 19 alcohols, six acids, three esters, three ketones, and two aldehydes. Moreover, production of acids and aldehydes were increased by 110 and 41%, respectively, at 37°C (the maximum temperature in liquor fermentation) compared with that at 30°C, indicating its excellent flavor productivity. *Z. bailii* MT15 is a diploid with genome size of 20.19 Mb. Comparative transcriptome analysis revealed that 12 genes related to amino acid transport were significantly up-regulated (2.41- to 5.11-fold) at 37°C. Moreover, genes *ARO8*, *ARO9*, and *ALDH4* involved in amino acid metabolism also showed higher expression levels (>1.71-fold) at 37°C. Increased substrate supply and a vigorous metabolism might be beneficial for the increased production of acids and aldehydes at 37°C. This work revealed the potential contribution of *Z. bailii* to various flavor compounds in food fermentation, and produced insights into the metabolic mechanisms of *Z. bailii* in flavor production.

Keywords: ***Zygosaccharomyces bailii,*** **food fermentation, flavor compounds, genome, transcriptome, Chinese *Maotai*-flavor liquor**

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Jyoti Prakash Tamang, Sikkim University, India
Valentina Bernini, Università degli Studi di Parma, Italy

***Correspondence:**
Qun Wu
wuq@jiangnan.edu.cn
Yan Xu
yxu@jiangnan.edu.cn

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *06 July 2017*
Accepted: *14 December 2017*
Published: *22 December 2017*

Citation:
Xu Y, Zhi Y, Wu Q, Du R and Xu Y (2017) Zygosaccharomyces bailii Is a Potential Producer of Various Flavor Compounds in Chinese Maotai-Flavor Liquor Fermentation. Front. Microbiol. 8:2609. doi: 10.3389/fmicb.2017.02609

INTRODUCTION

Zygosaccharomyces bailii is a yeast species widely present in various food fermentations, such as wine, tea, and vinegar fermentations (Teoh et al., 2004; Solieri et al., 2006; Garavaglia et al., 2015). In most food and beverage industries, *Z. bailii* is considered as a problematic spoilage yeast due to its high resistance to preservatives and high tolerance of various stresses (Stratford et al., 2013; Palma et al., 2015). Regardless of its association with spoilage, the potential beneficial effects of *Z. bailii* have also been proposed in food industries (Ciani et al., 2009; Domizio et al., 2011). In wine fermentation, owing to its high production of esters, *Z. bailii* in a mixed starter with *Saccharomyces cerevisiae* improved the production of ethyl esters (Garavaglia et al., 2015). The co-culture of *Z. bailii* with *S. cerevisiae* also increased the production of polysaccharides that improved the taste and body of wine (Domizio et al., 2011). Additionally, *Z. bailii* formed part of the tea fungus in Kombucha and Haipao tea fermentation that were rich in crude protein, crude fiber, and

lysine (Jayabalan et al., 2010). Nevertheless, the metabolic properties and genetic mechanisms of *Z. bailii* in food fermentations are still unclear and need to be systematically studied.

Chinese *Maotai*-flavor liquor is a popular alcoholic beverage even in other parts of the world (Xu and Ji, 2012). This liquor contains over 300 influential flavor compounds including alcohols, acids, esters, ketones, and aldehydes, which greatly contribute to its unique aroma and quality (Xu and Ji, 2012). *Maotai*-flavor liquor is produced from grains by a spontaneous and solid-state fermentation. Yeasts play essential roles in *Maotai*-flavor liquor fermentation (Wu, 2013; Wu et al., 2013). Among them, *S. cerevisiae* is one of the most important and contributes significantly to the quantity and quality of the liquor (Wu et al., 2012; Meng et al., 2015). *Z. bailii* has been found to be a dominant species in *Maotai*-flavor liquor fermentation, with proportions close to those of *S. cerevisiae* (Wu, 2013). Despite its large population, the metabolic activity of *Z. bailii* in the liquor fermentation is still unclear.

The present work is aimed to study the fermentation characteristics of *Z. bailii* MT15, including the genome sequencing and comparative transcriptome analysis to unravel its metabolic mechanisms at the relatively high temperature used in liquor fermentation. The results shed new light on the function of *Z. bailii* and provide a guide for its efficient use in food fermentation.

MATERIALS AND METHODS

Yeast Strains

Z. bailii MT15 and *S. cerevisiae* MT1 were previously isolated from the *Maotai*-flavor liquor fermentation process and were deposited in the China General Microbiological Culture Collection Center with accession number CGMCC 4745 (Xu et al., 2013), and the China Center for Type Culture Collection with accession number CCTCC M2014463 (Meng et al., 2015), respectively.

Fermentation Conditions

Sorghum extract was used as fermentation medium and was prepared according to the method below. Two kilogram of ground sorghum was added to 8 L of deionized water. Then the mixture was steamed for 2 h and subsequently saccharified by glucoamylase (5 U/L) at 60°C for 4 h. The supernatant was collected after being filtered through gauze and centrifuged at 8,000 × *g* for 15 min. The obtained sorghum extract was diluted with water to produce a final reducing sugar concentration of 75 ± 5 g/L before sterilization (Lu et al., 2015). Fermentation media were prepared with aliquots of 50 mL sorghum extract in 250 mL flasks, and were sterilized at 115°C for 30 min.

Z. bailii MT15 and *S. cerevisiae* MT1 were pre-cultured in sorghum extract at 30°C for 16 h to obtain the seed culture. Then they were inoculated into fermentation media with an initial concentration of 1×10^6 colony-forming units (CFU)/mL. For the comparison of fermentation characteristics between *Z. bailii* MT15 and *S. cerevisiae* MT1, the strains were, respectively, fermented at 30°C for 48 h, with shaking at 200 rpm. To study the effects of temperature on fermentation, *Z. bailii* MT15 was fermented at 30 and 37°C for 48 h, with shaking at 200 rpm. During the fermentation, 1 mL of samples were withdrawn at 8-h intervals to determine the cell numbers (**Figure 1A**). Each experiment was performed in triplicate.

Analytical Determinations and Statistical Analysis

Yeast cells in the seed culture and during the fermentation process were obtained by centrifugation, washed three times with sterile saline solution, diluted to the applicable concentration by saline solution, and counted by hemocytometer under microscope. To analyze the contents of ethanol, flavor compounds, and amino acids, fermentation broths were centrifuged at 8,000 × *g* for 10 min to remove cells. Ethanol was monitored by high-performance liquid chromatography (HPLC) with a refractive index detector (Varian 355 RI) (Meng et al., 2015). Flavor compounds were analyzed by the gas chromatography-mass spectrometry method (Kong et al., 2014). Free amino acids were detected by HPLC via a pre-column derivatization method (Zhang et al., 2014).

The data are presented in terms of arithmetic averages of three replicates and the error bars indicate the standard deviations. The statistical significance of the difference between the means of samples was tested by one-way analysis of variance (ANOVA) using SPSS Statistics 22.

Genomic Sequencing and Assembly

The *Z. bailii* MT15 genome was sequenced using the whole-genome shotgun sequencing approach on an Illumina Miseq platform (Quail et al., 2012). Four paired-end/mate-paired sequencing libraries were constructed with insert sizes of 450, 700 bp, 3, and 8 kb. The *de novo* assembler Newbler and SSPACE software packages were employed to assemble raw data into contigs and scaffolds (Boetzer et al., 2011; Nederbragt, 2014). The GapCloser program was used to close gaps (Boetzer and Pirovano, 2012). The genome sequence of *Z. bailii* MT15 was deposited in GenBank under the Whole Genome Shotgun project number SRR5452526.

The genes of *Z. bailii* MT15 were predicted with the CEGMA pipeline, combining Augustus, SNAP, and Glimmer gene prediction software (Delcher et al., 1999; Stanke and Waack, 2003; Korf, 2004; Parra et al., 2007). The functional annotation of each gene was based on the eggNOG and Swissprot databases (Boeckmann et al., 2003; Powell et al., 2012). The functional classifications were performed with the eggNOG database. Genes for tRNAs and rRNAs were predicted with tRNAscan-SE and RNAmmer 1.2 Server, respectively (Schattner et al., 2005; Lagesen et al., 2007).

cDNA Preparation and Transcriptome Analysis Using RNA Sequencing

For RNA extraction, *Z. bailii* MT15 cells were cultivated with three parallel at 30 and 37, respectively, for 24 h. Then, the three parallel were mixed in the same volume and harvested after centrifugation at 8,000 × *g* for 5 min at 4°C. Subsequently, the supernatant was removed and the cell pellet was washed three

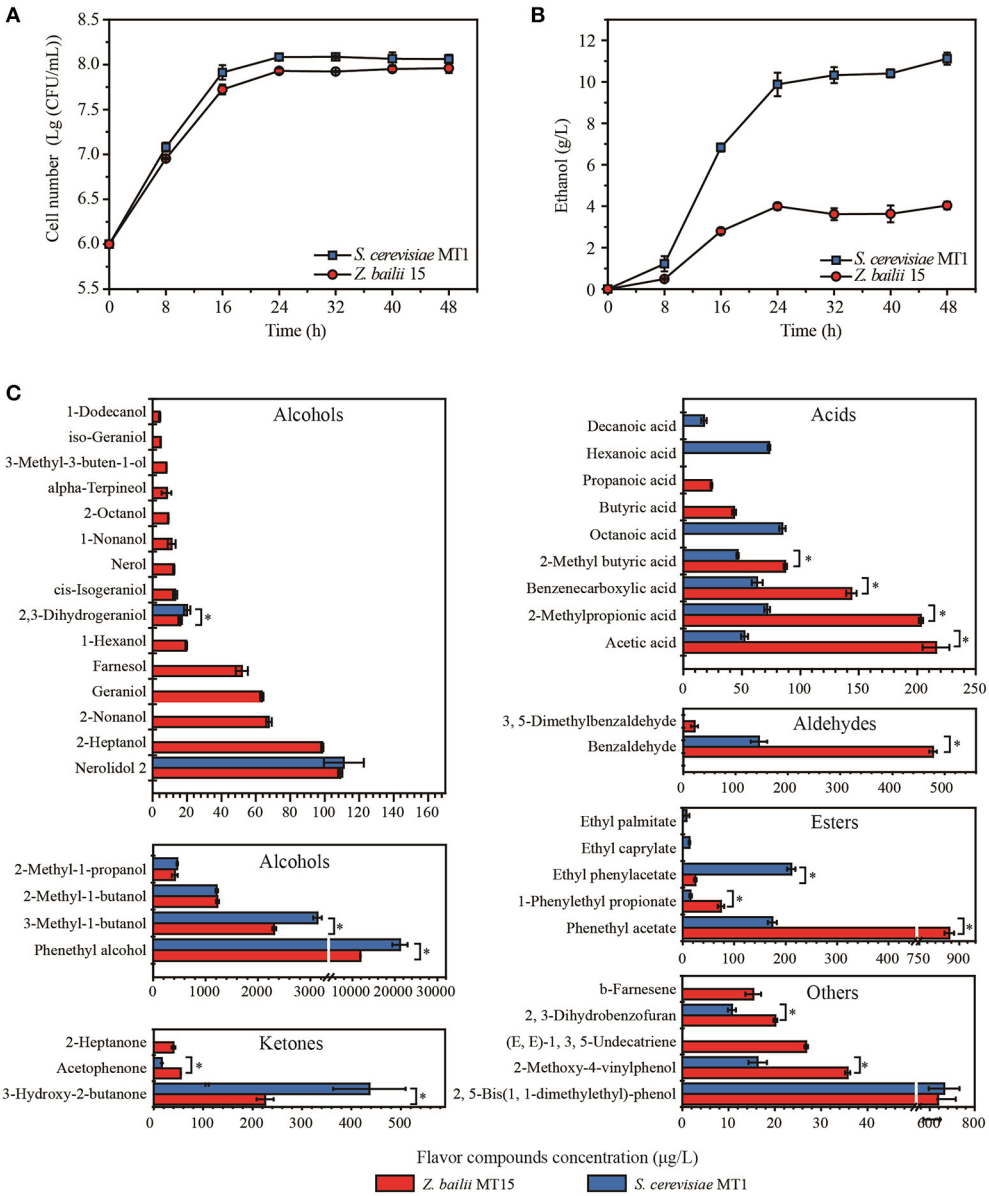


FIGURE 1 | Biomass **(A)**, ethanol content **(B)**, and flavor compounds production **(C)** of *Z. bailii* MT15 and *S. cerevisiae* MT1. Significance level of the one-way ANOVA: $^*P < 0.05$.

times with sterile saline solution on ice. Then the washed cell pellet was immediately frozen in liquid nitrogen. Total RNA was isolated using the Trizol Reagent (Invitrogen Life Technologies, Shanghai) according to the manufacturer's instructions. Quality and integrity of total RNA were determined using a Nanodrop spectrophotometer (Thermo Scientific, USA) and Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA). Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA) was used to remove the ribosomal RNAs. The mRNA was subsequently fragmented and used as a template for oligo (dT)-primed PCR.

The cDNA libraries were prepared employing standard techniques for subsequent Illumina sequencing using the mRNA-seq Sample Prep Kit (Illumina, San Diego, CA). The cDNA libraries were sequenced on an Illumina NextSeq 500 according to the manufacturer's instructions. Sequencing raw reads were pre-processed after filtering sequencing adapters, rRNA reads, short-fragment reads, and other low-quality reads. The remaining clear reads were mapped to the reference genome of *Z. bailii* MT15 using Bowtie2/Tophat2 software (Langmead and Salzberg, 2012; Kim et al., 2015) based on the local alignment algorithm. Gene expression level was normalized by calculating reads per kilobase per million reads (RPKM) (Mortazavi et al., 2008). Differential expression of all of transcripts was quantified using DESeq software, and the method of FDR (False Discovery

Rate) control was used to correct the results for multiple hypothesis testing (Anders and Huber, 2013). Significant DEGs were screened based on an FDR threshold of $\leq$0.05, and a Fold change $\geq$1.5. The RNA sequence data of *Z. bailii* MT15 at 30 and 37°C was deposited in the DNA Data Bank of Japan (DDBJ) with accession IDs DRX082892 and DRX082893, respectively.

RESULTS AND DISCUSSION

Metabolic Properties of *Z. bailii* MT15

The biomass and ethanol production of *Z. bailii* MT15 and *S. cerevisiae* MT1 was determined during the fermentation process. The cell number of *Z. bailii* MT15 and *S. cerevisiae* MT1 was 9.18×10^7 and 8.18×10^7 CFU/mL at the end of fermentation (**Figure 1A**), respectively, which showed no significantly statistical difference (Supplementary Table 1). The highest ethanol production of *Z. bailii* MT15 was 4.03 g/L, which was 63.76% less than that of *S. cerevisiae* MT1 (**Figure 1B**, Supplementary Table 2). Considering the large population of *Z. bailii* MT15, whose proportion can reach 78% of the total yeast population (Wu, 2013), we considered that this yeast also contributes to ethanol production in *Maotai*-flavor liquor fermentation.

Flavor compounds play essential roles in forming the unique flavor quality of *Maotai*-flavor liquor (Xu and Ji, 2012). The metabolic activity of *Z. bailii* MT15 in flavor production was studied at the end of fermentation and compared with that of *S. cerevisiae* MT1 (**Figure 1C**). Production of alcohols by *Z. bailii* MT15 was 16,333.78 μg/L, which was approximately half that of *S. cerevisiae* MT1. Production of ketones by *Z. bailii* MT15 was 319.95 μg/L, which was also less than that of *S. cerevisiae* MT1 (452.98 μg/L). However, the production of acids, esters, and aldehydes by *Z. bailii* MT15 was 1.75, 2.28, and 3.45 times of that of *S. cerevisiae* MT1, respectively.

Additionally, *Z. bailii* MT15 produced 38 flavor compounds including 19 alcohols, six acids, three esters, three ketones, two aldehydes, and five other compounds (**Figure 1C**). Among them, 19 flavor compounds were also detected in the fermentation broth of *S. cerevisiae* MT1. The amounts of 10 flavor compounds, including acetic acid, benzaldehyde, phenethyl acetate, 1-phenylethyl propionate, and acetophenone, produced by *Z. bailii* MT15 were significantly higher than those produced by *S. cerevisiae* MT1 (**Figure 1C**). Furthermore, *Z. bailii* MT15 was able to produce 19 unique flavor compounds compared with *S. cerevisiae* MT1, including 13 alcohols, two acids, one ketone, one aldehyde, and two other flavor compounds. Some of these unique flavor compounds were influential flavor compounds and contributed significantly to the flavor of the *Maotai*-flavor liquor, such as 2-heptanol (fruity flavor), 2-nonanol (fruity flavor), 1-nonanol (fruity flavor), 1-hexanol (floral, green scent), geraniol (sweet, rose-like scent), 2-heptanone (fruity, spicy, cinnamon scent), propionic acid (vinegar-like scent), and butyric acid (cheesy-like scent) (Xu and Ji, 2012).

These results demonstrated that *Z. bailii* MT15 could generate ethanol and various flavor compounds including alcohols, acids, esters, aldehydes, and ketones during liquor fermentation, which would contribute to the flavor and quality of *Maotai*-flavor liquor. Moreover, previous studies showed that *Z. bailii* contributed to the flavor complexity of wine and could be used as a mixed starter with *S. cerevisiae* to improve the production of ethyl esters in wine fermentation (Ciani et al., 2009; Garavaglia et al., 2015). Therefore, except for the capacity of flavor production, *Z. bailii* would interact with *S. cerevisiae* during liquor fermentation and positively affect the flavor and quality of *Maotai*-flavor liquor.

Effect of Temperature on Flavor Metabolism of *Z. bailii* MT15

Maotai-flavor liquor is produced by a spontaneous fermentation process with a relatively high temperature of up to about 37°C (Wu et al., 2013). However, little is known about the metabolic properties and mechanisms of *Z. bailii* under such a relatively high temperature in food fermentation. We therefore studied the metabolic activity of *Z. bailii* MT15 at 37°C, and compared it with that at 30°C.

The biomass of *Z. bailii* MT15 decreased by 32.46% at 37°C compared with 30°C. Meanwhile, few differences were observed in the types of flavor compounds produced, but the amounts of flavor compounds per unit cell were substantially different between the two temperatures (Dataset 1). As shown in **Figure 2**, no significant change was observed in the production of alcohols and esters per unit cell, and the production of ketones decreased by 28.30% at the higher temperature. By contrast, the production of acids and aldehydes per unit cell at 37°C was, respectively, 110 and 41% higher than that at 30°C. *Maotai*-flavor liquor contains large amounts of acids and aldehydes, which have positive effects on its unique sensory characteristics (Fan et al., 2011; Xu et al., 2017). The results suggested that production of acids and aldehydes in *Z. bailii* MT15 was enhanced under higher temperature. This enhancement would probably contribute to the unique sensory characteristics of *Maotai*-flavor liquor. Nevertheless, the metabolic mechanisms for its vigorous activities under higher temperature remains unclear. Therefore, we used omics' technology including genomic and transcriptomic analysis to unravel the metabolic features and mechanisms of *Z. bailii* MT15.

Genomic and Transcriptomic Analysis of *Z. bailii* MT15

The whole genome of *Z. bailii* MT15 was sequenced and compared with the genome of three other *Z. bailii* strains (*Z. bailii* CLIB 213^T, *Z. bailii* ISA1307, and *Z. bailii* IST302) (Galeote et al., 2013; Mira et al., 2014; Palma et al., 2017), and the results are shown in **Table 1**. The *Z. bailii* MT15 genome assembly resulted in 287 contigs (>527 bp) with an N50 value of 156,420 bp, and 95 scaffolds (>2020 bp) with an N50 value of 684,448 bp. The assembled genome was 20.19 Mb with a GC content of 42.38 mol%. The genome size of *Z. bailii* MT15 was 1.97-fold that of *Z. bailii* CLIB 213^T and 1.87-fold that of *Z. bailii* IST302, but was similar to that of *Z. bailii* ISA1307. A total of 9,498 genes were predicted with an average length of 1,461 bp, occupying 68.72% of the whole genome (**Table 1**).

The number of genes in the *Z. bailii* MT15 genome was nearly twice that of *Z. bailii* CLIB 213^T and *Z. bailii* IST302,

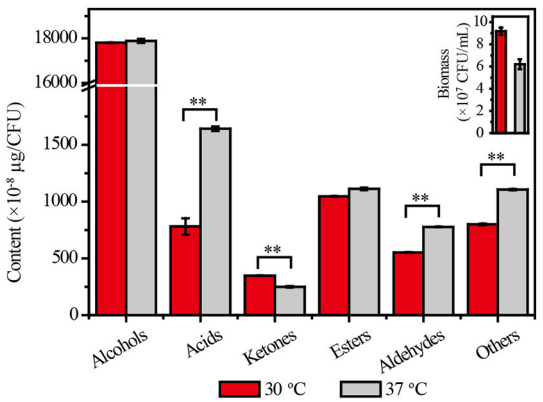


FIGURE 2 | Cell growth and production of flavor compounds by *Z. bailii* MT15 at 30 and 37°C. Significance level of the one-way ANOVA: $^{**}P < 0.01$.

TABLE 1 | General features of the *Z. bailii* MT15, *Z. bailii* CLIB 213^T, *Z. bailii* ISA1307, and *Z. bailii* IST302 genomes.

Feature	*Z. bailii* MT15	*Z. bailii* CLIB 213^T	*Z. bailii* ISA1307	*Z. bailii* IST302
Genome size (Mb)	20.19	10.27	21.14	10.77
GC content (mol%)	42.38	42.50	42.40	42.20
GC content in mRNA region (mol%)	–	–	43.80	NP
rRNA number	3	0	3	NP
tRNA number	507	162	514	NP
Coding sequence number	9498	4723	9925	5142
Average coding sequence length (bp)	1461	1475	1471	NP
Total coding sequence length (Mb)	13.88	6.97	14.76	NP
Coding sequence of genome (%)	68.72	67.97	69.80	NP

NP, Not published.

but was close to the number in *Z. bailii* ISA1307. It has been proven that *Z. bailii* ISA1307 is an interspecies hybrid strain that was generated in a stressful environment to improve strain robustness (Sipiczki, 2008; Mira et al., 2014). We analyzed the sequences of housekeeping genes including *RPB1*, *RPB2*, *TBB*, and *EFGM* in the *Z. bailii* MT15 genome, which are proposed to have a high capacity to discriminate *Zygosaccharomyces* species (Suh et al., 2013). The *RPB1*, *RPB2*, *TBB*, and *EFGM* genes were all duplicated in the *Z. bailii* MT15 genome, and the sequences of alleles of the four genes were different, which indicated that *Z. bailii* MT15 might be an interspecies hybrid strain. Furthermore, alleles of the four genes showed high sequence identity to the orthologous genes in *Z. bailii* CLIB 213^T and *Zygosaccharomyces parabailii* ATCC 60483 (Supplementary Table 3). Thus, it is likely that *Z. bailii* MT15 is an interspecies hybrid strain generated from *Z. bailii* and *Z. parabailii*, which was beneficial for its adaptation to the relatively high temperature environment in *Maotai*-flavor liquor fermentation.

RNA-Seq was employed to reveal the transcriptomic features of flavor metabolism in *Z. bailii* MT15 under heat stress at 30and 37°C. The results showed that 257 genes (2.71% of the *Z. bailii* MT15 genome) were differentially expressed (≥1.5-fold, $P < 0.05$), including 126 up-regulated genes (Dataset 2) and 131 down-regulated genes (Dataset 3). These differentially expressed genes (DEGs) were clustered according to eggNOG functional categories (**Figure 3**). Among these categories, general function prediction only (8.56% of DEGs), transcription (7.39% of DEGs), amino acid transport and metabolism (7.00% of DEGs), and carbohydrate transport and metabolism (5.84% of DEGs), contained the greatest number of DEGs (**Figure 3**). Amino acid metabolism is important for the production of ethanol and most flavor compounds in liquor; therefore, we further analyzed the DEGs involved in amino acid transport and metabolism.

Comparative transcriptome analysis showed that 12 genes related to amino acid transport and metabolism were significantly up-regulated at 37°C (**Figure 4A**). Among them, eight genes encoded amino acid permeases, including one general amino acid permease and seven permeases specific for arginine, proline, lysine, and γ-aminobutyric acid. Higher transcription levels of these genes would be important for the transport of more amino acids into yeast cells (Jauniaux and Grenson, 1990; Regenberg et al., 1999). The other four genes were found to be involved in amino acids biosynthesis, including *PTR2* and *OPT1* associated with the transport of small peptides, and $MEP2_1$ and $MEP2_2$ encoding ammonium transporters, would also contribute to a strengthened amino acid metabolism (Meister, 1957). Thus, we speculated that these up-regulated amino acid metabolism related genes would promote the absorption and utilization of amino acid at 37°C. To validate this speculation, we detected the absorption of the most conventional amino acids by *Z. bailii* MT15 at 30 and 37°C after 24 h of cultivation. As shown in **Figure 4B**, the absorption of 14 amino acids was obviously improved by 83% at 37°C, including the uptake of glutamate acid, aspartate, proline, alanine, and arginine. Among these, the uptake of glutamate acid increased most prominently from 1.14×10^{-9} to 2.11×10^{-9} mg/CFU. By contrast, the absorption of valine and cysteine decreased, and no significant difference was found in the phenylalanine absorption. Therefore, we can conclude that *Z. bailii* MT15 could absorb and utilize amino acids more efficiently at 37 than at 30°C, due to the strengthened amino acid associated pathway at higher temperature. The strengthened absorption and utilization of amino acids possibly contributed to the improvement of the corresponding flavor compounds production, in particular, the production of acids and aldehydes.

Metabolic Mechanisms of Production of Flavor Compounds in *Z. bailii* MT15

Z. bailii MT15 could produce various flavor compounds, especially alcohols, acids, and esters (**Figure 2**). During food fermentation, alcohols and acids are generated by yeasts from amino acids through the Ehrlich pathway and from

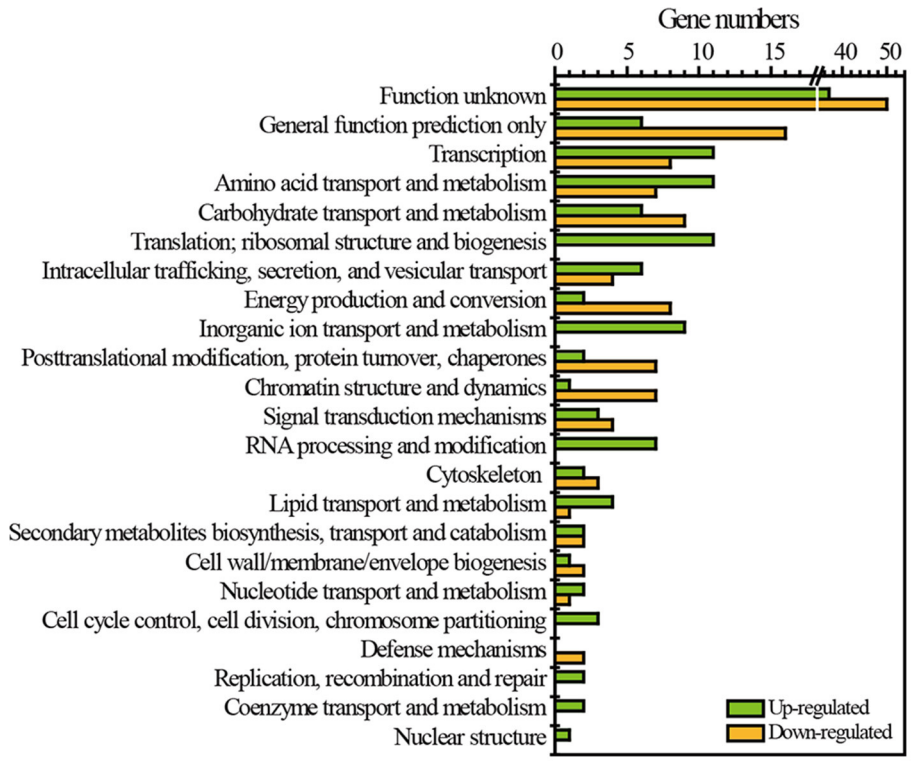


FIGURE 3 | Functional classes of DEGs of *Z. bailii* MT15 at 30 and 37°C using eggNOG functional categories.

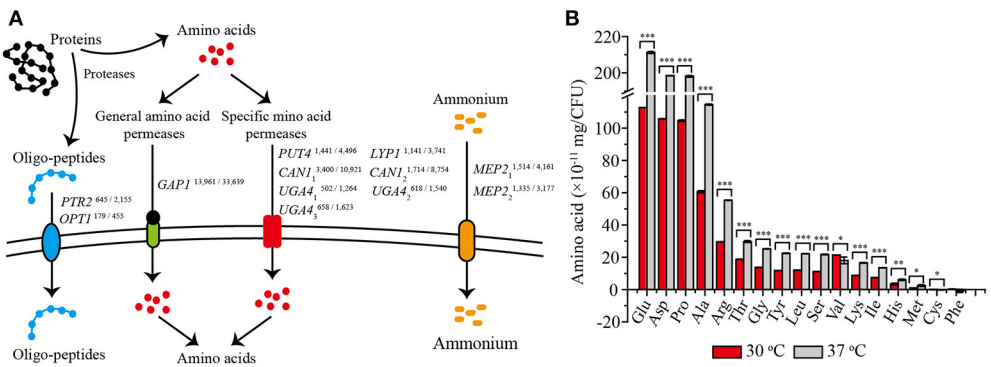


FIGURE 4 | Differentially expressed genes related to amino acid transport **(A)** and the absorption of 17 conventional amino acids **(B)** at 30 and 37°C after 24 h of cultivation. The RPKM value of each gene is presented in the upper right corner, and the numbers on the left and right represent the RPKM values of *Z. bailii* MT15 at 30 and 37°C, respectively. Significance level of the one-way ANOVA: *0.05 > P > 0.01; **0.01 > P > 0.001; ***P < 0.001.

sugars through the Harris pathway (**Figure 5**; Ehrlich, 1907; Chen, 1978). To unravel the molecular mechanisms of flavor metabolism in *Z. bailii* MT15, we compared its genome with that of strain *S. cerevisiae* MT1 and analyzed the transcription levels of genes involved in flavor metabolism at 37 and 30°C.

As shown in **Figure 5**, sugars could be converted to α-keto acids via the glycolysis and tricarboxylic acid (TCA) cycle pathways. In the *Z. bailii* MT15 genome, the copy number of genes involved in glycolysis and the TCA cycle was around twice those in the *S. cerevisiae* MT1 genome (Supplementary Table 4), which was expected as *Z. bailii* MT15 was considered to be an interspecies hybrid strain. In addition, α-keto acids could also be generated by various aminotransferases through the Ehrlich pathway (Hazelwood et al., 2008). In the *Z. bailii* MT15 genome, at least 16 different genes were annotated as aminotransferases including *AGX1*, *BCA1*, *ARO9*, *ARO8*, and *YGD3*. (Supplementary Table 5). Among them, *ARO9* and *ARO8*

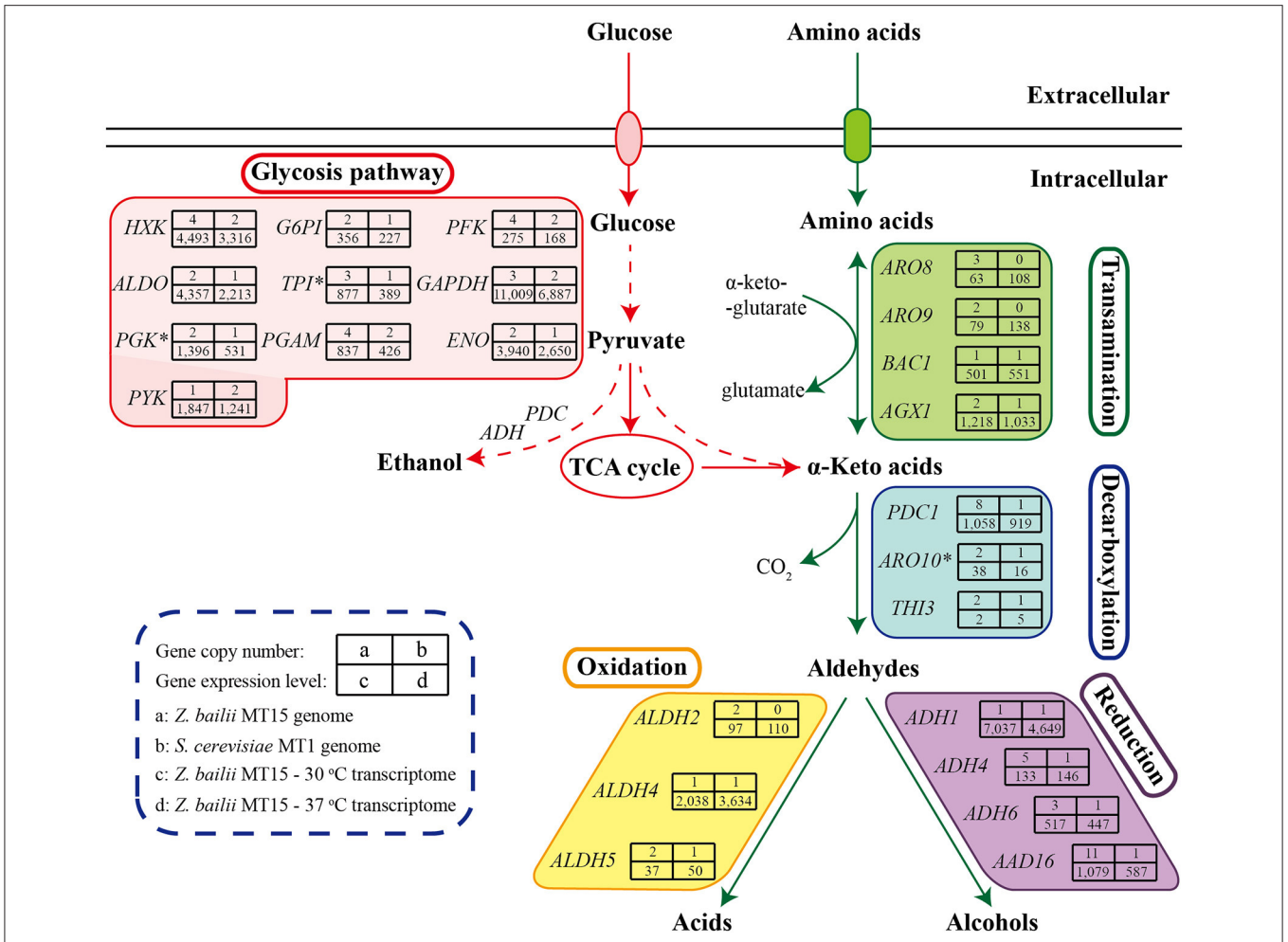


FIGURE 5 | Genome and transcriptome analysis of *Z. bailii* MT15 in flavor production. Catabolism of sugars and amino acids leads to the formation of various flavor compounds. Genes encoding the key enzymes required in each step are indicated. *This gene was a DEG (≥2-fold, $P < 0.05$) at 37°C.

were uniquely found in the *Z. bailii* MT15 genome and might be associated with the differences in flavor metabolism compared with *S. cerevisiae* MT1.

The α-keto acids were further decarboxylated to the corresponding aldehydes by α-keto acid decarboxylases including *PDC1*, *PDC5*, *PDC6*, *ARO10*, and *THI3* (ter Schure et al., 1998; Dickinson et al., 2003; Vuralhan et al., 2005). Among them, *PDC1*, *ARO10*, and *THI3* existed in both of the two strains, while *PDC5* and *PDC6* were uniquely found in the *S. cerevisiae* MT1 genome. Pyruvate decarboxylases (PDCs) participate in alcoholic fermentation by converting pyruvate to acetaldehyde and in amino acid metabolism through the Ehrlich pathway (ter Schure et al., 1998). We found 13 genes encoding PDC enzymes in the *Z. bailii* MT15 genome, whereas only six were found in the *S. cerevisiae* MT1 genome. Since the ethanol production of *Z. bailii* MT15 was less than that of *S. cerevisiae* MT1 (**Figure 1B**), more PDC genes in *Z. bailii* MT15 did not favor its alcoholic fermentation, but might be beneficial for the conversion of amino acids and reducing sugars to aldehydes by the Ehrlich pathway.

Aldehydes could subsequently be converted to higher alcohols and acids by alcohol dehydrogenases and aldehyde dehydrogenases, respectively (Hazelwood et al., 2008). So far, at least 16 genes encoding alcohol dehydrogenases have been found to catalyze the interconversion of aldehydes and alcohols (Hazelwood et al., 2008). However, only four of these genes (*ADH1*, *ADH4*, *ADH6*, and *AAD16*) were found in the *Z. bailii* MT15 genome, while 11 such genes were present in the *S. cerevisiae* MT1 genome. *Z. bailii* MT15 harboring fewer alcohol dehydrogenases genes may be the reason that it produced more aldehydes and less alcohols compared with *S. cerevisiae* MT1. Moreover, the genome analysis showed that three genes (*ALDH2*, *ALDH4*, and *ALDH5*) annotated as aldehyde dehydrogenase were found in the *Z. bailii* MT15 genome, and *ALDH2* was a specific gene compared with *S. cerevisiae* MT1. This would account for the higher acid productivity of *Z. bailii* MT15. Therefore, we can conclude that genome differences, particularly in the aspect of genes involved in amino acid metabolism between *Z. bailii* MT15 and *S. cerevisiae* MT1, would lead to the different metabolic features of the two strains (**Figure 1C**).

The transcriptome comparison of *Z. bailii* MT15 revealed that genes involved in the glycolysis pathway showed lower RPKM values at 37 than those at 30°C, which indicated that the glycolysis pathway might not be related to the increased production of acids and aldehydes at higher temperature. Moreover, in the TCA cycle, *ACON2* associated with the production of α-ketoglutaric acid was significantly up-regulated at 37°C, which would promote production of α-keto acids. Meanwhile, the RPKM values of *ARO8* and *ARO9*, which are involved in the transamination step of the Ehrlich pathway and regarded as broad-substrate-specificity aminotransferases (Iraqui et al., 1998), were higher at 37 than those at 30°C ($\geq$1.50-fold, $P \leq 0.25$). The higher expression of these genes might be beneficial for the transamination of amino acids to corresponding α-keto acids at 37°C. Furthermore, *ALDH4* also showed higher RPKM values at 37°C (1.78-fold, $P = 0.19$), which was consistent with the increased production of acids at 37°C (**Figure 2**). Therefore, the higher expression of genes, including *ACON2*, *ARO8*, *ARO9*, and *ALDH4* in the Harris pathway and Ehrlich pathway, might be beneficial for the increased production of acids and aldehydes at the relatively high fermentation temperature.

CONCLUSIONS

This study revealed the fermentation characteristics and potential function of *Z. bailii* MT15 that produces various flavor compounds including alcohols, acids, esters, aldehydes, and ketones. Its ability to generate acids and aldehydes is improved at the relatively high temperature used in liquor fermentation, which is beneficial for the complexity of the aroma and quality of the liquor. The genome and transcriptome analysis of *Z. bailii* MT15 revealed that amino acids metabolism plays important roles in flavor production. This work sheds new light on the metabolic characteristics of *Z. bailii* in flavor production during *Maotai*-flavor liquor fermentation that would be applicable to various food fermentations.

AUTHOR CONTRIBUTIONS

YX (first author), QW, and YZ drafted the manuscript. YX (first author), QW, and RD performed the physiological studies and genome sequencing and transcriptome analysis. QW and YX (last author) participated in the design of the study. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31371822, 31530055), the National Key R&D Program (2016YFD0400503), the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the 111 Project (no. 111-2-06).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02609/full#supplementary-material

REFERENCES

Anders, S., and Huber, W. (2013). *Differential Expression of RNA-Seq Data at the Gene Level—the DESeq package*. Available online at: www.bioconductor.org/packages/devel/bioc/vignettes/DESeq/inst/doc/DESeq.pdf (Accessed October 17, 2013).

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31, 365–370. doi: 10.1093/nar/gkg095

Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biol.* 13:R56. doi: 10.1186/gb-2012-13-6-r56

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–579. doi: 10.1093/bioinformatics/btq683

Chen, E. C. H. (1978). The relative contribution of Ehrlich and biosynthetic pathways to the formation of fusel alcohols. *Am. Soc. Brew. Chem.* 36, 39–43.

Ciani, M., Comitini, F., Mannazzu, I., and Domizio, P. (2009). Controlled mixed culture fermentation: a new perspective on the use of non-Saccharomyces yeasts in winemaking. *FEMS Yeast Res.* 10, 123–133. doi: 10.1111/j.1567-1364.2009.00579.x

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636–4641. doi: 10.1093/nar/27.23.4636

Dickinson, J. R., Salgado, L. E., and Hewlins, M. J. (2003). The catabolism of amino acids to long chain and complex alcohols in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 278, 8028–8034. doi: 10.1074/jbc.M211914200

Domizio, P., Romani, C., Comitini, F., Gobbi, M., Lencioni, L., Mannazzu, I., et al. (2011). Potential spoilage non-Saccharomyces yeasts in mixed cultures with *Saccharomyces cerevisiae*. *Ann. Microbiol.* 61, 137–144. doi: 10.1007/s13213-010-0125-1

Ehrlich, F. (1907). Über die Bedingungen der Fuselölbildung und über ihren Zusammenhang mit dem Eiweissaufbau der Hefe. *Ber. Dtsch. Chem. Ges.* 40, 1027–1047. doi: 10.1002/cber.190704001156

Fan, W., Shen, H., and Xu, Y. (2011). Quantification of volatile compounds in Chinese soy sauce aroma type liquor by stir bar sorptive extraction and gas chromatography-mass spectrometry. *J. Sci. Food Agric.* 91, 1187–1198. doi: 10.1002/jsfa.4294

Galeote, V., Bigey, F., Devillers, H., Neuvéglise, C., and Dequin, S. (2013). Genome sequence of the food spoilage yeast *Zygosaccharomyces bailii* CLIB 213T. *Genome Announc.* 1, e00606–e00613. doi: 10.1128/genomeA.00606-13

Garavaglia, J., Schneider, Rde. C., Camargo Mendes, S. D., Welke, J. E., Zini, C. A., Caramão, E. B., et al. (2015). Evaluation of *Zygosaccharomyces bailii* BCV 08 as a co-starter in wine fermentation for the improvement of ethyl esters production. *Microbiol. Res.* 173, 59–65. doi: 10.1016/j.micres.2015.02.002

Hazelwood, L. A., Daran, J. M., van Maris, A. J., Pronk, J. T., and Dickinson, J. R. (2008). The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl. Environ. Microbiol.* 74, 3920–3920. doi: 10.1128/AEM.02625-07

Iraqui, I., Vissers, S., Cartiaux, M., and Urrestarazu, A. (1998). Characterisation of *Saccharomyces cerevisiae* ARO8 and ARO9 genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily. *Mol. Gen. Genet.* 257, 238–248. doi: 10.1007/s004380050644

Jauniaux, J. C., and Grenson, M. (1990). GAP1, the general amino acid permease gene of *Saccharomyces cerevisiae*. Nucleotide sequence, protein similarity with the other bakers yeast amino acid permeases, and nitrogen catabolite repression. *Eur. J. Biochem.* 190, 39–44. doi: 10.1111/j.1432-1033.1990.tb15542.x

Jayabalan, R., Malini, K., Sathishkumar, M., Swaminathan, K., and Yun, S. E. (2010). Biochemical characteristics of tea fungus produced

during kombucha fermentation. *Food Sci. Biotechnol.* 19, 843–847. doi: 10.1007/s10068-010-0119-6

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2015). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Kong, Y., Wu, Q., Zhang, Y., and Xu, Y. (2014). *In situ* analysis of metabolic characteristics reveals the key yeast in the spontaneous and solid-state fermentation process of Chinese light-style liquor. *Appl. Environ. Microbiol.* 80, 3667–3676. doi: 10.1128/AEM.04219-13

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. doi: 10.1186/1471-2105-5-59

Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lu, X., Wu, Q., Zhang, Y., and Xu, Y. (2015). Genomic and transcriptomic analyses of the Chinese Maotai-flavored liquor yeast MT1 revealed its unique multi-carbon co-utilization. *BMC Genomics* 16:1064. doi: 10.1186/s12864-015-2263-0

Meister, A. (1957). *Biochemistry of the Amino Acids*. New York, NY: Academic Press.

Meng, X., Wu, Q., Wang, L., Wang, D., Chen, L., and Xu, Y. (2015). Improving flavor metabolism of *Saccharomyces cerevisiae* by mixed culture with *Bacillus licheniformis* for Chinese Maotai-flavor liquor making. *J. Ind. Microbiol. Biotechnol.* 42, 1601–1608. doi: 10.1007/s10295-015-1647-0

Mira, N. P., Münsterkötter, M., Dias-Valada, F., Santos, J., Palma, M., Roque, F. C., et al. (2014). The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res.* 21, 299–313. doi: 10.1093/dnares/dst058

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nederbragt, A. J. (2014). On the middle ground between open source and commercial software - the case of the Newbler program. *Genome Biol.* 15, 1–2. doi: 10.1186/gb4173

Palma, M., Münsterkötter, M., Peça, J., Güldener, U., and and, Sá-Correia, I. (2017). Genome sequence of the highly weak-acid-tolerant *Zygosaccharomyces bailii* IST302, amenable to genetic manipulations and physiological studies. *FEMS Yeast Res.* 17:fox025. doi: 10.1093/femsyr/fox025

Palma, M., Roque Fde, C., Guerreiro, J. F., Mira, N. P., Queiroz, L., and Sá-Correia, I. (2015). Search for genes responsible for the remarkably high acetic acid tolerance of a *Zygosaccharomyces bailii*-derived interspecies hybrid strain. *BMC Genomics* 16:1070. doi: 10.1186/s12864-015-2278-6

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., et al. (2012). eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40, 284–289. doi: 10.1093/nar/gkr1060

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341. doi: 10.1186/1471-2164-13-341

Regenberg, B., Düring-Olsen, L., Kielland-Brandt, M. C., and Holmberg, S. (1999). Substrate specificity and gene expression of the amino-acid permeases in *Saccharomyces cerevisiae*. *Curr. Genet.* 36, 317–328. doi: 10.1007/s002940050506

Schattner, P., Brooks, A., N., Lowe, T., M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689. doi: 10.1093/nar/gki366

Sipiczki, M. (2008). Interspecies hybridization and recombination in *Saccharomyces* wine yeasts. *FEMS Yeast Res.* 8, 996–1007. doi: 10.1111/j.1567-1364.2008.00369.x

Solieri, L., Landi, S., De Vero, L., and Giudici, P. (2006). Molecular assessment of indigenous yeast population from traditional balsamic vinegar. *J Appl. Microbiol.* 101, 63–71. doi: 10.1111/j.1365-2672.2006.02906.x

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19(Suppl. 2), ii215–ii225. doi: 10.1093/bioinformatics/btg1080

Stratford, M., Steels, H., Nebe-Von-Caron, G., Novodvorska, M., Hayer, K., and Archer, D. B. (2013). Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int. J. Food Microbiol.* 166, 126–134. doi: 10.1016/j.ijfoodmicro.2013.06.025

Suh, S.-O., Gujjari, P., Beres, C., Beck, B., and Zhou, J. (2013). Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int. J. Syst. Evol. Microbiol.* 63, 1922–1929. doi: 10.1099/ijs.0.048058-0

Teoh, A. L., Heard, G., and Cox, J. (2004). Yeast ecology of Kombucha fermentation. *Int. J. Food Microbiol.* 95, 119–126. doi: 10.1016/j.ijfoodmicro.2003.12.020

ter Schure, E. G., Flikweert, M. T., van Dijken, J. P., Pronk, J. T., and Verrips, C. T. (1998). Pyruvate decarboxylase catalyzes decarboxylation of branched-chain 2-oxo acids but is not essential for fusel alcohol production by *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 64, 1303–1307.

Vuralhan, Z., Luttik, M. A., Tai S. L., Boer, V. M., Morais, M. A., Schipper, D., et al. (2005). Physiological characterization of the ARO10-dependent, broad-substrate-specificity 2-oxo acid decarboxylase activity of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 71, 3276–3284. doi: 10.1128/AEM.71.6.3276-3284.2005

Wu, Q., Chen, L., and Xu, Y. (2013). Yeast community associated with the solid state fermentation of traditional Chinese Maotai-flavor liquor. *Int. J. Food Microbiol.* 166, 323–330. doi: 10.1016/j.ijfoodmicro.2013.07.003

Wu, Q., Xu, Y., and Chen, L. (2012). Diversity of yeast species during fermentative process contributing to Chinese Maotai-flavour liquor making. *Lett. Appl. Microbiol.* 55, 301–307. doi: 10.1111/j.1472-765X.2012.03294.x

Wu, X. (2013). *Diversity and Dynamics of Yeasts and Bacteria during the solid State Fermentative Process Contributing to Chinese Maotai-Flavor Liquor Making*. Master's thesis, Jiangnan University, Jiangsu.

Xu, M.L, Yu, Y., Ramaswamy, H., and Zhu, S. (2017). Characterization of Chinese liquor aroma components during aging process and liquor age discrimination using gas chromatography combined with multivariable statistics. *Sci. Rep.* 7:39671. doi: 10.1038/srep39671

Xu, Y., and Ji, K. (2012). 15–Moutai (Maotai): production and sensory properties. *Alcoholic Beverages* 25, 315–330. doi: 10.1533/9780857095176.3.315

Xu, Y., Wu, Q., and Chen, L. (2013). *Screening and Application of the Yeast CGMCC 4745 with High Ethanol Production and Low Fat Fusel Oil in Chinese Maotai-Flavor Liquor Fermentation*. Chinese Patent, CN102816706A.

Zhang, J., Fang, F., Chen, J., and Du, G. (2014). The arginine deiminase pathway of koji bacteria is involved in ethyl carbamate precursor production in soy sauce. *FEMS Microbiol. Lett.* 358, 91–97. doi: 10.1111/1574-6968.12542

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Xu, Zhi, Wu, Du and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 April 2017
doi: 10.3389/fmicb.2017.00688

A Novel RAYM_RS09735/RAYM_RS09740 Two-Component Signaling System Regulates Gene Expression and Virulence in *Riemerella anatipestifer*

Ying Wang[1], *Ti Lu*[1], *Xuehuan Yin*[1], *Zutao Zhou*[1,2], *Shaowen Li*[1,2], *Mei Liu*[1,2], *Sishun Hu*[1,2], *Dingren Bi*[1,2] *and Zili Li*[1,2]*

[1] *Department of Preventive Veterinary Medicine, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, China,* [2] *State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, China*

The Gram-negative bacterium *Riemerella anatipestifer* is an important waterfowl pathogen, causing major economic losses to the duck-producing industry. However, little is known of the virulence factors that mediate pathogenesis during *R. anatipestifer* infection. In this study, RAYM_RS09735 and RAYM_RS09740 were predicted to form a two-component signaling system (TCS) through bioinformatics analysis. This TCS was highly conserved across the *Flavobacteriaceae*. A mutant YMΔRS09735/RS09740 strain was constructed to investigate the role of the RAYM_RS09735/RAYM_RS09740 TCS in *R. anatipestifer* virulence and gene regulation. The median lethal dose (LD_{50}) of YMΔRS09735/RS09740 was found to be $>10^{11}$ CFU, equivalent to that of avirulent bacterial strains. The bacterial abundances of the YMΔRS09735/RS09740 strain in the heart, brain, liver, blood, and spleen were significantly lower than that of the wild-type *R. anatipestifer* YM strain. Pathological analysis using hematoxylin and eosin staining showed that, compared to the wild-type, the mutant YMΔRS09735/RS09740 strain caused significantly less virulence in infected ducklings. RNAseq and real-time PCR analysis indicated that the RAYM_RS09735/RAYM_RS09740 TCS is a PhoP/PhoR system. This is a novel type of TCS for Gram-negative bacteria. The TCS was also found to be a global regulator of expression in *R. anatipestifer*, with 112 genes up-regulated and 693 genes down-regulated in the YMΔRS09735/RS09740 strain (~33% genes demonstrated differential expression). In summary, we have reported the first PhoP/PhoR TCS identified in a Gram-negative bacterium and demonstrated that it is involved in virulence and gene regulation in *R. anatipestifer*.

Keywords: ***Riemerella anatipestifer***, ***RAYM_RS09735/RAYM_RS09740***, **two-component signaling system, virulence, RNAseq**

OPEN ACCESS

Edited by:
Florence Abram,
NUI Galway, Ireland

Reviewed by:
Dave Siak-Wei Ow,
*Bioprocessing Technology Institute (A*STAR), Singapore*
Robin Anderson,
Agricultural Research Service (USDA), USA

***Correspondence:**
Zili Li
lizili@mail.hzau.edu.cn

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *23 January 2017*
Accepted: *04 April 2017*
Published: *21 April 2017*

Citation:
Wang Y, Lu T, Yin X, Zhou Z, Li S, Liu M, Hu S, Bi D and Li Z (2017) A Novel RAYM_RS09735/RAYM_RS09740 Two-Component Signaling System Regulates Gene Expression and Virulence in Riemerella anatipestifer. Front. Microbiol. 8:688. doi: 10.3389/fmicb.2017.00688

INTRODUCTION

The disease riemerella anatipestifersis, caused by the Gram-negative bacterium *Riemerella anatipestifer*, occurs primarily in 1–8-week-old ducks but is most common in more susceptible 2–3-week-old ducklings. It is currently the most economically damaging bacterial infection affecting the global duck industry. Symptoms are characterized by fibrinous pericarditis, glissonitis, airbag

inflammation, and meningitis (Segers et al., 1993). A total of 21 *R. anatipestifer* serotypes have been identified, with no significant cross-protection reported, making it difficult to control the disease through vaccination (Pathanasophon et al., 1995, 2002). *R. anatipestifer* serotypes 1, 2, and 10 are responsible for most major outbreaks in China. In recent years, many *R. anatipestifer* strains with differing virulence have been isolated from duck farms in China (Yuan et al., 2011; Wang et al., 2015; Zhang et al., 2015; Song et al., 2016). Additionally, several virulence factors have been identified that associate with disease severity, including VapD (Chang et al., 1998), CAMP cohemolysin (Crasta et al., 2002), outer membrane protein A (OmpA; Hu et al., 2011), Nicotinamidase PncA (Wang et al., 2016), and putative genes associated with lipopolysaccharide (LPS) synthesis (Yu et al., 2016).

Bacteria sense and adapt to their environment via two-component signaling systems (TCSs). A paired TCS typically has a sensing histidine kinase (HK) coupled to a response regulator (RR). The sensing of a signal by the HK leads to autophosphorylation on a histidine residue. Subsequent transfer of the phosphate to an aspartate residue on the cognate RR facilitates the binding of the RR to its specific DNA. Each phosphorylated RR regulates specific genes that enable individual bacteria to sense environmental factors and respond to stresses (Stock et al., 2003; Mike et al., 2014). TCSs are present in nearly all sequenced bacterial genomes, as well as in some fungal, archaeal, and plant species (Skerker et al., 2008).

TCSs are involved in the regulation of a variety of important biological functions in bacteria. In particular, PhoP/PhoQ TCSs control the transcription of key virulent genes in several bacterial pathogens, including *Escherichia coli, Shigella* spp. (Tobe, 2008), *Yersinia pestis* (O'Loughlin et al., 2010), and *Salmonella typhimurium* (Tran et al., 2016). TCSs also control the expression of *Pneumococcal* surface antigen A (PsaA), subsequently regulating virulence and resistance to oxidative stress in *Streptococcus pneumonia* (McCluskey et al., 2004). Related two-component system also regulate genes that are essential to virulence and complex lipid biosynthesis. For example, SenX3, and RegX3 form a TCS that is involved in *Mycobacterium tuberculosis* virulence (Ryndak et al., 2008). The TCS is expressed during phosphate starvation and is required for phosphate uptake and aerobic respiration (Parish et al., 2003; Haydel et al., 2012). Additionally, in TB, the TCS PrrAB is required during early intracellular infection (Haydel et al., 2012) and the TCS MprAB responds to envelope stress by regulating stress-responding and virulence-associated genes (He et al., 2006; Pang et al., 2007).

The bacterium *R. anatipestifer* YM was isolated in Yunmeng, Hubei province, China, and is a highly virulent strain classified as serotype 1 (Zhou et al., 2011). We have previously used *in vivo*-induced antigen technology (IVIAT) to identify *in vivo*-induced protein antigens from *R. anatipestifer*. This predicted the involvement of a putative TCS. To research the function of this TCS in *R. anatipestifer*, we constructed a mutant with the putative TCS genes, *RAYM_RS09735* and *RAYM_RS09740*, deleted to investigate their biological characteristics. We found that RAYM_RS09735/RAYM_RS09740 form a PhoP/PhoR TCS, the first reported such TCS in Gram-negative bacteria. We confirmed that the RAYM_RS09735/RAYM_RS09740 TCS is an important global transcription regulator and regulates the expression of virulence-associated genes in *R. anatipestifer*. This may provide the theoretical basis for further study into the molecular pathogenesis of *R. anatipestifer* and facilitate the design of genetically engineered vaccines against *R. anatipestifer*.

MATERIALS AND METHODS

Analysis of RAYM_RS09735 and RAYM_RS09740 Homology in *R. anatipestifer*

The bacterial strains and plasmids used in this study are listed in **Table 1**. Primers used in this study are listed in **Table 2**. *In vivo*-induced antigen technology (IVIAT) was

TABLE 1 | Bacterial strains and plasmids used in this study.

Strains	Description	Source or references
STRAINS		
R. anatipestifer YM	*R. anatipestifer* wild-type strain (serotype 1)	Lab collection (Zhou et al., 2011)
YMΔRS09735/RS09740	Mutant strain, derived from *R. anatipestifer* YM, in which the *RAYM_RS09735/RAYM_RS09740* genes were inactivated by $Spec^R$ cassette insertion	This study
E.coli DH5α	F-φ80 lac ZΔM15 Δ(lacZYA-arg F) U169 endA1 recA1 hsdR17(rk-,mk+) supE44λ-thi -1 gyrA96 relA1 phoA	Lab collection
E.coli x7213	Thi-1 thr-1 leuB6 glnV44 fhuA21 lacY1 recA1 RP4-2-Tc::Mu λpir ΔasdA4 Δzhf-2::Tn10	Lab collection
PLASMIDS		
pMD-18T	Clone vector,Amp^R	Life Technology
pIC333	Source of $Spec^R$ cassette	Lab collection
pRE112	Suicide vector, sacB mobRP4 R6K ori Cm^R	Life Technology
pRE112-LSR	Suicide vector with $Spec^R$ and homology arms for homologous recombination	This tudy

TABLE 2 | Primers used in this study.

Name	Sequence (5′–3′)
PRIMERS FOR CONSTRUCTING SUICIDE PLASMID AND *RAYM_* RS09735/RAYM_RS09740 GENES MUTANT	
RAYM_RS09735/RAYM_RS09740 L-F	CGGTACC TTTTAACTATTGTATTTATTTAC
RAYM_RS09735/RAYM_RS09740 L-R	CTTTTAAAACTACTGTCCAGTAGATTGATAA
$Spec^R$ F	TATCAATCTACTGGACAGTAGTTTTAAAAG
$Spec^R$ R	AAAACAACTTCAAAACAGTGGAACGAAAAC
RAYM_RS09735/RAYM_RS09740 R-F	GTTTTCGTTCCACTGTTTTGAAGTTGTTTT
RAYM_RS09735/RAYM_RS09740 R-R	TGAGCTCGGAAGGAAGCGTCTTGCAATATT
PRIMERS FOR IDENTIFYING THE *RAYM_RS09735/RAYM_RS09740* GENES MUTANT	
Check F	AAAGGTATGGGAATGGAG
Check R	GCCGTCTGTAGCAAGGGT
16S rRNA F	AGAGTTTGATCCTGGCTCAGGATC
16S rRNA R	ACGGCTACCTTGTTACGACTTAGC
PRIMERS FOR qRT-PCR	
RS09730 F	GAGCTTTGGGGTAGGGTAGA
RS09730 R	TTCCACCACCCAATTCCTGT
RS09735 F	GGTATGGGAATGGAGCCTCA
RS09735 R	GCCCCAAACCTTGACCTTTT
RS09740 F	ACGGCGAGGAAGGTCTTAAA
RS09740 R	CGTCTGCACCTAACTGATAACC
RS09745 F	AAATCCGCAACTTCGAGACG
RS09745 R	CCTCTGGCGTAGGAAAAGGA
RS05555 F	AGCTCTCTTGGATTGGCACT
RS05555 R	ATGCTCCCAAATACTCGGCT
RS04585 F	AAACGACGCTCCAATGCTTT
RS04585 R	GGAGGTGGAGAAGGAGCAAT
RS07825 F	AGTTGTAGGCGACCGAGAAA
RS07825 R	GGGAGGAATAGCTTTTCGGG
RS03930 F	AATGAGTGTCTTCAGGGGCA
RS03930 R	ACAGTCTGACGCTGCCTATT
RS04770 F	GGTTCCGAGAATTGGTTGCT
RS04770 R	TAGCCCTCAGTTTTGGAGCA
RS08905 F	GTCGTCCCCTCCTAAGTGTG
RS08905 R	GACTGTGGTGGTGGTACTCA
RS02680 F	GCCTCTACCTTCAGTCCAGA
RS02680 R	AGCGAGTTATTCCTGGACCT
RS08330 F	TCCCACATTGTAGGCAGGAG
RS08330 R	TGGAACACCTGGGCAGATAG
RS03960 F	AATCCAAGAGACAACCGCAC
RS03960 R	ACTCGTCCGTTAGCACTTCT
RS04760 F	AGCGATGAGAAAAGCAGCTC
RS04760 R	TCTCGACTAGGACAATGCCG
pstS F	AGTGCTACCAGTGATGGATGA
pstS R	ATCCATTCCCAACCCCGAAA
hydrolase Nlp/P60 F	GCGTTGTAAGCGGCTTTACT
hydrolase Nlp/P60 R	ACTCACTGCCGCTCATAAGA
BLP F	CTGCAACATAGTTAGCTTCGG
BLP R	ACAGCTACTCCCTCTAACGC
DnaB F	AAACCCGCCCTAATAAACGC
DnaB R	GCTGGTTTGTGCTCCTTCTT

then used to characterize potential virulence factors that are expressed in ducks during infection with *R. anatipestifer*. A genomic DNA library of *R. anatipestifer* was screened, demonstrating *in vivo*-induced increased expression of genes with two ORFs, *RAYM_RS09735* and *RAYM_RS09740* (data not shown). RAYM_RS09735 and RAYM_RS09740 were predicted to be a histidine protein kinase (HK) and a response regulator (RR), respectively. To investigate *RAYM_RS09735* and *RAYM_RS09740* homology among different *R. anatipestifer* strains, the open reading frames of *RAYM_RS09735* and *RAYM_RS09740* were amplified from the wild-type *R. anatipestifer* YM strain. The PCR product was cloned into pMD-18T vectors and sequenced. Homologous amino acid sequences were identified by searching the GenBank database using BLASTX. The resulting alignments were used for the construction of a phylogenetic tree using neighbor-joining (NJ) and were further analyzed using MEGA v6.06 software.

Bacterial Strains and Animals

R. anatipestifer RA-YM was grown in trypticase soy broth (TSB) or on agar plates (Difco Laboratories, USA) at 37°C with 5% CO_2. *Escherichia coli X7213* was cultured in Luria Bertani (LB) broth containing 50 μg/mL diaminopimelic acid (DAP), with shaking at 37°C overnight (Roland et al., 1999). When needed, antibiotics were used at the following concentrations: 50 μg/mL spectinomycin (Spec), 25 μg/mL chloramphenicol (Cm), and 100 μg/mL ampicillin (Amp). A portion of the bacterial colonies grown in LB medium were stored at −80°C with 15% glycerol.

Plasmid Construction

The suicide vector used to create the *R. anatipestifer* YM *RAYM_RS09735/RAYM_RS09740* mutant was based on pRE112. Briefly, sequences 800 bp upstream and 800 bp downstream of *RAYM_RS09735* and *RAYM_RS09740* were amplified by PCR using the primer pairs RAYM_RS09735/RAYM_RS09740L F/R and RAYM_RS09735/RAYM_RS09740R F/R. These contained *Kpn*I and *Sac*I restriction sites. A spectinomycin resistance ($Spec^R$) cassette (1,185 bp) was PCR amplified from plasmid pIC333 using the primers $Spec^R$ F and $Spec^R$ R. These three fragments were then purified from an agarose gel and used as a PCR template at a 1:1:2 M ratio to join overlapping PCR products with the primers RAYM_RS09735L-F and RAYM_RS09740R-R. The final product was digested with *Kpn*I and *Sac*I enzymes and ligated into the pRE112 plasmid (Edwards et al., 1998) to yield suicide plasmid pRE112-LSR. This plasmid was used to delete *RAYM_RS09735* and *RAYM_RS09740*.

Construction of the *R. anatipestifer* RAYM_RS09735/RAYM_RS09740 Mutant

The *R. anatipestifer* YM RAYM_RS09735/RAYM_RS09740 mutant, termed YMΔRS09735/RS09740, was constructed via allelic exchange using the previously constructed suicide plasmid pRE112-LSR. The donor strain *X7213* was transformed with pRE112-LSR and grown overnight in LB medium supplemented with 50 μg/mL DAP, 50 μg/mL Spec, and 25 μg/mL Cm. The recipient *R. anatipestifer* YM strain was cultured in TSB medium to an OD_{600} of 0.4–0.5. One milliliter of the donor strain and 3 mL of receptor strain were centrifuged at 3,000 rpm/min for 5 min, and the pellet re-suspended in 1 mL TSB. Mixed cultures were then incubated on a TSA plate supplemented with 50 μg/mL DAP at 37°C for 24 h, facilitating the conjugation of the donor and receptor strains. Cultures were next streaked onto a TSA plate containing 50 μg/mL spectinomycin (Spec) to isolate the putative *R. anatipestifer* YM conjugants from the mixed strains. Single colonies were re-purified on TSA plates supplemented with 50 μg/mL Spec. The *R. anatipestifer RAYM_RS09735/RAYM_RS09740* mutant strain was screened and validated by PCR.

Characteristics of the YMΔRS09735/RS09740 Strain

The wild-type strain *R. anatipestifer* or the mutant strain YMΔRS09735/RS09740 was grown in TSB at 37°C for 12 h with shaking, respectively. Equal amounts of YM culture were transferred into fresh TSB (without serum) at a ratio of 1:100 (vol/vol) and incubated at 37°C with shaking at 200 rpm. Bacterial growth was measured as described previously (Hu et al., 2011) by counting the number of bacterial CFUs at 2 h intervals for 14 h.

Determination of Bacterial Virulence and Survival *In vivo*

To determine whether deletion of *RAYM_RS09735* and *RAYM_RS09740* influenced *R. anatipestifer* virulence, the median bacterial lethal dose (LD_{50}) of the mutant strain YMΔRS09735/RS09740 was determined (Hu et al., 2010). Healthy Cherry Valley ducks were purchased from Chunjiang Duck Company (China) and hosted in an isolated animal room. All animal experiments and procedures were approved by the Research Ethics Committee, Huazhong Agricultural University, Hubei, China. A total of 75 ducks were divided randomly into 15 groups (five ducks per group). The respective bacterial strains were then injected into duck flippers at a dose of 10^5–10^{11} CFU. Ducks in groups 1–7 were injected with 10^5–10^{11} CFU of YM bacteria, and ducks in groups 8–14 were injected with 10^5–10^{11} CFU of YMΔRS09735/RS09740 strain (Zou et al., 2015). The control group was injected with an equivalent volume of PBS.

For pathological histological examination, ducks from each group were sacrificed 48 h after injection with 5×10^6 CFU of each bacterial strain diluted in 0.5 mL PBS. Heart, liver, brain, and spleen samples were collected. Tissues were then fixed in formalin, sectioned, and stained with hematoxylin and eosin.

Twelve 12-day-old ducks were randomly divided into three groups (four ducks per group) and infected with either 5×10^6 CFU of the wild-type *R. anatipestifer* strain or the mutant strain YMΔRS09735/RS09740. Blood, heart, liver, brain, and spleen samples were collected 48 h after injection. Tissues were homogenized in 5 mL PBS, serially diluted, and plated on TSA plates to assess the number of viable bacteria.

Extraction of Total RNA from *R. anatipestifer In vitro*

The mutant strain YMΔRS09735/RS09740 or wild-type strain YM was grown in TSB to log phase, respectively, and then harvested in no more than 3 mL culture by centrifugation at 4000–5000 × g for 5–10 min at 4°C. Total RNA was extracted using a Bacterial RNA Kit (Omega). A NANODROP 2000c (Nanodrop) was used to measure the concentration and quality of bacterial RNA.

RNAseq Library Construction

A total of 1 μg RNA per sample was used for RNA sample preparations. Sequencing libraries were generated using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB), following the manufacturer's recommendations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). The first strand of cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (RNase H–). Second strand cDNA synthesis was then subsequently performed using DNA Polymerase I and RNase H. Any remaining overhangs were converted into blunt ends via exonuclease/polymerase. After adenylation of the 3′ ends of DNA fragments, adaptors with a hairpin loop structure were ligated in preparation for hybridization. To select cDNA fragments by length, library fragments were purified with a AMPure XP system (Beckman Coulter). Next, 3 μL USER Enzyme (NEB) was incubated with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. PCR was performed using a Phusion High-Fidelity DNA polymerase, a universal PCR primer set and an Index (X) primer. Finally, PCR products were purified and the library quality assessed using an Agilent Bioanalyzer 2100 system (Agilent).

RNAseq Differential Expression Analysis

HTSeq v0.6.1 was used to count the number of reads that mapped to each gene and then gene expression was calculated using an RPKM method (Reads Per kb per Million reads). Differential expression analysis was performed using DESeq. A q-value (or FDR) <0.001& and a $\log_2$ fold-change >1 were set as the thresholds for determining significantly differential expression.

Go Enrichment Analysis

Gene Ontology (GO) enrichment analysis of differentially expressed genes was performed using the Bioconductor package GOseq, with a gene length bias correction. GO functional analysis provided GO functional classification annotation for DEGs, as

well as GO functional enrichment analysis. The Gene Ontology database used can be found at http://www.geneontology.org/.

KEGG Pathway Enrichment Analysis

As different genes cooperate with each other to exercise biological functions, pathway-based analysis can help further understand the biological functions of genes. We used KOBAS to test for the statistical enrichment of differential expression genes in the KEGG pathway data set (http://www.genome.jp/kegg/).

Quantitative Real-time PCR Analysis

For quantitative real-time PCR (qRT-PCR) validation experiments, 10 genes were randomly selected to assess the RNAseq data (**Table 2**). For this analysis, 1 μg of RNA was reverse-transcribed to cDNA using the PrimeScript™ RT regent kit with gDNA Eraser (Takara), according to manufacturer's instructions. cDNA was diluted 10-fold and used for real-time PCR analysis using a Bio-Rad CFX96™ System and signal detection protocols in accordance with the manufacturer's instructions (TaKaRa). *DnaB* was used as an endogenous control. Primers used for the qRT-PCR are described in **Table 2**. Data analysis was performed using GraphPad Prism v 5.0 Software (GraphPad).

Statistical Analysis and Data Records

Student's *t*-tests were used to compare gene expression data. *P*-values of ≤ 0.05 were considered significant. RNAseq original data were uploaded to the NCBI Short Read Archive (SRA) with study accession number SRP096616 (http://www.ncbi.nlm.nih.gov/sra).

RESULTS

Identification of the RAYM_RS09735 and RAYM_RS09740 Genes in *R. anatipestifer*

Homologous amino acid sequences were identified by searching the GenBank database using BLASTX. A phylogenetic tree was constructed containing 19 amino homologous acid sequences (**Figures 1A,B**). RAYM_RS09735 and RAYM_RS09740 share more than 98% sequence identity with seven other *R. anatipestifer* strains. In addition, RAYM_RS09735 and RAYM_RS09740 shared 70% identity with species of *Flavobacteriaceae*, including *Cloacibacterium*, *Epilithonimonas*, and *Chryseobacterium*. Our results demonstrated that RAYM_RS09735 and RAYM_RS09740 are not only highly conserved in *R. anatipestifer*, but also conserved across the *Flavobacteriaceae* in general. Functional assessment predicted RAYM_RS09735 and RAYM_RS09740 to

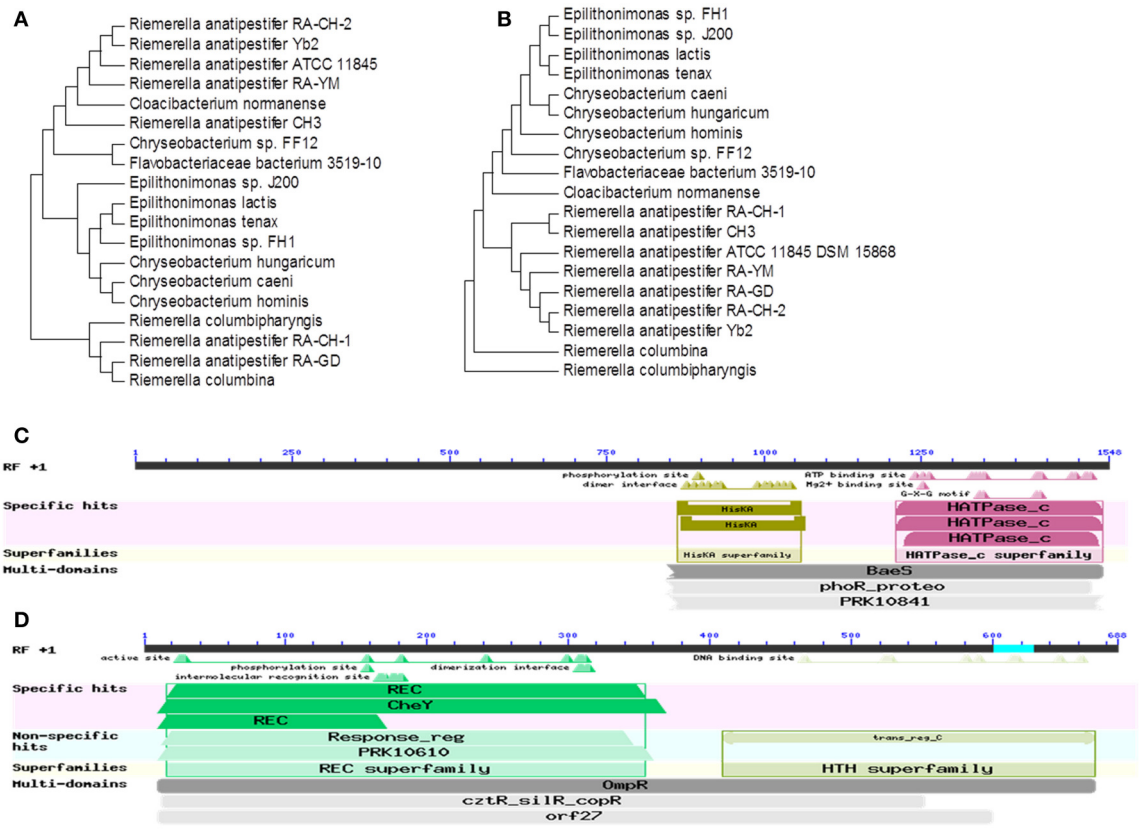


FIGURE 1 | Analys is of the RAYM_RS09735 and RAYM_RS09740 homology in *R. anatipestifer*. The bootstrap percentage from 1,000 replicates is indicated at each node. Rooted neighbor-joining tree of the amino acid sequences of RAYM_RS09735 **(A)** and RAYM_RS09740 **(B)**. Bioinformatics analysis the conserved domains of RAYM_RS09735 **(C)** and RAYM_RS09740 **(D)**.

be elements of a two-component signaling system (TCS). TCS are typically composed of a sensor with histidine kinase activity and a cytoplasmic transcriptional regulator. RAYM_RS09735 was identified as a BaeS family histidine kinase, while RAYM_RS09740 was predicted to be an OmpR family transcriptional regulator (**Figures 1C,D**). Both members of the TCS had the same promoter. Additionally, RAYM_RS09735 was predicted to be a phosphate regulon sensor protein (PhoR).

Characterization of a Mutant YMΔRS09735/RS09740 Strain

The *RS09735/RS09740* gene was deleted from the chromosome of an *R. anatipestifer* YM strain via allelic exchange. The gene was replaced with a $Spec^R$ cassette, allowing successful transfects to be drug selected. The mutant strain was further validated by PCR amplification of *RAYM_RS09735*, *RAYM_RS09740*, and 16S rRNA fragments from transconjugants (**Figure 2A**). Real-time PCR analysis confirmed that *RAYM_RS09735* and *RAYM_RS09740* transcription was completely abolished in the mutant strain. However, inactivation of the *RAYM_RS09735* and *RAYM_RS09740* genes led to significantly increased transcription of the chromosomally upstream gene *RAYM_RS09730* and the downstream gene *RAYM_RS09745* (**Figure 2B**). The *RAYM_RS09735* and *RAYM_RS09740* deletion mutant strain was designated YMΔRS09735/RS09740. Growth curve measurements revealed that growth in TSB was similar between the wild-type YM and mutant strains (**Figure 2C**). Transmission electron microscopy also showed that there were no significant changes in bacterial morphology in the RA-YM *RAYM_RS09735/RAYM_RS09740* mutant strain (data not shown).

Pathogenicity of the YMΔRS09735/RS09740 Mutant

Bacterial virulence was evaluated by median lethal doses (LD_{50}) in 12-day-old Cherry Valley ducks. The LD_{50} for the mutant strain was greater than 10^{11} CFU, which was more than a 10^3-fold attenuation in virulence compared to the wild-type YM strain (4×10^7 CFU). These LD_{50} values demonstrate that the strain is almost avirulent in ducklings (**Table 3**). To further investigate the role of RAYM_RS09735 and RAYM_RS09740 in *in vivo* systemic infections, the bacterial loads in the liver, spleen, heart, brain, and blood from infected ducks were quantified. This showed that, in these organs, bacterial capacity was significantly decreased compared to ducks infected with the wild-type strain (**Figure 3**). Pathological histological analysis was also performed 48 h post-infection comparing the wild-type YM and YMΔRS09735/RS09740 strains. Hematoxylin and eosin staining revealed that the capacity of the YMΔRS09735/RS09740 mutant to damage the heart, brain, and spleen was significantly decreased relative to the wild-type strain, although mild liver damage was still observed (**Figure 4**). These results show that RAYM_RS09735 and RAYM_RS09740 significantly affect the virulence of *R. anatipestifer*.

Function and GO Enrichment Analysis of the Differentially Expressed Genes

Differentially expressed genes (DEGs) between the mutant YMΔRS09735/RS09740 and wild-type strains were identified

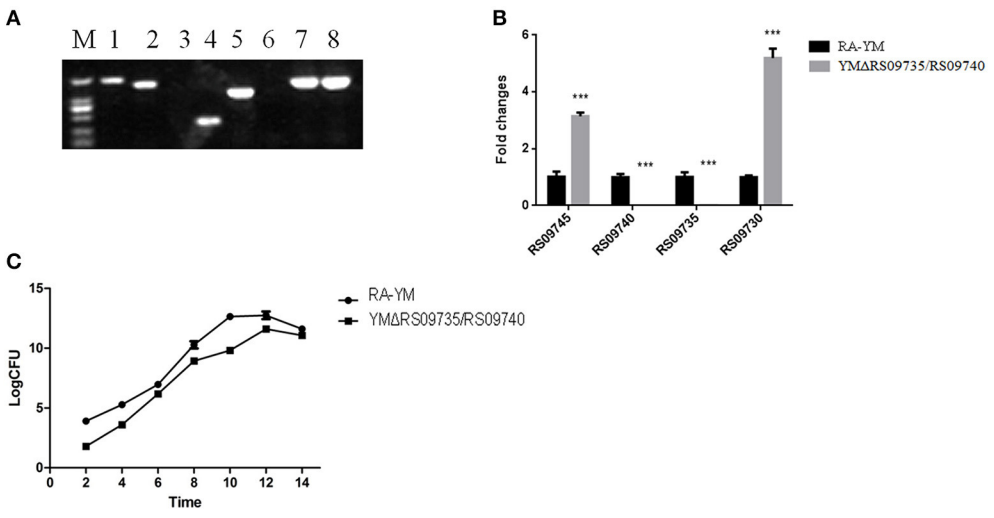


FIGURE 2 | Characteristics of YMΔRS09735/RS09740 strain. (A) PCR amplification. M, DL2000; Lane 1, *RAYM_RS09735/RAYM_RS09740* was amplified from YMΔRS09735/RS09740 strain; Lane 2, *RAYM_RS09735/RAYM_RS09740* was amplified from RA-YM strain; Lane 3, delta genes were amplified from YMΔRS09735/RS09740 strain; Lane 4, delta genes were amplified from RA-YM strain; Lane 5, $Spec^R$ cassette was amplified from YMΔRS09735/RS09740 strain; Lane 6, $Spec^R$ cassette was amplified from RA-YM strain; Lane 7, 16S rRNA fragment was amplified from YMΔRS09735/RS09740 strain; Lane 8, 16S rRNA fragment was amplified from RA-YM strain. **(B)** Real-time PCR analysis. The flanks mRNA levels of *RAYM_RS09735/RAYM_RS09740* genes were measured. The changes of transcription were expressed as fold expression; **(C)** Bacterial growth curves. The mutant strain YMΔRS09735/RS09740 or wild-type strain RA-YM was grown on TSB medium, and growth of each strain was monitored by measuring the CFU/ml. ***$P < 0.001$.

TABLE 3 | Virulence of the wild-type *R. anatipestifer* YM and YMΔRS09735/RS09740 strain.

Strains	Challenge dose (CFU)	Survival/total	LD50 (CFU)
R.anatipestifer	10^5	5/5	4×10^7
	10^6	4/5	
	10^7	3/5	
	10^8	0/5	
	10^9	0/5	
	10^{10}	0/5	
	10^{11}	0/5	
YMYMΔRS09735/RS09740	10^5	5/5	$>10^{11}$
	10^6	5/5	
	10^7	5/5	
	10^8	5/5	
	10^9	5/5	
	10^{10}	5/5	
	10^{11}	5/5	

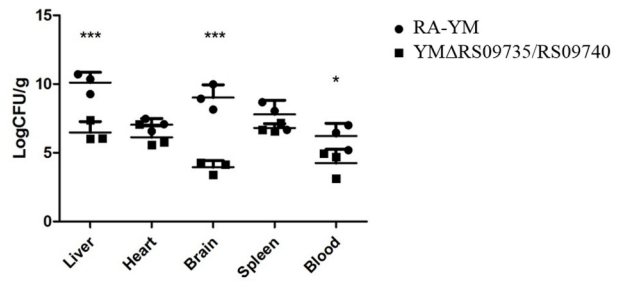


FIGURE 3 | Bacterial loadings in *R. anatipestifer* YM and YMΔRS09735/RS09740 in infected duckling tissues. Twelve 12-day-old ducks infected with the wild-type *R. anatipestifer* strain or mutant strain YMΔRS09735/RS09740. Blood, heart, liver, brain, and spleen samples were collected at 48 h after injected flippers. $^*P < 0.05$, $^{***}P < 0.001$.

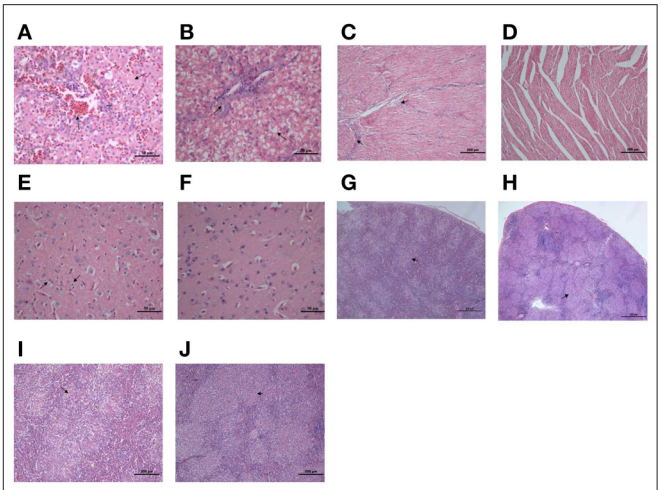


FIGURE 4 | Histopathological analysis of the duck tissues. Liver, heart, brain, and spleen samples were collected from the infected ducks with *R. anatipestifer* YM strain **(A,C,E,G,I)** or YMΔRS09735/RS09740 strain **(B,D,F,H,J)** in 48 h post-challenge, fixed in 10% formalin, and subjected to hematoxylin and eosin staining for a histopathological examination. The "arrow" represents the obvious pathologic change in tissues.

using RNAseq. In total, 805 genes were found to be differentially expressed, with 112 genes upregulated (13.9%), and 693 genes downregulated (86.1%) in the mutant YMΔRS09735/RS09740 strain compared to the wild-type strain (**Figure 5A**). Of the ∼2,000 genes identified in the *R. anatipestifer* genome, more than one third were differentially expressed. Further analysis of the RNAseq data using KEGG pathways predicted that RAYM_RS09735 and RAYM_RS09740 are components of a PhoP/PhoR TCS (**Figure 5B**). Gene Ontology (GO) enrichment analysis showed that differentially expressed genes can be found in a wide-variety of biological processes, cellular components, and molecular functions, including 21 distinct pathways (**Figure 6**). RNAseq was validated using qRT-PCR analysis of randomly selected 10 genes. Overall, the changes in expression of these randomly genes tested by qRT-PCR agreed with the direction determined by RNAseq (**Figure 7A**). Linear regression analysis was performed examining the fold-changes of the gene expression ratios between RNAseq and qRT-PCR, showing a significant positive correlation (**Figure 7B**).

We found that 11 genes had more than 4-fold higher expression in the mutant strain compared to the wild-type, eight of which encode hypothetical proteins. The three other genes encoded a carbohydrate-binding protein, a glycan metabolism protein (RagB), and a phosphate subunit transfer protein (PstS), respectively. We also found that several transcription factors, components of the CRISPR system, and additional putative proteins were also upregulated, with most upregulated genes involved in bacterial metabolism. In addition, 20 genes were found to be downregulated more than 4-fold. Of these, 13 genes encoded hypothetical proteins, while the other seven encoded RAYM_RS09735, RAYM_RS09740, a lipoprotein, a peptidoglycan hydrate (Nlp/P60), a DNA-binding protein, von Willebrand factor A, ATPase AAA, and an uncharacterized conserved protein. Other downregulated genes included several hypothetical proteins, transcription factors, and metabolic genes, in addition to multiple molecular chaperones and TonB-dependent receptors.

Finally, we performed real-time PCR analysis of three predicted virulence-associated genes expression in the YMΔRS09735/RS09740 mutant strain (*pstS*, *BLP*, and *Nlp/P60*). We found that the mRNA expression of *pstS* was 8-fold higher (**Figure 8A**), consistent with our RNAseq data. *Bacterial lipoprotein* (*BLP*) and *hydrolase Nlp/P60* were found to be significantly downregulated (**Figure 8B**).

DISCUSSION

Bacterial two-component regulatory systems (TCSs), consisting of a sensing histidine kinase and a response regulator, mediate gene expression in response to environmental stimuli

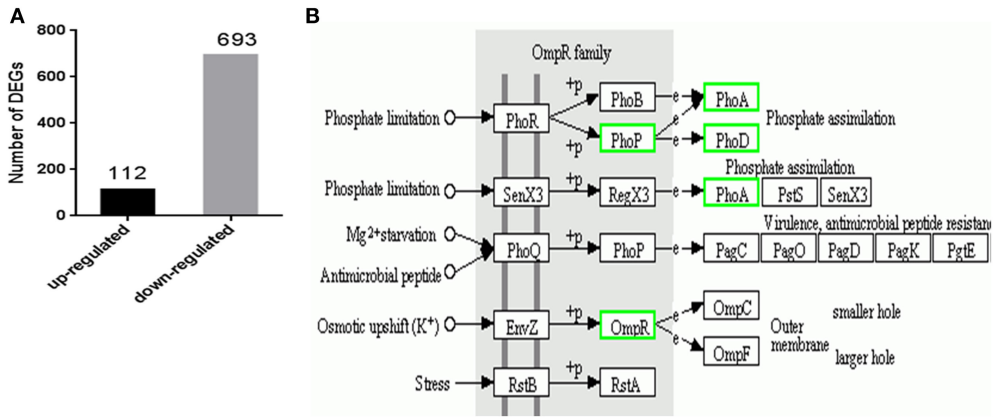


FIGURE 5 | Differential expression genes and functional classification of YMΔRS09735/RS09740 mutant strain. (A) RNASeq analysis was performed to identify differentially expressed genes in mutant strain YMΔRS09735/RS09740. **(B)** RAYM_RS09740 in KEGG pathway from RNASeq analysis.

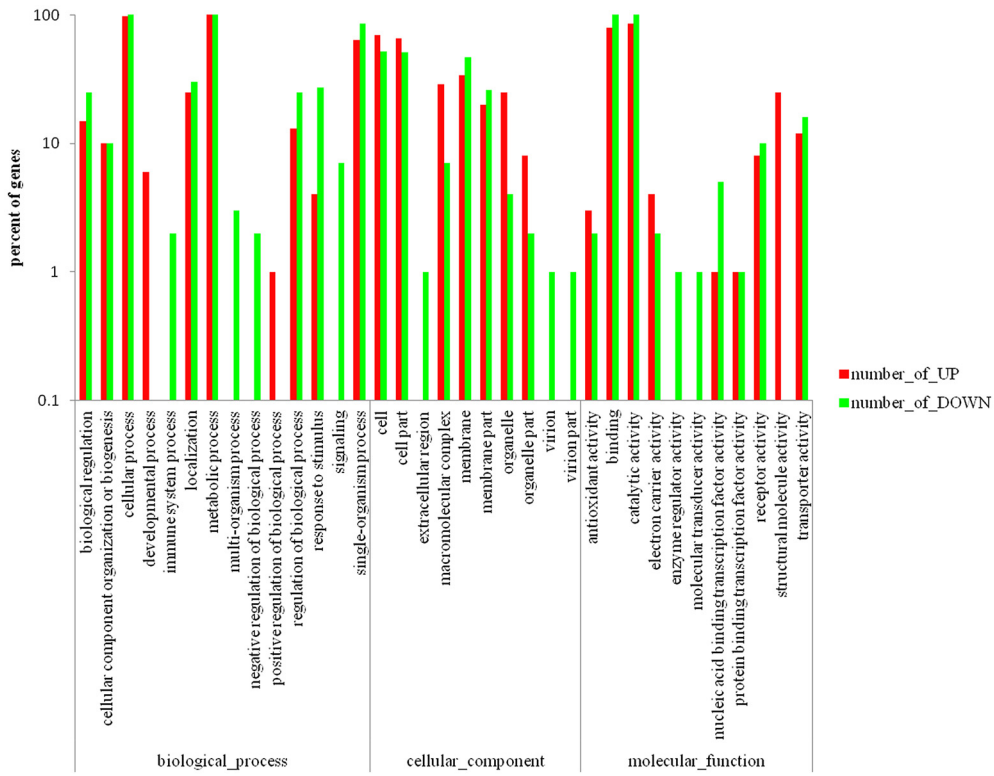


FIGURE 6 | Gene ontology enrichment analysis of the differentially expressed genes scattered to biological-process, cellular-component, and molecular-function, which including 21 pathways.

(Stock et al., 2003). To achieve this, histidine kinases sense environmental signals and autophosphorylate. This additional phosphate is then transferred to an aspartic acid residue on the corresponding response regulator (Hoch, 2000; Alm et al., 2006). The phosphorylated response regulator elicits a diverse range of downstream responses, including an enhancement of its DNA binding ability. This allows the TCS to modulate target gene expression (West and Stock, 2001; Lin et al., 2015). Using an *in vivo* antigen technology (IVIAT) to examine ducks infected with *R. anatipestifer*, we have identified a putative TCS involving the genes *RAYM_RS09740* and *RAYM_RS09735*. Bioinformatics analysis of these two genes, and their proteins, revealed that *RAYM_RS09740* and *RAYM_RS09735* encode a histidine kinase and response regulator, respectively. We also found that, like

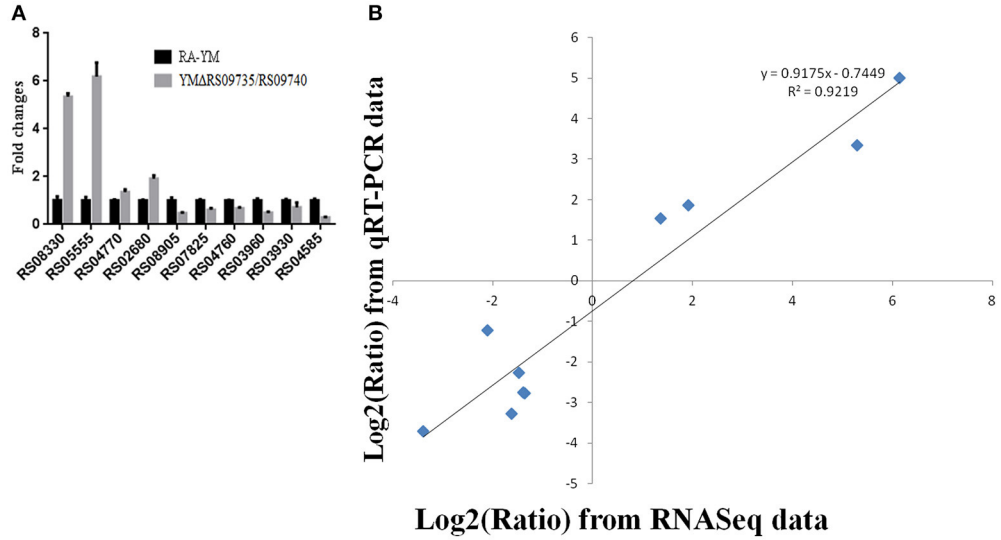


FIGURE 7 | Real-time PCR verification of differentially expressed genes in YMΔRS09735/RS09740 mutant strain. (A) The fold expression changes of randomly selected 10 genes to validity of the RNAseq results. **(B)** Coefficient analysis of fold changes data between qRT-PCR and RNAseq. Ten different expression genes were selected from qRT-PCR.

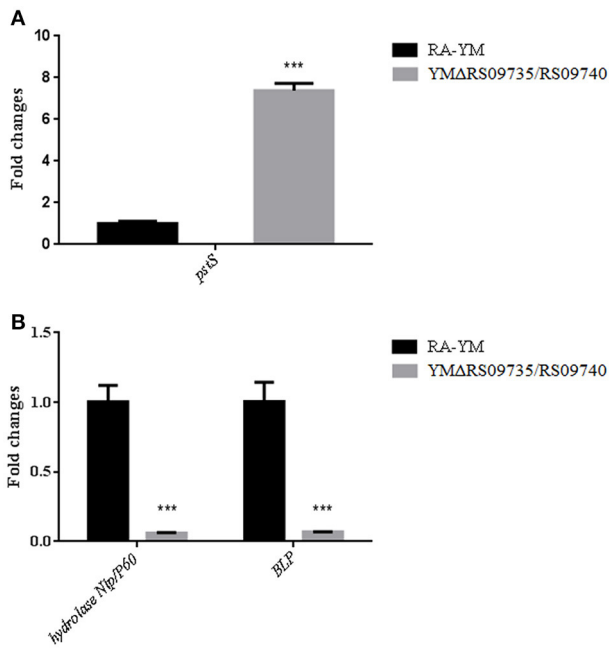


FIGURE 8 | Real-time PCR verification of three predicted virulence-associated genes expression in YMΔRS09735/RS09740 mutant strain. (A) The fold expression changes of *pstS* gene in mutant strain. **(B)** The fold expression changes of two predicted virulent genes *hydrolase Nlp/P60* and *BLP* in mutant strain. ***$P < 0.001$.

most TCSs, *RAYM_RS09735* and *RAYM_RS09740* have the same promoter and co-transcribe as an operon (Aggarwal et al., 2016). RAYM_RS09735 and RAYM_RS09740 possess all of the characteristic TCS domains essential to their biochemical activities and responses. RAYM_RS09735 and RAYM_RS09740 were predicted to belong to the OmpR and BaeS families, respectively, and RAYM_RS09735 was further predicted to be a phosphate regulon sensor protein (PhoR).

Our study aimed to investigate the RAYM_RS09735/RAYM_RS09740 TCS and explore its potential functions in *R. anatipestifer*. To better evaluate the role of RAYM_RS09735 and RAYM_RS09740 in *R. anatipestifer*, we constructed the mutant strain YMΔRS09735/RS09740 in which the *RAYM_RS09735* and *RAYM_RS09740* genes were mutated by deleting a 1400-bp fragment from the wild-type RA-YM strain. RT-PCR and qRT-PCR were used to validate the mutant strain (**Figures 2A,B**). We found that upstream and downstream genes were significantly down-regulated in the YMΔRS09735/RS09740 mutant, although the strain showed the same growth characteristics as the wild-type in TSB (**Figure 2C**).

LD_{50} analysis of the two strains demonstrated the *RAYM_RS09735/RAYM_RS09740* mutant had significantly reduced virulence. Indeed, the strain was almost avirulent to ducklings (**Table 3**). A normal *R. anatipestifer* infection is characterized by septicemia and the ability to colonize and develop in the tissues of their host. We found that the bacterial loads in the liver, spleen, heart, brain, and blood of ducks infected with the mutant YMΔRS09735/RS09740 strain were significantly lower than that of wild-type RA-YM infected ducklings (**Figure 3**). In addition, pathological histological analysis using hematoxylin and eosin staining showed that the mutant strain also induced significantly less damage to the heart, brain, and spleen, compared to the wild-type strain. There was still slight damage to the liver (**Figure 4**). Our results indicate that the *RAYM_RS09735* and *RAYM_RS09740* genes are important mediators of *R. anatipestifer* virulence during infection in

ducklings. Therefore, RAYM_RS09740 was determined to be a response regulator, with a combined promoter affecting downstream genes. The protein also altered the expression of target genes. Our study therefore indicated that virulence factors in *R. anatipestifer* could be regulated by RAYM_RS09740.

RAYM_RS09735 and RAYM_RS09740 share up to 70% sequence identity with *Flavobacteriaceae* species, such as *Cloacibacterium*, *Epilithonimonas*, and *Chryseobacterium*. These species have been previously reported to exhibit drug resistance. Molecular docking analysis identified inhibitors of various RRs in OmpR family to affect this observed drug resistance. Therefore targeting TCSs has become an attractive option for the development of new drugs, particularly to combat *M. tuberculosis* (Banerjee et al., 2016). These studies also demonstrate that TCSs can be used as the theoretical basis for the study of drug resistance mechanisms in *Flavobacteriaceae*.

To investigate how RAYM_RS09735 and RAYM_RS09740 regulate gene expression, RNAseq analysis was performed to identify differentially expressed genes in the mutant YMΔRS09735/RS09740 strain. This revealed 112 genes that were upregulated (13.9%) and 693 genes that were down-regulated (86.1%) relative to the wild-type strain (approximately one-third of genes had differential expression). To validate the RNAseq results, we randomly assessed the differential expression of 10 genes using qRT-PCR, confirming the accuracy of the RNAseq data (**Figure 7**). Gene ontology (GO) enrichment analysis showed that differentially expressed genes were involved in a wide-variety of biological processes, cellular components, and molecular functions, including 21 distinct pathways (**Figure 6**). Therefore, we hypothesize that the RAYM_RS09735/RAYM_RS09740 TCS is a global expression regulator that controls the expression of virulence genes, thus affecting the pathogenicity of *R. anatipestifer*.

KEGG pathway analysis showed that RAYM_RS09740 is a PhoP protein, indicating that RAYM_RS09735 and RAYM_RS09740 form a PhoP/PhoR TCS. PhoP/PhoR TCSs have been implicated in several biological processes in Gram-positive bacteria, such as *B. subtilis* (Hulett et al., 1994), *Streptomyces lividans* (Sola-Landa et al., 2003), and *M. tuberculosis* (Perez et al., 2001). Specifically, PhoP/PhoR has been shown to control respiration(Birkey et al., 1998), cell wall metabolism (Minnig et al., 2005), culture metabolism (Thomas et al., 2012), and biofilm formation (Bluskadosh et al., 2013). Most importantly, in many pathogens PhoP/PhoR TCSs have also been shown to regulate pathogenesis (Perez et al., 2001; Ryndak et al., 2008). PhoP and PhoQ TCS components have been reported in many Gram-negative bacteria, including *E. coli* (Eguchi et al., 2007), *Edwardsvilla* (Lv et al., 2012), and *Salmonella* (Tran et al., 2016). PhoP/PhoR is also an important TCS in *M. tuberculosis* and is a key player in virulence (Ryndak et al., 2008). However, PhoP/PhoR TCSs have not been previously observed in Gram-negative bacteria. Our data shows that a putative PhoP/PhoR TCS is involved in *R. anatipestifer* virulence, although further phosphotransfer profiling will be required to validate RAYM_RS09735/RAYM_RS09740 as a PhoP/PhoR TCS.

Our data also suggest that RAYM_RS09740 can act as a transcription factor and plays a role in regulating genes expression, although there are likely other TCSs present in *R. anatipestifer*. This is based on our observation that, after the deletion of *RAYM_RS09735* and *RAYM_RS09740*, other TCSs regulated gene expression. We hypothesize that cross-talk and -regulation may occur between a multitude of two-component systems in *R. anatipestifer*, similarly to other organisms. For example, the transcription factor NemR is a redox-regulated transcriptional repressor in *E. coli* (Gray et al., 2013) and a carbohydrate-responsive regulatory protein (BadR) is a transcriptional repressor of *rpoS* in *Borrelia burgdorferi* (Miller et al., 2013).

The PhoP/PhoR TCS we identified was initially predicted to be part of the phosphate (Pi) stress-response in *R. anatipestifer*. PhoR acts to sense environmental phosphate (Pi) levels in bacteria through the ABC-type phosphate-specific transport (Pst) system and the protein PhoU. PstS is a periplasmic protein that binds Pi with high affinity and PhoP induces the promoter activity of *pstS*. We found with RNAseq and qRT-PCR that the mRNA expression of *pstS* was 8-fold higher (**Figure 8A**) in the mutant strain and different from other Gram-negative bacteria. Therefore, RAYM_RS09735 and RAYM_RS09740 may act more like the PhoP/PhoR system of *M. tuberculosis*. Additionally, *M. tuberculosis* PhoP/PhoR does not respond to phosphate starvation and is similar to the PhoP/PhoQ TCS in *Salmonella*. In *Salmonella* PhoP/PhoQ TCS, PhoQ senses low Mg^{2+} levels, and PhoP activates expression of genes encoding high-affinity Mg^{2+} transport systems (Walters, 2006). We aim to investigate the environmental factors sensed by *R. anatipestifer* PhoP/PhoR in future work.

Further, DEG analysis showed that the expression of several other signal transduction system genes was altered in the YMΔRS09735/RS09740 strain. It is likely that PhoP can influence the expression of transcription factors in other signal transduction system and there is cross-talk or cross regulation between PhoP/PhoR and other TCSs in *R. anatipestifer*. As more than one third of genes demonstrated differential expression, it is unlikely that the PhoP/PhoR TCS directly regulates all of these DEGs. However, the TCS we identified may affect wider gene expression by influencing the expression of other transcription factors in a variety of signal transduction systems.

Additionally, the CRISPR-Cas bacterial immune system has been shown to be present in *R. anatipestifer* (Chamnongpol et al., 2003; Zhu et al., 2016) and we found that the expression of *cas9*, *cas1*, and *cas2* were all different in the YMΔRS09735/RS09740 mutant, compared to the wild-type. Recent studies have reported that quorum sensing has a role in controlling the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system (Høyland-Kroghsbo et al., 2016) and the TCS BfiSR affects the production of proteins involved in virulence, post-translational modification, and quorum sensing (Petrova and Sauer, 2010). We therefore reasoned that TCS might also affect the CRISPR-Cas immune system of *R. anatipestifer*.

We found 13 genes encoding hypothetical proteins that were downregulated more than 4-fold. At present, M949_RS01915, M949_1360, AS87_01735, AS87_03730, M949_1556, and AS87_04050 are all hypothetical proteins but have been reported to associate with virulence in *R. anatipestifer*.

Bacterial lipoprotein (BLP) and the hydrolase Nlp/P60 were both found to be significant down-regulated in the mutant YMΔRS09735/RS09740 strain (**Figure 8B**). Nlp/P60 is involved in cell growth and division, autolysis, and invasion (Xu et al., 2015). Bacterial lipoproteins on the cell surface include an important class of virulence factor that is expressed by many pathogens. BLP can promote TLR2 expression, TLR2-induced NF-κB activation, and IL-6 production, leading to immune inflammation (Buddelmeijer, 2015). The clinical symptoms of *R. anatipestifersis* are characterized by fibrinous pericarditis, glissonitis, airbag inflammation, and meningitis. Our pathological results demonstrated that the damage induced by the mutant YMΔRS09735/RS09740 strain to the heart, brain, and spleen was significantly less severe than the wild-type strain (**Figure 4**). We therefore hypothesize that *R. anatipestifer* reduced expression of BLP and the hydrolase Nlp/P60, subsequently affecting inflammation in ducklings.

Other downregulated genes we identified included transcription factors, metabolic genes, multiple molecular chaperones, and TonB-dependent receptors. The primary role of DNA binding proteins and transcription factors is to regulate the expression of the genes and RAYM_RS09740 may regulate transcription by modifying the expression of these transcription factors. Molecular chaperone proteins assist in the folding of nascent polypeptide chains and prevent the aggregation of denatured proteins (Hartl and Hayerhartl, 2002). We found that transcript for the molecular chaperones *dnaJ, dnaK, grpE, clpP, tir,* and *skp* were all downregulated in the mutant YMΔRS09735/RS09740 strain, suggesting that the RAYM_RS09735/RAYM_RS09740 TCS may regulate the expression of the chaperone genes to influence the severity of *R. anatipestifersis*. The DnaK molecular chaperone system includes DnaK, DnaJ, and GrpE (Mayer et al., 2000; Ben-Zvi and Goloubinoff, 2001; Genevaux et al., 2007) and is related to stress tolerance and virulence in *Salmonella*. DnaK can also regulate the expression of virulence islands and a DnaK mutation strain was found to be largely avirulent (Takaya et al., 2004). ClpP was found to be a virulence-related factor in *Actinobacillus pleuropneumoniae*, with a *clpP* gene deletion strain demonstrating decreased biofilm production (Xie et al., 2013). We also found that TonB-dependent receptors were downregulated. These receptors have been reported to associate with hemin iron acquisition and *R. anatipestifer* virulence.

In this study, RAYM_RS09735 and RAYM_RS09740 were predicted to be components of a PhoP/PhoR TCS in *R. anatipestifer* using bioinformatics analysis. Furthermore, deletion of the *RAYM_RS09735* and *RAYM_RS09740* genes significantly decreased *R. anatipestifer* virulence. RNAseq analysis showed that RAYM_RS09740 is a global expression regulator in *R. anatipestifer*. We also demonstrated that the RAYM_RS09735/RAYM_RS09740 TCS contributes to virulence, signal translation, and the CRISPR-Cas adaptive immune system of *R. anatipestifer*. BLP and the hydrolase Nlp/P60 were predicted to be two virulence factors involved. Although common in Gram-positive bacteria, the predicted PhoP/PhoR TCS we have putatively characterized is the first such TCS identified in Gram-negative bacteria. Our study is also the first to report this TCS in *R. anatipestifer*, providing new insight into the role of TCSs in regulating *R. anatipestifer* virulence.

ETHICS STATEMENT

All the animal experiments were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals from Research Ethics Committee, Huazhong Agricultural University, Hubei, China. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

AUTHOR CONTRIBUTIONS

Funding acquisition: ZL. Investigation: YW. Methodology: YW, TL, and XY. Project administration: SL, ML, SH, and DB. Resources: ZZ. Supervision: ZL. Validation: ZL. Visualization: YW. Writing—original draft: YW. Writing—review's editing: ZL.

FUNDING

This work was supported by grants from the Wuhan Science and Technology Bureau (No. 2015020101010070 to ZL) and Natural Science Foundation of Hubei Province (No. 2015CFB268 to ZL).

REFERENCES

Aggarwal, S., Somani, V. K., Gupta, V., Kaur, J., Singh, D., Grover, A., et al. (2016). Functional characterization of PhoPR two component system and its implication in regulating phosphate homeostasis in *Bacillus anthracis*. *Biochim. Biophys. Acta* 1861, 2956–2970. doi: 10.1016/j.bbagen.2016.09.022

Alm, E., Huang, K., and Arkin, A. (2006). The Evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS Comput. Biol.* 2:e143. doi: 10.1371/journal.pcbi.0020143

Banerjee, S. K., Kumar, M., Alokam, R., Sharma, A. K., Chatterjee, A., Kumar, R., et al. (2016). Targeting multiple response regulators of *Mycobacterium tuberculosis* augments the host immune response to infection. *Sci. Rep.* 6:25851. doi: 10.1038/srep25851

Ben-Zvi, A. P., and Goloubinoff, P. (2001). Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *J. Struct. Biol.* 135, 84–93. doi: 10.1006/jsbi.2001.4352

Birkey, S. M., Liu, W., Zhang, X., Duggan, M. F., and Hulett, F. M. (1998). Pho signal transduction network reveals direct transcriptional regulation of one two-component system by another two-component regulator: *Bacillus subtilis* PhoP directly regulates production of ResD. *Mol. Microbiol.* 30, 943–953. doi: 10.1046/j.1365-2958.1998.01122.x

Bluskadosh, I., Zilka, A., Yerushalmi, G., and Banin, E. (2013). The effect of pstS and phoB on quorum sensing and swarming motility in *Pseudomonas aeruginosa*. *PLoS ONE* 8:e74444. doi: 10.1371/journal.pone.0074444

Buddelmeijer, N. (2015). The molecular mechanism of bacterial lipoprotein modification—How, when and why? *FEMS. Microbiol. Rev.* 39, 246–261. doi: 10.1093/femsre/fuu006

Chamnongpol, S., Cromie, M., and Groisman, E. A. (2003). Mg^{2+} sensing by the Mg^{2+} sensor PhoQ of *Salmonella enterica*. *J. Mol. Biol.* 325, 795–807. doi: 10.1016/S0022-2836(02)01268-8

Chang, C. F., Hung, P. E., and Chang, Y. F. (1998). Molecular characterization of a plasmid isolated from *Riemerella anatipestifer*. *Avian Pathol.* 27, 339–345. doi: 10.1080/03079459808419349

Crasta, K. C., Chua, K. L., Subramaniam, S., Frey, J., Loh, H., and Tan, H. M. (2002). Identification and characterization of CAMP cohemolysin as a potential virulence factor of *Riemerella anatipestifer*. *J. Bacteriol.* 184, 1932–1939. doi: 10.1128/JB.184.7.1932-1939.2002

Edwards, R. A., Keller, L. H., and Schifferli, D. M. (1998). Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene* 207, 149–157. doi: 10.1016/S0378-1119(97)00619-7

Eguchi, Y., Itou, J., Yamane, M., Demizu, R., Yamato, F., Okada, A., et al. (2007). B1500, a small membrane protein, connects the two-component systems EvgS/EvgA and PhoQ/PhoP in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18712–18717. doi: 10.1073/pnas.0705768104

Genevaux, P., Georgopoulos, C., and Kelley, W. L. (2007). The Hsp70 chaperone machines of *Escherichia coli*: a paradigm for the repartition of chaperone functions. *Mol. Microbiol.* 66, 840–857. doi: 10.1111/j.1365-2958.2007.05961.x

Gray, M. J., Wholey, W. Y., Parker, B. W., Kim, M., and Jakob, U. (2013). NemR is a bleach-sensing transcription factor. *J. Biol. Chem.* 288, 13789–13798. doi: 10.1074/jbc.M113.454421

Høyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., et al. (2016). Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc. Natl. Acad. Sci. U.S.A.* 114, 131–135. doi: 10.1073/pnas.1617415113

Hartl, F. U., and Hayerhartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852–1858. doi: 10.1126/science.1068408

Haydel, S. E., Malhotra, V., Cornelison, G. L., and Clark-Curtiss (2012). The prrAB two-component system is essential for *Mycobacterium tuberculosis* viability and is induced under nitrogen-limiting conditions. *J. Bacteriol.* 194, 354–361. doi: 10.1128/JB.06258-11

He, H., Hovey, R., Kane, J., Singh, V., and Zahrt, T. C. (2006). MprAB is a stress-responsive two-component system that directly regulates expression of sigma factors SigB and SigE in *Mycobacterium tuberculosis*. *J. Bacteriol.* 188, 2134–2143. doi: 10.1128/JB.188.6.2134-2143.2006

Hoch, J. A. (2000). Two-component and phosphorelay signal transduction. *Curr. Microbiol.* 3, 165–170. doi: 10.1016/S1369-5274(00)00070-9

Hu, Q., Han, X., Zhou, X., Ding, C., Zhu, Y., and Yu, S. (2011). OmpA is a virulence factor of *Riemerella anatipestifer*. *Vet. Microbiol.* 150, 278–283. doi: 10.1016/j.vetmic.2011.01.022

Hu, Q., Han, X., Zhou, X., Ding, S., Ding, C., and Yu, S. (2010). Characterization of biofilm formation by *Riemerella anatipestifer*. *Vet. Microbiol.* 144, 429–436. doi: 10.1016/j.vetmic.2010.02.023

Hulett, F., Lee, J., Shi, L., Sun, G., Chesnut, R., Sharkova, E., et al. (1994). Sequential action of two-component genetic switches regulates the PHO regulon in *Bacillus subtilis*. *J. Bacteriol.* 176, 1348–1358. doi: 10.1128/jb.176.5.1348-1358.1994

Lin, M. F., Lin, Y. Y., and Lan, C. Y. (2015). The role of the two-component system baesr in disposing chemicals through regulating transporter systems in *Acinetobacter baumannii*. *PLoS ONE* 10:e0132843. doi: 10.1371/journal.pone.0132843

Lv, Y., Xiao, J., Liu, Q., Wu, H., Zhang, Y., and Wang, Q. (2012). Systematic mutation analysis of two-component signal transduction systems reveals EsrA-EsrB and PhoP-PhoQ as the major virulence regulators in *Edwardsiella tarda*. *Vet. Microbiol.* 157, 190–199. doi: 10.1016/j.vetmic.2011.12.018

Mayer, M. P., Rüdiger, S., and Bukau, B. (2000). Molecular basis for interactions of the DnaK chaperone with substrates. *Biol. Chem.* 381, 877–885. doi: 10.1515/BC.2000.109

McCluskey, J., Hinds, J., Husain, S., Witney, A., and Mitchell, T. J. (2004). A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* 51, 1661–1675. doi: 10.1111/j.1365-2958.2003.03917.x

Mike, L. A., Choby, J. E., Brinkman, P. R., Olive, L. Q., Dutter, B. F., Ivan, S. J., et al. (2014). Two-component system cross-regulation integrates *Bacillus anthracis* response to heme and cell envelope stress. *PLoS Pathog.* 10:e1004044. doi: 10.1371/journal.ppat.1004044

Miller, C. L., Karna, S. L., and Seshu, J. (2013). Borrelia host adaptation Regulator (BadR) regulates rpoS to modulate host adaptation and virulence factors in *Borrelia burgdorferi*. *Mol. Microbiol.* 88, 105–124. doi: 10.1111/mmi.12171

Minnig, K., Lazarevic, V., Soldo, B., and Mauël, C. (2005). Analysis of teichoic acid biosynthesis regulation reveals that the extracytoplasmic function sigma factor σM is induced by phosphate depletion in *Bacillus subtilis* W23. *Microbiology* 151, 3041–3049. doi: 10.1099/mic.0.28021-0

O'Loughlin, J. L., Spinner, J. L., Minnich, S. A., and Kobayashi, S. D. (2010). Yersinia pestis two-component gene regulatory systems promote survival in human neutrophils. *Infect. Immun.* 78, 773–782. doi: 10.1128/IAI.00718-09

Pang, X., Vu, P., Byrd, T. F., Ghanny, S., Soteropoulos, P., Mukamolova, G. V., et al. (2007). Evidence for complex interactions of stress-associated regulons in an mprAB deletion mutant of *Mycobacterium tuberculosis*. *Microbiology* 153, 1229–1242. doi: 10.1099/mic.0.29281-0

Parish, T., Smith, D. A., Roberts, G., Betts, J., and Stoker, N. G. (2003). The senX3-regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. *Microbiology* 149, 1423–1435. doi: 10.1099/mic.0.26245-0

Pathanasophon, P., Phuektes, P., Tanticharoenyos, T., Narongsak, W., and Sawada, T. (2002). A potential new serotype of *Riemerella anatipestifer* isolated from ducks in Thailand. *Avian Pathol.* 31, 267–270. doi: 10.1080/03079450220136576

Pathanasophon, P., Sawada, T., and Tanticharoenyos, T. (1995). New serotypes of *Riemerella anatipestifer* isolated from ducks in Thailand. *Avian Pathol.* 24, 195–199. doi: 10.1080/03079459508419059

Perez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B., and Martin, C. (2001). An essential role for phoP in *Mycobacterium tuberculosis* virulence. *Mol. Microbiol.* 41, 179–187. doi: 10.1046/j.1365-2958.2001.02500.x

Petrova, O. E., and Sauer, K. (2010). The novel two-component regulatory system BfiSR regulates biofilm development by controlling the small RNA rsmZ through CafA. *J. Bacteriol.* 192, 5275–5288. doi: 10.1128/JB.00387-10

Roland, K., Curtiss, R., and Sizemore, D. (1999). Construction and evaluation of a delta cya delta crp *Salmonella typhimurium* strain expressing avian pathogenic *Escherichia coli* O78 LPS as a vaccine to prevent airsacculitis in chickens. *Avian Dis.* 43, 429–441. doi: 10.2307/1592640

Ryndak, M., Wang, S., and Smith, I. (2008). PhoP, a key player in *Mycobacterium tuberculosis* virulence. *Trends Microbiol.* 16, 528–534. doi: 10.1016/j.tim.2008.08.006

Segers, P., Mannheim, W., Vancanneyt, M., De Brandt, K., Hinz, K. H., Kersters, K., et al. (1993). *Riemerella anatipestifer* gen. nov., comb. nov., the causative agent of septicemia anserum exsudativa, and its phylogenetic affiliation within the Flavobacterium-Cytophaga rRNA homology group. *Int. J. Syst. Bacteriol.* 43, 768–776. doi: 10.1099/00207713-43-4-768

Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., et al. (2008). Rewiring the specificity of two-component signal transduction systems. *Cell* 133, 1043–1054. doi: 10.1016/j.cell.2008.04.040

Sola-Landa, A., Moura, R., and Martin, J. (2003). The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6133–6138. doi: 10.1073/pnas.0931429100

Song, X. H., Zhou, W. S., Wang, J. B., Liu, M. F., Wang, M. S., Cheng, A. C., et al. (2016). Genome sequence of *Riemerella anatipestifer* strain RCAD0122, a multidrug-resistant isolate from ducks. *Genome Announc.* 4, e00332–e00316. doi: 10.1128/genomeA.00332-16

Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2003). Two-component signal transduction. *Annu. Rev. Biochem.* 69, 183–215. doi: 10.1146/annurev.biochem.69.1.183

Takaya, A., Tomoyasu, T., Matsui, H., and Yamamoto, T. (2004). The DnaK/DnaJ chaperone machinery of *Salmonella enterica* serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. *Infect. Immun.* 72, 1364–1373. doi: 10.1128/IAI.72.3.1364-1373.2004

Thomas, L., Hodgson, D. A., Wentzel, A., Nieselt, K., Ellingsen, T. E., Moore, J., et al. (2012). Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and phoP mutant grown in batch culture. *Mol. Cell Proteomics.* 11:M111.013797. doi: 10.1074/mcp.m111.013797

Tobe, T. (2008). The roles of two-component systems in virulence of pathogenic *Escherichia coli* and Shigella spp. *Adv. Exp. Med. Biol.* 631, 189–199. doi: 10.1007/978-0-387-78885-2_13

Tran, T. K., Han, Q. Q., Shi, Y., and Guo, L. (2016). A comparative proteomic analysis of *Salmonella typhimurium* under the regulation of the

RstA/RstB and PhoP/PhoQ systems. *Biochim. Biophys. Acta* 1864, 1686–1695. doi: 10.1016/j.bbapap.2016.09.003

Walters, S. B. E. A. (2006). The *Mycobacterium tuberculosis* PhoPR twocomponent system regulates genes essential for virulence and complex lipid biosynthesis. *Mol. Microbiol.* 60, 312–330. doi: 10.1111/j.1365-2958.2006.05102.x

Wang, X., Ding, C., Wang, S., Han, X., and Yu, S. (2015). Whole-genome sequence analysis and genome-wide virulence gene identification of *Riemerella anatipestifer* strain Yb2. *Appl. Environ. Microbiol.* 81, 5093–5102. doi: 10.1128/AEM.00828-15

Wang, X., Liu, B., Dou, Y., Fan, H., Wang, S., Li, T., et al. (2016). The *Riemerella anatipestifer* AS87_01735 gene encodes nicotinamidase PncA, an important virulence factor. *Appl. Environ. Microbiol.* 82, 5815–5823. doi: 10.1128/AEM.01829-16

West, A. H., and Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends. Biochem. Sci.* 26, 369–376. doi: 10.1016/S0968-0004(01)01852-7

Xie, F., Zhang, Y., Li, G., Zhou, L., Liu, S., and Wang, C. (2013). The ClpP protease is required for the stress tolerance and biofilm formation in *Actinobacillus pleuropneumoniae*. *PLoS ONE* 8:e53600. doi: 10.1371/journal.pone.0053600

Xu, Q., Mengin-Lecreulx, D., Liu, X. W., Patin, D., Farr, C. L., Grant, J. C., et al. (2015). Insights into substrate specificity of NlpC/P60 cell wall hydrolases containing bacterial SH3 domains. *MBio* 6, e02327–e02314. doi: 10.1128/mBio.02327-14

Yu, G., Wang, X., Dou, Y., Wang, S., Tian, M., Qi, J., et al. (2016). *Riemerella anatipestifer* M949_1360 gene functions on the lipopolysaccharide biosynthesis and bacterial virulence. *PLoS ONE* 11:e0160708. doi: 10.1371/journal.pone.0160708

Yuan, J., Liu, W., Sun, M., Song, S., Cai, J., and Hu, S. (2011). Complete genome sequence of the pathogenic bacterium *Riemerella anatipestifer* strain RA-GD. *J. Bacteriol.* 193, 2896–2897. doi: 10.1128/JB.00301-11

Zhang, T., Zhang, R., Luo, Q., Wen, G., Ai, D., Wang, H., et al. (2015). Genome sequence of avirulent *Riemerella anatipestifer* strain RA-JLLY. *Genome Announc.* 3, e00895–e00815. doi: 10.1128/genomeA.00895-15

Zhou, Z., Peng, X., Xiao, Y., Wang, X., Guo, Z., Zhu, L., et al. (2011). Genome sequence of poultry pathogen *Riemerella anatipestifer* strain RA-YM. *J. Bacteriol.* 193, 1284–1285. doi: 10.1128/JB.01445-10

Zhu, D. K., Yang, X. Q., He, Y., Zhou, W. S., Song, X. H., Wang, J. B., et al. (2016). Comparative genomic analysis identifies structural features of CRISPR-Cas systems in *Riemerella anatipestifer*. *BMC Genomics* 17:689. doi: 10.1186/s12864-016-3040-4

Zou, J., Wang, X., Tian, M., Cao, S., Hou, W., Wang, S., et al. (2015). The M949_1556 gene plays a role on the bacterial antigenicity and pathogenicity of *Riemerella anatipestifer*. *Vet. Microbiol.* 177, 193–200. doi: 10.1016/j.vetmic.2015.03.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Wang, Lu, Yin, Zhou, Li, Liu, Hu, Bi and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 September 2017
doi: 10.3389/fmicb.2017.01812

Studying the Differences of Bacterial Metabolome and Microbiome in the Colon between Landrace and Meihua Piglets

Shijuan Yan[1†], Cui Zhu[1†], Ting Yu[1†], Wenjie Huang[1], Jianfeng Huang[1,2], Qian Kong[1], Jingfang Shi[1], Zhongjian Chen[1], Qinjian Liu[1], Shaolei Wang[1], Zongyong Jiang[1,3] and Zhuang Chen[1*]*

[1] *Agro-biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou, China,* [2] *Brain Science Institute, South China Normal University, Guangzhou, China,* [3] *Ministry of Agriculture Key Laboratory of Animal Nutrition and Feed Science, Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou, China*

OPEN ACCESS

Edited by:
Florence Abram,
NUI Galway, Ireland

Reviewed by:
Biswarup Sen,
Tianjin University, China
Jayashree Ray,
Lawrence Berkeley National Laboratory, United States

****Correspondence:***
Zongyong Jiang
jiangzy@gdaas.cn
Zhuang Chen
chenzhuang@agrogene.ac.cn

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *18 July 2017*
Accepted: *05 September 2017*
Published: *21 September 2017*

Citation:
Yan S, Zhu C, Yu T, Huang W, Huang J, Kong Q, Shi J, Chen Z, Liu Q, Wang S, Jiang Z and Chen Z (2017) Studying the Differences of Bacterial Metabolome and Microbiome in the Colon between Landrace and Meihua Piglets. Front. Microbiol. 8:1812. doi: 10.3389/fmicb.2017.01812

This study was conducted to compare the microbiome and metabolome differences in the colon lumen from two pig breeds with different genetic backgrounds. Fourteen weaned piglets at 30 days of age, including seven Landrace piglets (a lean-type pig breed with a fast growth rate) and seven Meihua piglets (a fatty-type Chinese local pig breed with a slow growth rate), were fed the same diets for 35 days. Untargeted metabolomics analyses showed that a total of 401 metabolites differed between Landrace and Meihua. Seventy of these 401 metabolites were conclusively identified. Landrace accumulated more short-chain fatty acids (SCFAs) and secondary bile acids in the colon lumen. Moreover, expression of the SCFAs transporter (solute carrier family 5 member 8, *SLC5A8*) and receptor (G protein-coupled receptor 41, *GPR41*) in the colon mucosa was higher, while the bile acids receptor (farnesoid X receptor, *FXR*) had lower expression in Landrace compared to Meihua. The relative abundances of 8 genera and 16 species of bacteria differed significantly between Landrace and Meihua, and were closely related to the colonic concentrations of bile acids or SCFAs based on Pearson's correlation analysis. Collectively, our results demonstrate for the first time that there were differences in the colonic microbiome and metabolome between Meihua and Landrace piglets, with the most profound disparity in production of SCFAs and secondary bile acids.

Keywords: microbiome, metabolome, pig breeds, colon, short chain fatty acids, bile acids

INTRODUCTION

The gastrointestinal tract is a multi-function organ that harbors a dynamic microbiota population that interacts with the nutritional, physiological, and immunological functions of the host (Brestoff and Artis, 2013). Gut microbiota are influenced by many factors such as genetics, environment, diet, diseases, and lifestyle (Ananthakrishnan, 2015). Previous research suggests that genetic information of gut microbes is transmissible through generations (Goodrich et al., 2016). Furthermore, host genes related to immunity and diet could select for particular species of bacteria and archaea across different individuals, and it is possible that this effect is inherited. Support for inheritance comes from research showing that monozygotic twins possess a much more similar gut

microbial community than dizygotic ones when they are raised in the same conditions (Ridaura et al., 2013).

Recent research has led to increasing recognition of the association between gut microbiota and metabolites and host physiology. Gut microbiota can influence nutrient digestion and absorption (Turnbaugh et al., 2006), lipid metabolism (Li F. et al., 2013), and hormone biosynthesis (Clarke et al., 2014) in their hosts through key functional metabolites (Clarke et al., 2014; Levy et al., 2016), which include short chain fatty acids (SCFAs), bile acids, indoles, vitamins, and polyamines (Yan et al., 2016). These metabolites can initiate various physiological and immunological responses once recognized and taken-up by host cells (Malmuthuge and Guan, 2016). For example, G protein-coupled receptors (GPR41, GPR43, and GPR109A) are activated by SCFAs to influence host physiology (Sivaprakasam et al., 2016). Microbiota-derived bile acids can modulate the metabolic activities of the host through activation of bile acid receptors such as farnesoid X receptor (FXR) and Takeda G protein-coupled receptor 5 (TGR5) (Fiorucci et al., 2009; Wahlstrom et al., 2016).

Recent studies have identified differences in the plasma and serum metabolomic profiles between two heavy pig breeds (Bovo et al., 2016), as well as differences in the colonic bacterial abundances and bacterial metabolites between fatty and lean pigs (Jiang et al., 2016). Moreover, evidence has indicated that the gut microbiomes shaped by host diet or host genotype, and can affect postnatal development of gut tissues and host metabolic health (Ha et al., 2014). Previous studies have shown the fecal microbial composition displayed diverse difference among different pig breeds (Pajarillo et al., 2014, 2015; Yang et al., 2014; Xiao et al., 2017). Microbial metabolites are major mediators linking host health, physiology, and pathology through regulation of many biological effects. However, differences in microbial communities and their metabolic activity in different pig breeds with different growth rates are largely unknown, especially in terms of how gut microbiota-derived metabolites relate to pig physiology.

Meihua is a Chinese fatty-type pig breed with a slow growth rate, while Landrace is a lean-type pig breed with a fast growth rate (Li Z. et al., 2013). In the present study, we therefore aimed to explore the differences in the colonic luminal metabolome and microbiome between Landrace and Meihua piglets by integrating taxonomic and metabolomic profiling analyses. Moreover, the relative gene expression of receptors and transporters of certain gut microbiota-derived metabolites (including SCFAs and bile acids) were determined. The results of this study may provide new insights in developing dietary intervention strategies to improve human health and animal production by manipulating host-microbiome-metabolites interactions.

MATERIALS AND METHODS

Animals and Sample Collection

A total of seven Landrace piglets (10.53 ± 0.52 kg initial BW) and seven Meihua piglets (3.71 ± 0.44 kg initial BW) weaned at 30 days of age were used in this study ($n = 7$). All piglets were reared under the same conditions and housed in pens with plastic slatted flooring in the research farm (Qujiang District, Shaoguan, China) of Agro-biological Gene Research Center, Guangdong Academy of Agricultural Sciences. Piglets were provided food and water *ad libitum*. Piglets were fed with the same commercial diets (Guangzhou Kingcard Biology Technology Co. Ltd., China), which contained 20.0% crude protein, 3,370 kcal of digestible energy, and 1.3% lysine. The experiment lasted 35 days. Piglets at 65 days of age were weighed again and the average daily weight gain of piglets within each breed was calculated. At the end of the experiment, all piglets were slaughtered for sample collections as previously described (Zhu et al., 2017). Fresh ileum and colon mucosa, and colon contents were collected and shock-frozen in liquid nitrogen, then stored at −80°C until analysis. All experimental procedures were carried out with the approval of the Animal Care and Use Committee in Guangdong Academy of Agricultural Sciences, China.

Untargeted Metabolomic Study Based on Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS)

Colon contents (100 mg) from each piglet ($n = 7$) were extracted with 1,000 μL extraction solution (methanol: acetonitrile: $ddH_2O = 2:2:1$) and 20 μL L-2-chlorophenylalanine (1 mg/mL stock in ddH_2O, as the internal standard), and ultrasound treated for 5 min with 25 KHz intensity (SB-5200D, NingBoScientz Biotechnology Co., Ltd.). The extracted mixture was centrifuged for 15 min (16,090 g, 4°C) after incubating for 1 h at −20°C. Then, 0.5 mL supernatant was transferred to a vacuum concentrator and dried for 30 min without heating. Finally, 100 μL acetonitrile-water solution (1:1, v/v) was used to reconstitute the dry extracts, and 60 μL supernatant was transferred into a glass vial for LC-MS/MS analysis.

Two microliters of supernatant from each sample was injected into the LC-MS/MS system with HPLC (1290, Agilent Technologies) tandem TripleTOF 6600 (Q-TOF, AB Sciex) MS. The metabolome was separated with a UPLC BEH Amide column (1.7 μm, 2.1 × 100 mm, Waters). The mobile phase consisted of 25 mM NH_4OAc and 25 mM NH_4OH in water (pH 9.75) (A) and acetonitrile (B) and was eluted with the following gradient: 0 min, 85% B; 2 min, 75% B; 9 min, 0% B; 14 min, 0% B; 15 min, 85% B; 20 min, 85% B, which was delivered at 0.3 mL/min. A mass spectrometer (Q-TOF, AB Sciex) was used to acquire MS/MS spectra on an information-dependent basis during an LC-MS/MS experiment. In this mode, the acquisition software (Analyst TF 1.7, AB Sciex) continuously evaluated the full scan survey MS data as it collected and triggered the acquisition of MS/MS spectra depending on preselected criteria. In each cycle, six precursor ions with intensity >100 were chosen for fragmentation at a collision energy of 35 V (15 MS/MS events with product ion accumulation time of 50 ms each). Electrospray ion (ESI) source conditions were set as following: ion source gas 1 as 60, ion source gas 2 as 60, curtain gas as 30, source temperature

550°C, ion spray voltage floating 5,500 or −4,500 V in positive or negative modes, respectively.

Metabolomic Data Processing and Multivariate Statistical Analysis

The MS raw data (.d) files were converted to mzXML format using ProteoWizard, and processed using the R (version 3.3.2) package XCMS. The preprocessing results generated a data matrix that consisted of the retention time (RT), mass to-charge ratio (m/z) values, and normalized peak intensity. 3,864 and 3,840 peaks were detected under LC-MS (ESI+) and LC-MS (ESI-) after pre-processing the detected signals, respectively. The R package CAMERA and an in-house database were used for peak annotation after XCMS data processing (Wang et al., 2016). Further, multivariate statistical analyses were conducted using the SIMCA software package (V14, Umetrics AB, Umea, Sweden) on the resulting three-dimensional data matrix.

After the data matrix was mean-centered and scaled to the pareto variance, principal components analysis (PCA) and orthogonal partial least squares-discriminant analysis (OPLS-DA) were carried out to generate an overview of the variables of colon contents and to visualize the differences between pig breeds, respectively (Wheelock and Wheelock, 2013). For model diagnosis, an appropriate OPLS-DA model must meet at the condition of *P*-value (CV-ANOVA) < 0.05. In addition, the parameters R^2Y and Q^2 are used to evaluate the predictive ability and fitting level of the model resulting from internal validation. After model diagnosis, metabolites for separating the models were selected with the following requirements: variable importance in the projection (VIP)>1 and $|p(corr)| \geq 0.5$ with 95% jack-knifed confidence intervals. The Student's *t*-test was applied to further analyze intergroup significance of the selected metabolites.

SCFA Quantification by Gas Chromatography-Mass Spectrometry (GC-MS)

The extraction and quantification of SCFAs in the colon contents were performed as previously described (Sun et al., 2017).

Bile Acid Quantification by LC-MS/MS

Colon content from piglets of two breeds ($n = 7$) was extracted as previously reported (Cai et al., 2012), with some modifications. Briefly, 100 mg of colon content for each sample was extracted with 1 mL of 10 mM ammonium acetate in 70% methanol followed by vortexing for 30 s twice before and after shaking for 20 min, then centrifuging at 13,000 g at 4°C for 10 min. Finally, 200 μL of extract solution was used for quantification by LC-MS/MS. The bile acid extract solution (10 μL) was injected into a C18 column (AQUITY UPLC BEH 130, 1.7 μm, 2.1 by 100 mm, Waters) at a flow rate of 0.1 mL/min, and the column temperature was maintained at 40°C. The solution was separated by reversed phase ultra-fast LC (Shimadzu, Kyoto) with a multi-step linear gradient elution using solution A (10 mM ammonium acetate–ammonium hydroxide buffer at pH 8.0) and solution B (10 mM ammonium acetate in acetonitrile–methanol solution, 3:1) over 30 min. The gradient profile was set as follows: 30–65 % B over 6 min; 65–72% B over 8 min; 72–90% B over 1 min; 90–90% B over 5 min; 90–30% B over 0.1 min. Then the column was equilibrated with 30% mobile phase B for 10 min. The eluate was then introduced into the ESI source of a tandem triple quadrupole MS analyzer (API4000, AB Sciex, Foster City, CA), and cholic acid (CA), deoxycholic acid (DCA), chenodexycholic acid (CDCA), and lithocholic acid (LCA) authentic compounds were quantified in multiple reaction monitoring (MRM) mode using optimized MS/MS conditions (Table S1). MS conditions were as follows: source, Turbo IonSpray; ion polarity, negative; IonSpray voltage, 4,500 V; source temperature, 550°C; gas, nitrogen; curtain gas, 25 psi; nebulizing gas (GS1), 55 psi; collision gas (GS2), 55 psi; scan type, MRM; Q1 resolution: unit; Q3 resolution: unit. Analyst 1.5.2 software (AB Sciex, Foster City, CA) was used to control the instrument and to acquire and process all MS data.

Gene Expression Study by Quantitative Real-Time PCR (qRT-PCR)

Total RNA was extracted from frozen colon and ileum mucosal tissues ($n = 3$) using TranzolUp reagent (TransGen Biotech, China) according to the manufacturer's instructions. The concentration and quality of extracted total RNA were determined by a NanoDrop-ND2000 spectrophotometer (Thermo Fisher Scientific Inc., Germany). The integrity of total RNA was further checked by gel electrophoresis on a 1% agarose gel for visualization of complete 28 and 18S bands. Genomic DNA was eliminated by treatment with DNase I (TransGen Biotech, China). Complementary DNA (cDNA) was then synthesized from 1 μg of total RNA using an M-MLV First Strand Kit (Invitrogen, USA) following the instructions of the manufacturer. The qRT-PCR for gene expression was performed in duplicate using a ChamQTMSYBR®qPCR Master Mix (Vazyme Biotech Co., China) on a CFX connect system (Bio-Rad, USA). Specificity of the amplification was confirmed by the melting curve. Primers were designed using Primer Premier 5.0 software (Applied Biosystems, USA) and were synthesized by Generay Biotech Co. (Shanghai, China). Primer sequences, annealing temperatures (Tm), and product lengths of target genes are listed in Table S2. The fold change of the target genes was normalized to housekeeping gene (β-actin) and was calculated using the $2^{-\Delta\Delta CT}$ method.

ATP Quantification by HPLC

Colon content (25 mg) from piglets of two breeds ($n = 7$) was extracted with 1 mL of 0.3 M $HClO_4$, and sonicated once for 5 min after vortex mixing, followed by centrifugation for 5 min (16,090 g, 4°C). The supernatant (160 μL) was then transferred to a new tube containing 2 M KOH solution and equilibrated at 4°C for 3 h. After centrifugation (5 min, 16,090 g, 4°C), the supernatant was passed through a membrane filter (0.22 μm), and 150 μL was transferred into a 2 mL glass vial for HPLC analysis. HPLC analysis was performed using methanol: 50 mM KH_2PO_4 buffer solution (9:91, v:v) (pH 6.5) as the mobile phase with a flow rate of 0.5 mL/min, and detected using a UV

detector at a wavelength of 259 nm. Finally, data processing was performed using Chromeleon version 6.8 (Thermo Fisher).

Bacterial DNA Extraction, PCR Amplification, High-Throughput Sequencing, and Bioinformatics Analysis

Microbial genomic DNA from piglets of two breeds ($n = 7$) was extracted from 200 mg of each colonic sample using QIAamp DNA stool minikit (Qiagen, Germany) according to the manufacturer's instructions. 2.5 μL diluted DNA sample (5 ng/μL) was used for 25 μL PCR reaction mixtures. 10 μL of primers (forward primer, 341F 5″-CCTACGGGAGGCAGCAG-3″; reverse primer 806R 5″-GGACTACHVGGGTWTCTAAT-3″ with attaching12 bp barcode sequences) at 1 μM concentration were used to amplify a portion of the V3–V4 region of bacterial 16S rRNA genes with 12.5 μL TaKaRa ExTaq polymerase mixtures. PCR amplicon products were purified using AMPure XP beads (Biomek, USA) and were checked for quality on an Agilent 2100 bioanalyzer (Agilent, USA). Amplicons were paired-end sequenced on the Illumina MiSeq platform using 2 × 250 bp MiSeq reagent kit v3 (Illumina, USA). Raw reads were submitted to the Sequence Read Archive of the NCBI (accession number SRP 095863).

In order to obtain high quality sequences, head or tail bases with qualities lower than Q30 were trimmed, and sequence lengths shorter than 100 bp were removed. Fastq-join (v1.3.1) (Aronesty, 2013) was used to combine paired-end reads. Assembled sequences were analyzed with QIIME software (v1.8.0) (Caporaso et al., 2010) to obtain operational taxonomic units (OTU) using the closed-reference OTU picking method with default OTU clustering tool UCLUST (uclust v1.2.22) such that each clustered OTU was at the 97% similarity level. Representative sequences of OTUs were selected based on the maximum length and were aligned to Greengenes 16S rRNA gene database (v13.8) with the RDPII classifier (v2.2) (Wang et al., 2007) to obtain taxonomic assignments. For species level identification, all OTUs were aligned to bacterial genome sequences in GenBank using the BLAST (Basic Local Alignment Search Tool) algorithm (blast v2.2.25) and parsed with the following criteria:(1) best hit; (2) cutoffs of 90% identity and 400 bp alignment length; (3) in accordance with the axonomic assignments from QIIME. Alpha-diversity indices (Chao1, Shannon, PD whole tree, and observed species) were calculated based on a subset of randomly selected sequences from each sample. Beta-diversity of weighted UniFrac-based PCoA (principal coordinate analysis) was calculated to show the group differences. Beta-diversity statistical analyses were tested using PERMANOVA based on Bray-Curtis dissimilarities and 999 permutations in R (v3.2.0). Microbial function was predicted using PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (v1.1.0) (Langille et al., 2013). Differences among groups were compared with the software STAMP (Parks et al., 2014). Two-side *t*-test and Benjamini-Hochberg FDR correction from STAMP were used in two-group analysis. Differences were considered significant at $P < 0.05$. The correlation coefficients between metabolic compounds and bacterial compositions at both genus and species level were calculated using Pearson's correlation test in R software (v3.2.2).

RESULTS

Metabolic Differences of Gut Microbiota between Landrace and Meihua Piglets

In the PCA score-plots (**Figures 1A,B**), the two pig breeds were distributed in separate groups based on the first two principal components, which indicate significant differences in the metabolites between them. The score-plot of OPLS-DA, a method of supervised pattern recognition, further revealed that the colonic metabolome of Meihua piglets could be clearly distinguished from that of Landrace piglets (**Figures 1C,D**). Parameters for evaluating the predictive ability and fitting level of the model resulting from internal validation suggested that the three models possessed a satisfactory fit with good predictive power. These parameters were: values of R^2Y and $Q^2 > 0.7$, and *P*-values of CV-ANOVA <0.05, which meant the groups within the models were reliable and significantly different (**Figures 1C,D**). Further, based on a VIP > 1 in 95% jack-knifed confidence intervals and $|P(\text{corr})| \geq 0.5$, a total of 222 and 179 biomarker metabolites, which were detected in the LC-MS/MS (ESI+) and LC-MS/MS (ESI-) respectively, were selected according to the OPLS-DA S-plot (**Figures 1E,F**).

In total, relative levels of 401 biomarker metabolites differed significantly between Landrace and Meihua piglets. Detailed information about the 401 biomarker metabolites is in Data S1. In-depth analyses using an in-house MS/MS database established by authentic standards allowed for conclusive identification of 70 biomarker metabolites (**Table 1**). Those 70 compounds included bile acids, free amino acids, dipeptides, lipids, nucleotides, organic acids, nitrogen-containing compounds, and so on. In Landrace piglets, the relative levels of five bile acids (deoxycholic acid, lithocholic acid, glycocholate, ursodeoxycholic acid, and taurolithocholic acid), four sphingolipids [Lyso PE (20:3n6/0:0), PE (20:3/0:0), phytosphingosine, and sphinganine], two fatty acids (stearic and 16-hydroxypalmitic acid), four benzene derivatives (hydroquinone, 3-methylphenyacetic acid, 1,4-dihydroxybenzene and 3-hydroxybenzoate), and 11 other compounds in the colon lumen were significantly higher than those in Meihua piglets ($P < 0.05$). However, the relative levels of six free amino acids (glutamine, glutamate, D-proline, L-proline, citrulline, and tyrosine), eight dipeptides (Phe-Thr, Lys-Leu, Pro-Val, IIe-Pro, His-IIe, Lys-Pro, and Phe-Val), five nucleotides (thymine, uracil, adenine, deoxyguanosine, and 2-hydroxyadenine), and 16 other compounds were significantly lower in Landrace piglets compared to Meihua piglets ($P < 0.05$).

The metabolome view map generated by the software Metaboanalyst 3.0 revealed that nine relevant pathways were significantly enriched for those 70 metabolites based on *P*-value (<0.05) or impact value (>0.1) (Figure S1). The impact values for those seven pathways, including arginine and proline metabolism, pyrimidine metabolism, alanine-aspartate-glutamate metabolism, tyrosine metabolism, D-Glutamine

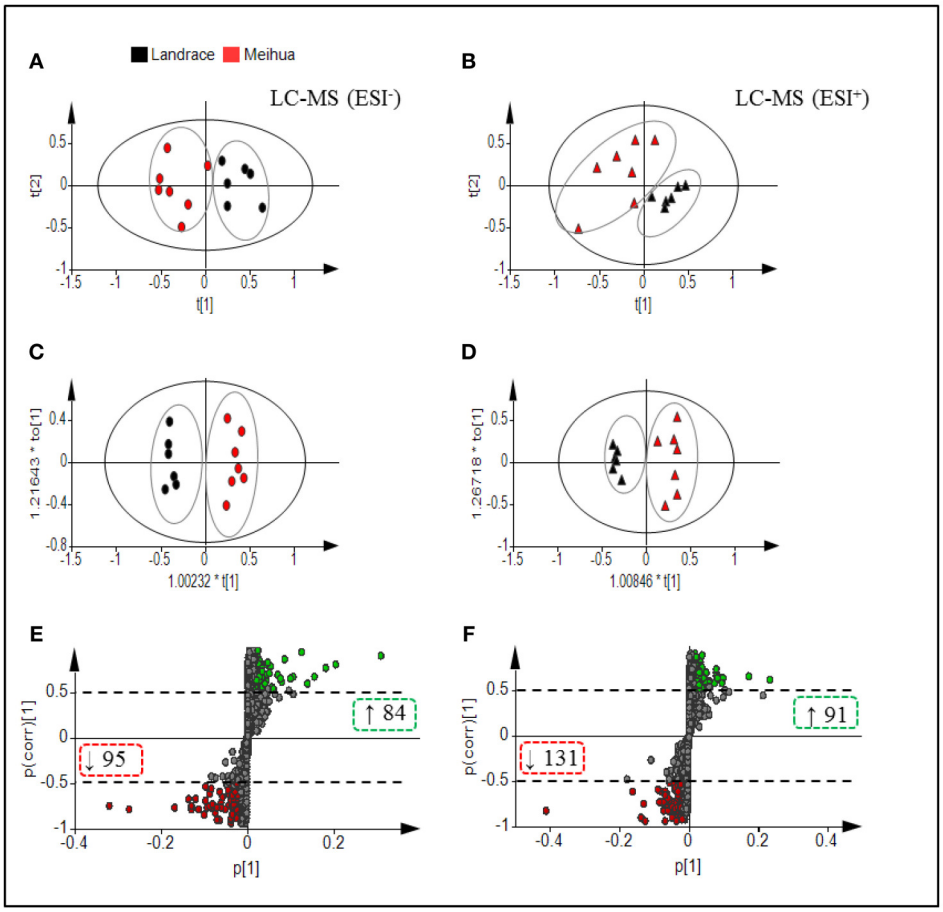


FIGURE 1 | Multivariate statistical analysis of untargeted metabolomic data obtained using the LC-MS/MS approach. PCA Score plot of colonic metabolomic data for Landrace (black) and Meihua (red) piglets obtained by **(A)** LC-MS (ESI−) and **(B)** LC-MS (ESI+). **(C)** OPLS-DA Score plot of colonic metabolomic data obtained by LC-MS (ESI−); $R^2Y = 0.976$, $Q^2 = 0.816$; and *P*(CV-ANOVA) = 0.0049. **(D)** OPLS-DA Score plot of colonic metabolomic data obtained by LC-MS (ESI+) data; $R^2Y = 0.963$, $Q^2 = 0.73$ and *P* (CV-ANOVA) = 0.021. **(E)** S-plot of LC-MS (ESI-) data with 3,480 metabolite signals detected. **(F)** S-plot of LC-MS (ESI+) data with 3,864 metabolite signals detected. Red circles in S-plots are model-separated metabolites following the conditions of VIP >1 and | *P* (corr)| ≥ 0.5 with 95% jack-knifed confidence intervals. Red or green rectangles in S-plots identify the numbers and tendency of metabolites to separate in the model when Meihua piglets are compared with Landrace piglets.

and D-glutamate metabolism, sphingolipid metabolism, and pantothenate-CoA biosynthesis, were 0.199, 0.164, 0.384, 0.152, 0.139, 0.140, and 0.180, respectively (Table S3).

High Levels of SCFAs and Secondary Bile Acids Accumulated in the Colon Lumen of Landrace Piglets

In order to validate differences in the gut metabolome between the two pig breeds, we used a targeted metabolomics approach to quantify the concentrations of some microbiota-derived metabolites (bile acids and SCFAs) in the colon lumen. Based on the GC-MS results, we found that levels of acetic acid, propionic acid, butyric acid, and valeric acid were much higher in the colon lumen of Landrace compared with Meihua ($P < 0.05$), but their levels of isovaleric acid did not differ (**Figure 2A**). The ratio of each SCFAs component was recorded as followed: 48.8 and 46.2% for acetic acid, 29.0 and 32.5% for propionic acid, 16.4 and 16.2% for butyric acid, 2.0 and 2.0% for isovaleric acid and, 3.8 and 3.1% for valeric acid in Landrace and Meihua, respectively (Figure S2B).

Based on the untargeted metabolomics approach, metabolites from bile acid metabolism were different between these two pig breeds (**Table 1**). Due to these differences, we further quantified the primary and secondary bile acids by LC-MS/MS. The primary bile acid (CA), which was derived from the conjugated endogenous bile acids by specific gut bacteria, was much higher in Meihua than Landrace ($P < 0.05$), while the secondary bile acids (DCA and LCA) derived from the primary bile acids were significantly lower in Meihua compared with Landrace, particularly for DCA ($P < 0.05$) (**Figure 2B**). As expected, the average daily weight gain of Landrace piglets during days 30–65 was higher than that of Meihua piglets ($P < 0.05$) (**Figure 2D**).

Higher Expression of SCFA Transporters and Receptor Genes but Lower Expression of Bile Acid Receptor Genes Found in the Colon Mucosa of Landrace Piglets

To further explore whether SCFA levels in the colon lumen were associated with the expression of transporter and receptor genes of SCFAs in piglets, qRT-PCR was performed on the colon mucosa tissues of the piglets. As expected, relative mRNA expression of three receptor genes activated by SCFAs (*GPR41*, *GPR43,* and *GPR109A*) (Kasubuchi et al., 2015), were higher in colon tissues of Landrace (**Figure 3A**). In addition, one of the transporters of SCFAs, *SLC5A8*, was also highly expressed in Landrace when compared to Meihua piglets (**Figure 3A**). The result indicated that the high levels of SCFAs in the colon lumen of Landrace might be positively correlated with the expression of *GPR41*and *SLC5A8* in colon mucosa.

For bile acids in the colon lumen, Landrace piglets had higher levels of the secondary bile acids, LCA, and DCA, when compared to Meihua piglets. Accordingly, we found that the expression levels of *FXR*, but not *TGR5*, were up-regulated in colon and ileum mucosa of Meihua, the opposite pattern of the levels of secondary bile acids in colon content (**Figures 3B,C**).

Differences in Colonic Luminal Microbiome between Landrace and Meihua Piglets

To further assess whether differences in gut microbiota are the causal factor for the differences in colonic luminal metabolomes between Landrace and Meihua piglets, high-throughput sequencing was performed to analyze 16S rRNA of bacteria from both pig breeds. A total of 792,455 quality-filtered sequences were obtained with an average of 56,604 sequences per sample in colonic microbiota. Then, four alpha diversity indices including observed species, Chao1, PD whole tree, and Shannon, were estimated. Community diversity index (PD whole tree) and community richness index (Chao1) of colonic microbes in Landrace piglets were significantly higher than those in Meihua piglets ($P < 0.05$, Table S4). The PCoA indicated that groups were not distinctly clustered separately in distribution of microbiota at the colonic contents (Figure S3). The relationships of gut microbiota between groups were calculated using permutational multivariate analysis of variance (PERMANOVA) based on Bray–Curtis distance. Results showed that community structures in the colon were not significantly different between Landrace and Meihua piglets at the genus level ($P > 0.05$, Table S4).

Firmicutes and Bacteroidetes were the most predominant phyla in the colons of both Landrace and Meihua piglets, comprising more than 88% of the total sequences. There was no significant difference in bacterial phyla in the colonic contents between two breeds. At the genus level, the most abundant genus in both pig breeds was *Prevotella*, followed by the genera *Lactobacillus*, *Bacteroides*, and *Streptococcus*. The proportion of *Streptococcus*, *CF231*, *Bulleidia*, and *Chlamydia* in Landrace were 2.6, 1.8, 9.2, and 6.6-fold higher than those in Meihua, respectively ($P < 0.05$) (**Figure 4A**). The proportion of *Bacteroides*, *YRC22*, and *Holdemania* in Landrace were 0.8, 0.9, and 0.7-fold lower than those in Meihua, respectively ($P < 0.05$) (**Figure 4A**). The genus *Neisseria* was found in Landrace piglets, but was absent from Meihua piglets.

We further assessed differences in the bacterial community at the species level using the BLAST algorithm to align the assembled 16s rRNA sequences to the bacterial genome and parsed the alignment results using the filtered setting (see section Materials and Methods). A combined total of 217 bacterial species were identified in the colons of two pig breeds. *Prevotella dentalis* and *Prevotella melaninogenica* were the most common species and both accounted for more than 10% of all colonic microbes in Landrace and Meihua piglets. Significant differences in four *Clostridium* species were observed between the two breeds. The relative abundances of *Clostridium saccharoperbutylacetonicum*, *Clostridium* sp., *Clostridium ljungdahlii*, and *Clostridium sticklandii* were higher in Landrace piglets than Meihua piglets ($P < 0.05$) (**Figure 4B**). The relative proportions of *Streptococcus pasteurianus*, *Streptococcus lutetiensis*, *Streptococcus suis*, *Lactobacillus helveticus*, and *Chlamydia trachomatis* were also higher in Landrace piglets than Meihua piglets ($P < 0.05$) (**Figure 4B**). However, *Bacteroides thetaiotaomicron*, *Bacteroides fragilis*, *Lactobacillus sanfranciscensis*, *Enterococcus faecalis*, and *Staphylococcus saprophyticus* were lower in Landrace piglets than Meihua piglets ($P < 0.05$) (**Figure 4B**). The species *Neisseria meningitides* exists in Landrace but was not present in Meihua piglets.

The PICRUSt analysis was used for predicting the potential functions of the intestinal microbiota. Based on the level 2 of KEGG Pathway analysis, it showed that the colon luminal microbial carbohydrate metabolism pathway in Meihua piglets was more abundant than in Landrace piglets (Figure S4A). Furthermore, pathway enrichments at KEGG level 3 also showed that Meihua piglets had higher enrichment of the pathways involved in carbohydrate metabolism, including (1) galactose, fructose, and mannose metabolism; (2) other glycan degradation; and (3) starch and sucrose metabolism (Figure S4B). Less obviously but significantly, the abundance of pathways for biosynthesis of primary and secondary bile acids in Meihua piglets were higher than in Landrace piglets (Figure S4B). In addition, more enrichment was detected for sphingolipid metabolism and the insulin signaling pathway in Meihua piglets (Figure S4B). However, colonic microbiota in Landrace had higher enrichment of pathways involved in branched chain amino acid degradation, butanoate metabolism, as well as in flagellar assembly, secretion system, bacterial motility proteins, and bacterial chemotaxis processes (Figure S4B).

Finally, we compared the abundance of bile salt hydrolase (KO1442) in the colonic microbiota between the two pig breeds. Bile salt hydrolase is the key enzyme involved in metabolism of primary bile acids. We observed that the abundance of bile salt hydrolase (KO1442) from colonic microbiota in Landrace piglets was significantly lower compared to Meihua piglets ($P < 0.05$) (**Figure 2C**). This was consistent with the higher level of the unconjugated primary bile acid CA in Meihua piglets.

TABLE 1 | Identified metabolites for discriminating between Landrace and Meihua piglets based on the untargeted metabolomics study.

Metabolic pathway	RT[a] (min)	Detection mode	VIP	p (corr)	Mean-Landrace	Mean-Meihua	*t*-test[b]
BILE ACIDS METABOLISM							
Deoxycholic acid	6.522	LC-MS ESI-	9.6825	0.76131	1.56E – 02	5.55E – 03	**
Lithocholic acid	5.988	LC-MS ESI+	2.05897	0.61714	4.42E – 04	6.17E – 05	*
Taurolithocholic acid	7.772	LC-MS ESI+	1.05204	0.81984	0.000142	0.0000724	***
Glycocholate	4.721	LC-MS ESI-	1.39793	0.803073	0.00029897	0.00010597	**
Ursodeoxycholic acid	4.71187	LC-MS ESI+	1.75393	0.812378	0.000297555	0.0000774	**
DIPEPTIDES							
Phe-Thr	3.852	LC-MS ESI+	1.14398	0.69232	0.0000206	0.000125	*
Lys-Leu	4.337	LC-MS ESI+	1.35231	0.66845	0.000128	0.000287	***
Pro-Val	3.586	LC-MS ESI+	1.27253	0.61089	0.0000627	0.000205	*
Ile-Pro	3.82	LC-MS ESI+	3.29598	0.68484	0.000668	0.00149	**
His-Pro	3.919	LC-MS ESI+	1.49293	0.60608	0.00013	0.000337	*
Phe-Pro	4.386	LC-MS ESI+	1.87255	0.54983	0.000369	0.00072	
His-Ile	8.80625	LC-MS ESI+	1.35118	0.73756	8.55865E−05	0.00020965	**
Lys-Pro	14.627	LC-MS ESI+	2.1594	0.799666	0.000314194	0.00062943	**
Phe-Val	8.19602	LC-MS ESI+	1.00549	0.696012	0.00001605	0.0000976	*
ORGANIC ACIDS METABOLISM							
2-Butenoic acid	1.684	LC-MS ESI+	3.80601	0.57964	0.00147	0.00286	
Dodecanoic acid	3.32	LC-MS ESI-	1.07499	0.66646	0.00025	0.000392	*
Alpha-Linolenic acid	1.14592	LC-MS ESI-	7.29765	0.934984	0.00935933	0.0139623	***
16-Hydroxypalmitic acid	4.521	LC-MS ESI-	6.0069	0.85129	0.00384	0.000565	***
Stearic acid	1.38912	LC-MS ESI+	1.65239	0.744532	0.000699651	0.00049507	**
4-Guanidinobutyric acid	10.7132	LC-MS ESI+	2.07103	0.593309	0.000571072	0.00101013	
3-Hydroxybutyric acid	4.118	LC-MS ESI+	1.28956	0.59916	0.000142	0.000296	*
Levulinic acid	4.169	LC-MS ESI+	2.5701	0.6198	0.000219	0.000835	*
Succinate	1.95	LC-MS ESI-	1.10273	0.55747	0.000175	0.00067	*
SPHINGOLIPID METABOLISM							
LysoPE (20:3n6/0:0)	8.405	LC-MS ESI+	1.00886	0.71241	0.000117	0.000046	**
PE (20:3/0:0)	8.039	LC-MS ESI+	3.80258	0.74022	0.0017	0.000719	**
Phytosphingosine	5.005	LC-MS ESI+	1.59896	0.79725	0.000224	0.000061	***
Sphinganine	5.038	LC-MS ESI+	2.26242	0.74951	0.000463	0.000108	**
ARGININE AND PROLINE METABOLISM							
Glutamate	2.434	LC-MS ESI-	1.01355	0.51585	0.000458	0.000603	*
Glutamine	2.451	LC-MS ESI+	1.04735	0.61003	0.000143	0.00023	*
5-Aminopentanoic acid	1.968	LC-MS ESI+	5.46205	0.62789	0.00168	0.00427	*
D-Proline	9.50385	LC-MS ESI+	2.95699	0.593271	0.000792095	0.00157409	*
L-Proline	9.5367	LC-MS ESI+	1.89764	0.515471	0.000137675	0.00050401	
N-Acetylputrescine	2.185	LC-MS ESI+	2.55803	0.60163	0.000572	0.00124	
Citrulline	2.335	LC-MS ESI+	1.20243	0.66665	0.000295	0.000184	**
N-Acetylputrescine	10.1657	LC-MS ESI+	2.55803	0.601628	0.000572109	0.00123815	
N-Acety-L-glutamate	3.134	LC-MS ESI-	1.65456	0.70721	0.00009	0.000502	
TYROSINE METABOLISM							
Tyrosine	3.035	LC-MS ESI+	1.40856	0.75924	0.000173	0.000313	**
Tyramine	2.001	LC-MS ESI+	4.91784	0.72541	0.00143	0.00322	**
4-Methoxyphenylacetic acid	5.519	LC-MS ESI-	1.14392	0.53676	0.000119	0.000328	
Hydroquinone	1.03603	LC-MS ESI-	1.99776	0.610798	0.000521502	0.0000483	*
3-Methylphenylacetic acid	2.484	LC-MS ESI-	8.00808	0.68299	0.00928	0.0018	*
1,4-Dihydroxybenzene	1.817	LC-MS ESI-	1.08224	0.65869	0.000224	0.0000857	*
3-Hydroxybenzoate	3.367	LC-MS ESI+	1.67021	0.67459	0.00121	0.00097	*
Dopamine	2.268	LC-MS ESI+	1.32803	0.74421	0.000145	0.000272	**

(Continued)

TABLE 1 | Continued

Metabolic pathway	RT[a] (min)	Detection mode	VIP	p (corr)	Mean-Landrace	Mean-Meihua	*t*-test[b]
PYRIMIDINE METABOLISM							
Cytosine	1.868	LC-MS ESI+	2.17035	0.81631	0.000729	0.000396	**
Cytidine	4.068	LC-MS ESI+	1.44061	0.79162	0.000326	0.000176	**
Thymine	2.084	LC-MS ESI-	1.57467	0.71573	0.000215	0.000502	*
Uracil	1.85	LC-MS ESI-	4.91099	0.86249	0.00181	0.00408	***
Adenine	2.234	LC-MS ESI-	4.31482	0.66043	0.000225	0.00261	*
Deoxyguanosine	4.435	LC-MS ESI-	1.10168	0.53988	0.000123	0.00029	*
2-Hydroxyadenine	2.534	LC-MS ESI+	3.75296	0.53183	0.00105	0.00224	*
OTHERS							
O-Acetylcholine	11.3929	LC-MS ESI+	5.50725	0.597001	0.00689022	0.00420787	*
D-Lyxose	4.43453	LC-MS ESI-	1.55	0.632079	0.000296192	0.00061133	*
Glutaraldehyde	1.384	LC-MS ESI+	1.26127	0.60349	0.000153	0.000299	*
Glycyrrhetinic acid	1.28742	LC-MS ESI-	1.96603	0.639127	0.000175096	0.00067003	*
L-Carnitine	2.702	LC-MS ESI+	4.59919	0.64673	0.00116	0.00309	*
Ketoisocaproic acid	2.151	LC-MS ESI-	4.03199	0.69535	0.000726	0.0026	*
2-Oxoadipic acid	2.667	LC-MS ESI-	1.01241	0.8472	0.000114	0.000214	***
Beta-Glycyrrhetinic acid	7.822	LC-MS ESI-	1.96603	0.63913	0.000175	0.00067	*
2-Indolecarboxylic acid	7.71052	LC-MS ESI+	2.03021	0.545726	0.000760285	0.00039665	*
Norvaline	1.968	LC-MS ESI+	2.54301	0.57594	0.000856	0.000268	*
Oleanolic acid	7.589	LC-MS ESI-	1.64358	0.68042	0.0156	0.00555	**
p-Chlorophenylalanine	3.334	LC-MS ESI+	3.43195	0.69057	0.00429	0.00329	*
Betaine aldehyde	2.702	LC-MS ESI+	2.65383	0.5971	0.00105	0.000471	*
Picolinic acid	2.067	LC-MS ESI+	1.56513	0.74737	0.000282	0.000105	**
Pyrrolidine	1.201	LC-MS ESI+	3.27502	0.71124	0.0011	0.000269	**
Erucamide	5.64	LC-MS ESI+	1.50832	0.65965	0.000702	0.000507	*
Nicotinamide	2.051	LC-MS ESI+	3.28287	0.52826	0.00127	0.000147	
3-Mercapto-2-butanone	8.95645	LC-MS ESI+	1.43044	0.761824	0.000148586	0.00029097	**
1H-Purin-6-amine, N,N-dimethyl-Aminobutyrate	1.92688	LC-MS ESI+	1.61983	0.588964	0.0000449	0.00026979	*

*RT[a], Retention time. t-test[b], Student t-test of the relative levels of compounds detected in the colon lumen of Landrace piglets and Meihua piglets (n = 7). *p < 0.05, **p < 0.01 and ***p < 0.001, respectively. Asterisks in red were those metabolites with higher relative levels in Landrace piglets compared with Meihua piglets, otherwise the levels were lower in Landrace piglets (n = 7).*

Correlation between Microbial Communities and their Metabolites

Correlations between metabolites and 8 genera (**Figure 5**) or 16 species (Figure S5) of bacteria with significant differences between Landrace and Meihua piglets were obtained via Pearson's correlation analysis. As shown in **Figure 5**, the relative higher abundances of *Streptococcus*, *CF231*, *Bulleidia*, *Chlamydia*, and *Neisseria* were positively associated with higher concentrations of microbial metabolites in Landrace, including bile acids, SCFAs, and lipids ($P < 0.05$). The relative lower abundances of *Bacterioides*, *Holdemania*, and *YRC22* were negatively associated with the higher concentrations of bile acids, SCFAs, and lipids in Landrace, when compared to Meihua piglets ($P < 0.05$).

Furthermore, the relative higher abundances of 16 bacterial species in Landrace (Figure S5), including *C. saccharoperbutylacetonicum*, *Clostridium* sp., *C. ljungdahlii*, *C. sticklandii*, *S. pasteurianus*, *S. lutetiensis*, *S. suis*, *Lactobacillus delbrueckii*, *L. helveticus*, and *C. trachomatis* were positively correlated with the higher concentrations of bile acids, SCFAs, and lipids ($P < 0.05$). The relative lower abundances of *N. meningitides*, *B. thetaiotaomicron*, *B. fragilis*, *S. saprophyticus, E. faecalis*, and *L. sanfranciscensis* were negatively correlated with lower concentrations of metabolites such as amino acids, dipeptides, purines, and pyrimidines, when compared to Meihua piglets ($P < 0.05$) (Figure S5).

DISCUSSION AND CONCLUSION

The host gut–microbial relationship is of great importance to host phenotype, physiology, and health status (Lalles, 2016). However, whether or not the development of piglets with different genetic backgrounds will be influenced by differences in the intestinal microbiome and its metabolites requires further investigations. In-depth analyses of gut microbiome and metabolic activities using "omics" approaches will help identify

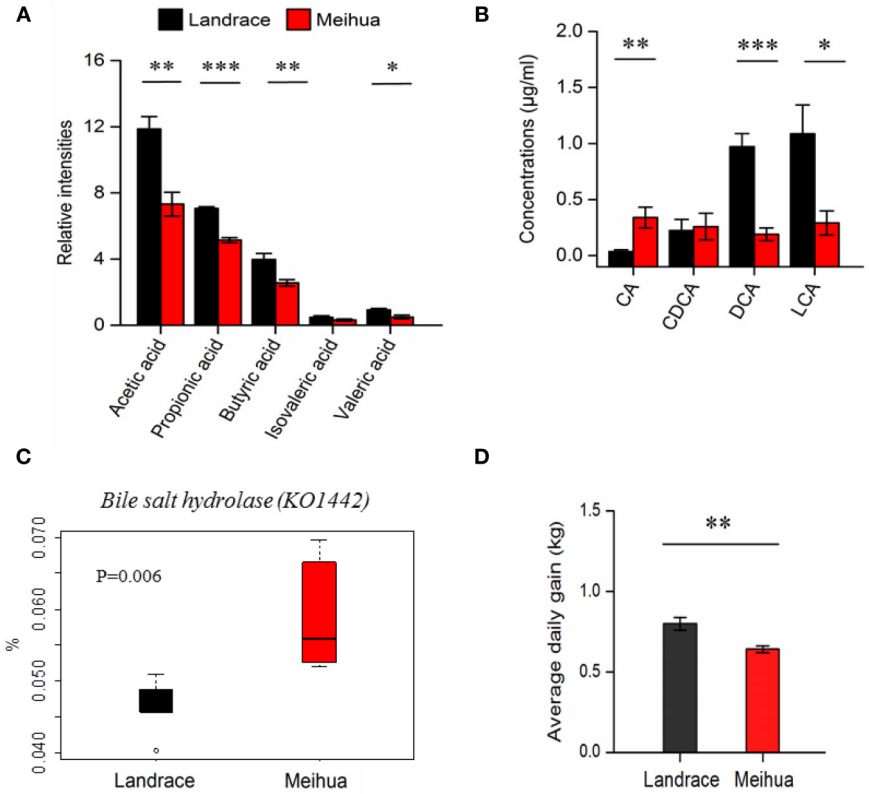


FIGURE 2 | Quantification of short chain fatty acids (SCFAs) and bile acids in the colon lumen using the targeted metabolomics approach and average daily weight gain of piglets. **(A)** The relative levels of SCFAs in the colonic contents samples from Landrace and Meihua piglets. **(B)** The absolute levels of bile acids in the colonic contents samples from Landrace and Meihua piglets. CA, cholic acid; DCA, deoxycholic acid; CDCA, chenodexycholic acid; LCA, lithocholic acid. Error bar indicated the mean value ± SEM of each experimental group. **(C)** The relative abundance of bile salt hydrolase (K01442) from the colonic microbiota in Landrace and Meihua piglets. Based on the PICRUSt prediction and KEGG analysis, the KO term named K01224 was enriched in the pathways of primary bile acid biosynthesis and secondary bile acid biosynthesis. **(D)** Average daily weight gain of Landrace and Meihua piglets between days 30 and 65 of age. Results are shown as mean ± SE. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$, respectively.

microbial biomarkers that facilitate the animal production and phenotype identification. Here, we investigated the composition of the microbiome and its metabolites in the colon contents of two pig breeds, using a combination of 16S rRNA gene high-throughput sequencing and MS-based metabolomics techniques. We found that the microbiome and metabolome in the colon lumen were significantly different between Meihua and Landrace piglets, breeds with varied growth rates (**Figure 2D**). The relative abundances of 8 genera and 16 species of bacteria differed significantly between Landrace and Meihua piglets. A total of 401 metabolites in the colon lumen differed significantly in relative levels between the two pig breeds. Seventy of these metabolites were definitively identified, including: bile acids, dipeptides, sphingolipids, amino acids, nucleotides, and many hydrophilic molecules. Differences between concentrations of bile acids and SCFAs in the colon lumen were further validated through targeted metabolomics approaches as well as gene expression of their transporters and receptors determined by qPCR.

The SCFAs, one of the most abundant microbial metabolites, are mainly produced by colonic bacteria through fermentation of carbohydrates, but a small portion (∼5%) are produced from protein or amino acids that are unabsorbed or undigested in the small intestine. Evidence has shown that the production of intestinal SCFAs can be influenced by dietary factors and food intake patterns (Rios-Covian et al., 2016). Whether the genetic variance of pigs or exposure to the pre-weaning diet led to differences in colonic SCFAs levels needs to be investigated. Our results have shown that the concentrations of most SCFAs, including acetic acid, propionic acid, butyric acid, and valeric acid, were higher in Landrace than in Meihua piglets. Our results were similar to previous reports that Bama mini-pig (a fatty-type Chinese local pig strain) had lower concentrations of total SCFAs in colon content than Landrace (Jiang et al., 2016). Previous study has also demonstrated that Chinese native Lantang pigs produced more SCFAs in the large intestine than Duroc pigs (Cheng et al., 2017). It is well known that acetate, propionate, and butyrate account for more than 90% of the total SCFAs in the colon (Rios-Covian et al., 2016). Similarly, our results indicate that the proportions of these three SCFAs in the colon of Landrace and Meihua piglets reached 94.2 and 94.9%, respectively (Figure S2B).

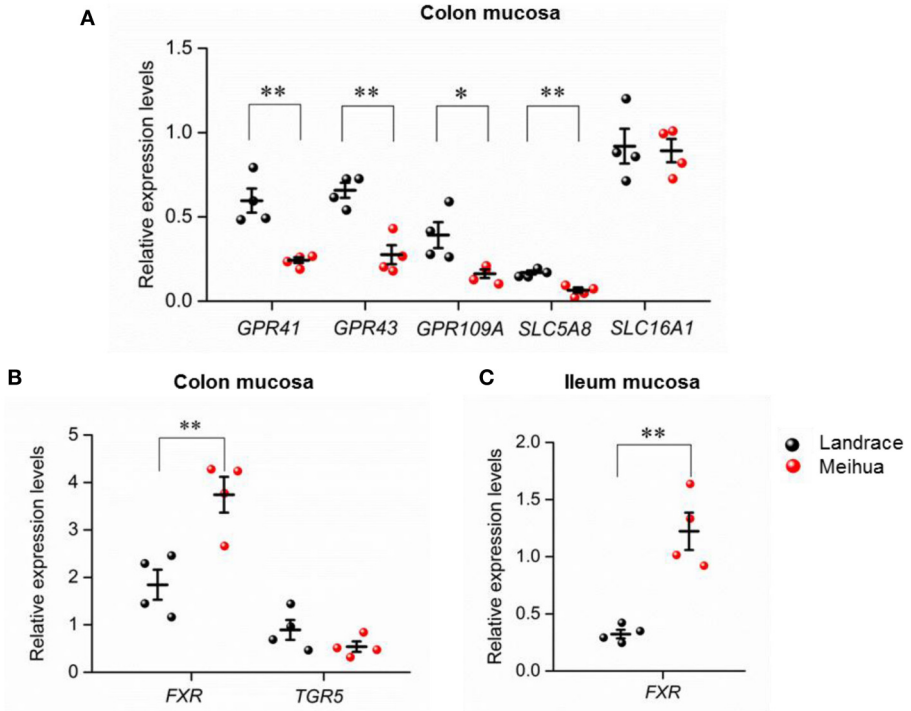


FIGURE 3 | Expression levels of genes related to SCFAs and bile acids in different tissues of Landrace and Meihua piglets. **(A)** Comparison of expression levels of SCFA-receptor genes in colon mucosa between Landrace and Meihua piglets. GPR41, GPR43, and GPR109A are receptors of SCFAs, while SLC5A8 and SLC16A1 are transporters of SCFAs. Comparison of expression levels of bile acid-related genes in colon mucosa **(B)** and ileum mucosa tissues **(C)** between Landrace and Meihua piglets. FXR and TGR5 are both receptors of bile acids. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$, respectively. Results are shown as mean ± SE.

The higher levels of SCFAs in Landrace were also supported by an increased abundance of SCFAs-producing bacteria. Even though we did not observe differences between bacterial phyla from the colonic lumen of Landrace and Meihua piglets, we found that *Clostridium* species like *C. saccharoperbutylacetonicum*, *Clostridium* sp., *C. ljungdahlii*, and *C. sticklandii* were all more abundant in colonic microbiota of Landrace piglets. Notably, differences in genera and species of bacteria are believed to contribute to specific metabolic functions (Bauer et al., 2016). Bacterial genera such as the *Clostridium* clusters IV and XIVa of Firmicutes, including species of *Eubacterium*, *Roseburia*, *Faecalibacterium*, and *Coprococcus*, are involved in SCFAs production (Nicholson et al., 2012; Van den Abbeele et al., 2013), and might thus impact swine health and development (Park et al., 2014).

Using KEGG pathway analysis, we found that carbohydrate metabolism by microbes was enriched in the colon tract of Meihua piglets, and intermediate metabolites involved in carbohydrate metabolism were identified by metabolomic analysis. We also found that ATP levels in the colon content of Meihua piglets were significantly higher than in Landrace piglets (Figure S2A), which indicates that the ability of gut microbiota to metabolize carbohydrates may not indicate high efficiency in production of SCFAs.

Microbial-derived SCFAs are almost totally absorbed by colonocytes, either through diffusion or through transport by monocarboxylate transporters like SLC16A1 and SLC5A8 (Ganapathy et al., 2013), with a small portion (<10%) excreted in the feces (Boets et al., 2015). The transporter SLC5A8, which can transport a variety of SCFAs, mediates beneficial effects of SCFAs especially when the concentration of SCFAs is low in the colon lumen (Ganapathy et al., 2013). In our study, the expression of *SLC5A8*, but not *SLC16A1*, in the colon mucosa was found to be higher in Landrace than Meihua piglets, which is consistent with the higher concentration of SCFAs mentioned above. Our results were also in accordance with a previous study demonstrating lower levels of mRNA and proteins of SLC5A8 in germ-free mice in contrast to wild-type mice (Cresci et al., 2010). Importantly, the result of *SLC5A8* expression was in agreement with the higher richness and diversity of colon luminal microbiota in Landrace compared to Meihua piglets, as shown by 16S sequencing.

SCFAs are not only the critical energy sources for colonocytes (Xiong et al., 2004), but also serve as key regulators of the intestinal epithelial barrier and gut immunity (Rios-Covian et al., 2016). These functions of SCFAs may facilitate better growth performance in Landrace compared to Meihua piglets. In addition, the biological functions of SCFAs depend on their specific receptors (GPR41, GPR43, and GPR109A) to affect host physiological processes, including regulation of energy metabolism in mammals (Kasubuchi et al., 2015; Hu et al., 2016; Rios-Covian et al., 2016). In our study, Landrace had higher expression of *GPR41* in the colon mucosa compared to Meihua

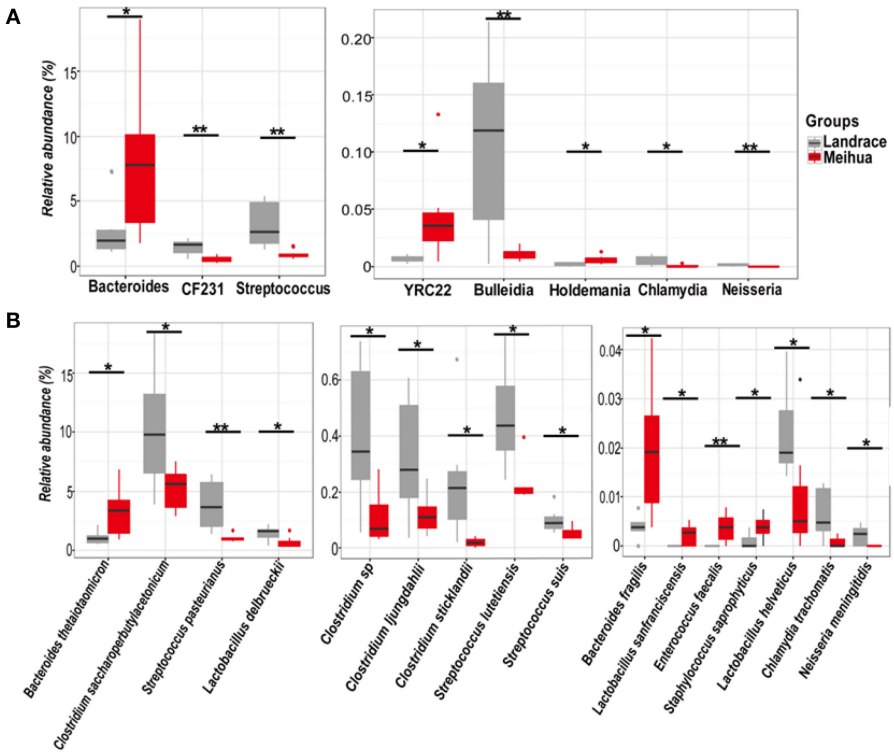


FIGURE 4 | Relative abundances of bacterial genera **(A)** and species **(B)** in the colon differed significantly between Landrace (gray) and Meihua (red) piglets. *$P <$ 0.05, and **$P < 0.01$.

piglets. This result was in accordance with the higher levels of lipid and lipid-like molecules, and lower concentrations of amino acids and dipeptides detected in the colon contents of Landrace by untargeted metabolomics analyses. However, another study has shown that the fatty-type Bama mini-pig expressed higher mRNA levels of *GPR41* and *GPR43* in colonic tissue compared to lean-type Landrace (Jiang et al., 2016). The discrepancy between the previous results and ours may involve differences in fatty-type pig breed used, experimental design, dietary composition, and sampling sites.

Apart from the differences in SCFAs production and expression of *SLC5A8* and *GPR41* between Landrace and Meihua piglets, another important finding of our study was the significant difference of secondary bile acids between the two pig breeds. During metabolism of bile acids, taurine- or glycine-conjugated bile acids were found to escape from the distal ileum when reabsorbed to enterohepatic circulation. However, before these bile acids escape, they are deconjugated by bile salt hydrolase secreted by gut microbiota including *Lactobacillus*, *Bifidobacteria*, *Enterobacter*, *Bacteroides,* and *Clostridium* (Wahlstrom et al., 2016). Primary bile acids are then oxidized or dehydroxylated by other microbiota-produced enzymes like hydroxy-steroid dehydrogenases (Ridlon et al., 2006) from bacteria involved in secondary bile acid fermentation including *Clostridium* (clusters XIVa and XI) and *Eubacterium* (Wahlstrom et al., 2016). Here, we showed that the abundance of bacterial genes (bile salt hydrolase, KO1442) related to secondary bile acid biosynthesis was higher in the colonic lumen of Meihua piglets, while the absolute quantification of secondary bile acids (DCA and LCA) in the colon contents was lower. Our results were in accordance with the previous results that a higher amount of bile acids was associated with lower butyrate concentrations in the rat cecum (Islam et al., 2011), probably due to inhibition of the proliferation of butyrate-producing bacteria or of the butyrate synthesis metabolic pathways (Ha et al., 2014).

The mRNA expression of bile acid receptor *FXR*, which is activated by unconjugated primary and secondary bile acids and inhibited by conjugated bile acids (Wahlstrom et al., 2016), was significantly higher in the colon and ileum of Meihua piglets than those of Landrace. Previous studies also demonstrated that the inhibition of FXR by gut microbiota was tightly linked to decreased hepatic lipid synthesis (Zhang et al., 2016) and alleviated obesity phenotypes (Li F. et al., 2013). Compared to wild type mice, FXR-deficient mice have a higher relative ratio of *Firmicutes* to *Bacteroidetes* in the intestine, as well as reduced obesity (Parseus et al., 2017). In addition, bile acids in the gut lumen can also directly regulate microbiotic functioning by serving as an antibiotic for bile sensitive bacteria or as a promoter for bile acid-metabolizing bacterial communities (Hofmann and Eckmann, 2006; Wahlstrom et al., 2016). In the present study, a higher abundance of *Bacteroides* spp., which are advantageous to

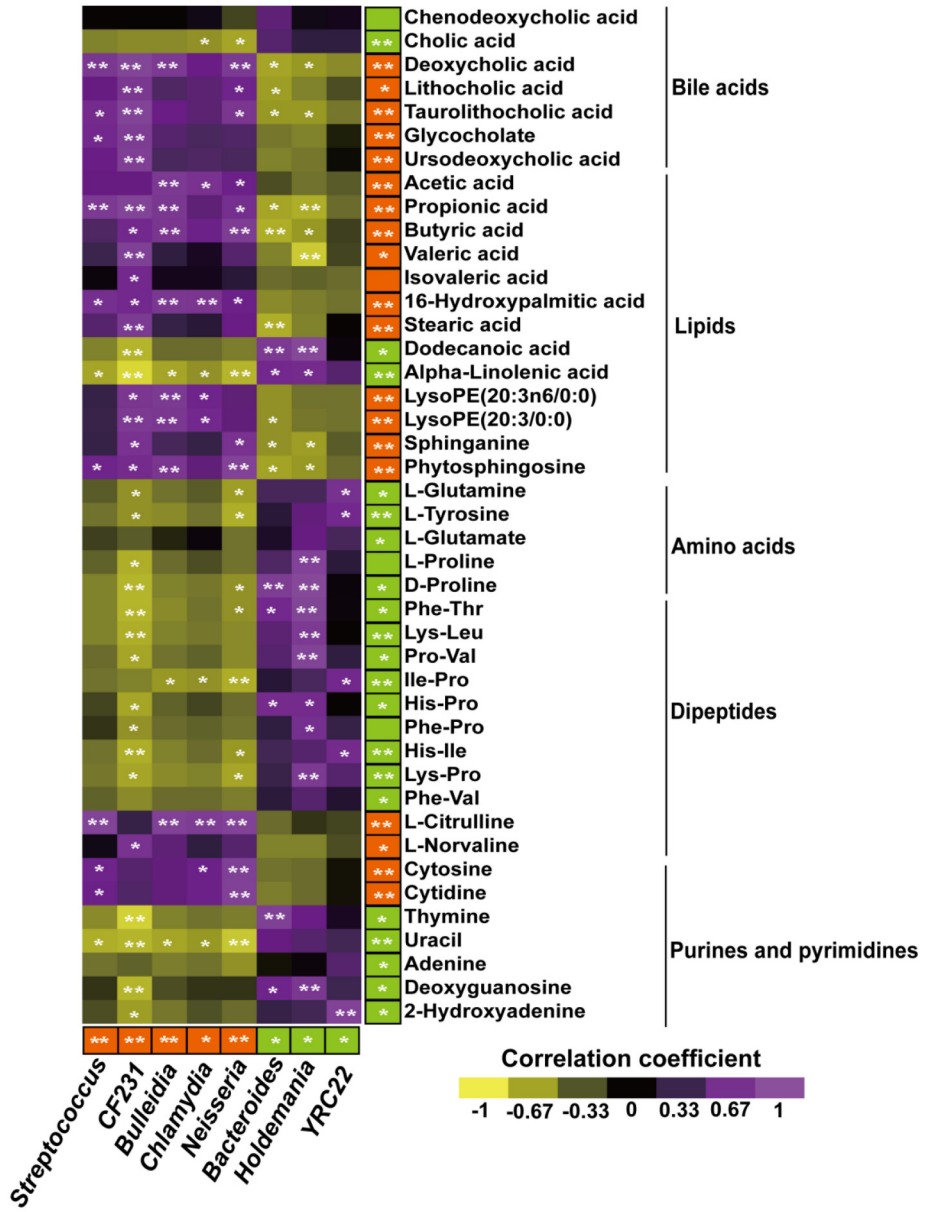


FIGURE 5 | Pearson's correlation analysis of colonic metabolites and colonic bacterial genus-level taxa in Landrace and Meihua samples. Bacterial genera and colonic metabolites enriched in Landrace or Meihua samples are colored red and green, respectively. An asterisk in the colored box indicates that a genera or metabolite is significantly enriched in Landrace (red) or Meihua (green) samples. Correlations with $P \leq 0.05$ are shown. Yellow represents a significant negative correlation ($P < 0.05$), purple represents a significant positive correlation ($P < 0.05$), and black represents no significant correlation ($P > 0.05$). Orange color with *represents a higher value in Landrace ($P < 0.05$), and **$P < 0.01$.

bile acid biosynthesis (Wahlstrom et al., 2016), were found in the colon lumen of Meihua compared to Landrace piglets.

By performing Pearson's correlation analyses, the relative abundances of bacteria at either the genus or species levels were found to be closely associated with the concentrations of specific microbial metabolites in the colonic lumen. For example, Firmicutes species like *Clostridium* sp., *L. delbrueckii*, *L. helveticcus*, and *Streptococcus lutetiensis*, were positively correlated with secondary bile acids as well as SCFAs. The diversity of gut microbiota species and the abundance of SCFA-producing bacteria are believed to be associated with energy harvesting and body weight. Recent study has also shown that there is a possible link between the intestinal microbiota and feed efficiency in pigs (McCormack et al., 2017). Landrace pigs are known for good growth performance and a high ratio of lean meat (Li Z. et al., 2013). In accordance with this characteristic, we found that the average daily weight gain of Landrace piglets was higher than that of Meihua piglets. Collectively, the higher

production of SCFAs and lower amount of secondary bile acids in response to differences in the colonic microbiome may be positively correlated with a faster growth rate in Landrace pigs. Further mechanisms underlying inheritance, diets, and environment in regulating host phenotypes should be explored, focusing on SCFAs and bile acids as well as their receptors.

In conclusion, the microbial communities and metabolome profiles in the colon lumen were influenced by host genetics, and displayed significant differences between the fatty-type Meihua and the lean-type Landrace piglets. In the present study, significant differences in the production of SCFAs and secondary bile acids, as well as expression of their receptors, in the colon between Landrace and Meihua piglets were clearly demonstrated for the first time. The integration of results from the gut luminal metabolome and microbiome not only provide an understanding of the metabolic differences of gut microbiota between these two pig breeds, but also may have great potential as biomarkers for some metabolic diseases in human.

AUTHOR CONTRIBUTIONS

ZJ and ZhuC conceived and designed the experiments, SY, CZ, and TY analyzed the data and drafted the manuscript, WH, JH, QK, JS, ZhoC, QL, and SW performed the experiments and analyzed the data.

ACKNOWLEDGMENTS

This work was jointly supported by the Science and Technology Program of Guangzhou City, China (201607020035), the Presidential Foundation of the Guangdong Academy of Agricultural Sciences, China (201420), the Science and Technology Program of Guangdong Province, China (2016B070701013), and the Hundred Outstanding Talents Training Program at Guangdong Province, China. We thank Prof. Chun-Ming Liu (Institute of Crop Sciences, Chinese Academy of Agricultural Sciences), Dr. Li Wang and Dr. Xuefen Yang (Institute of Animal Science, Guangdong Academy of Agricultural Sciences) for providing useful comments on this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01812/full#supplementary-material

REFERENCES

Ananthakrishnan, A. N. (2015). Epidemiology and risk factors for IBD. *Nat. Rev. Gastroenterol. Hepatol.* 12, 205–217. doi: 10.1038/nrgastro.2015.34

Aronesty, E. (2013). Comparison of sequencing utility programs. *Open Bioinform. J.* 7:18. doi: 10.2174/1875036201307010001

Bauer, P. V., Hamr, S. C., and Duca, F. A. (2016). Regulation of energy balance by a gut-brain axis and involvement of the gut microbiota. *Cell Mol. Life Sci.* 73, 737–755. doi: 10.1007/s00018-015-2083-z

Boets, E., Deroover, L., Houben, E., Vermeulen, K., Gomand, S. V., Delcour, J. A., et al. (2015). Quantification of *in vivo* colonic short chain fatty acid production from inulin. *Nutrients* 7, 8916–8929. doi: 10.3390/nu7115440

Bovo, S., Mazzoni, G., Galimberti, G., Calo, D. G., Fanelli, F., Mezzullo, M., et al. (2016). Metabolomics evidences plasma and serum biomarkers differentiating two heavy pig breeds. *Animal* 10, 1741–1748. doi: 10.1017/S1751731116000483

Brestoff, J. R., and Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. *Nat. Immunol.* 14, 676–684. doi: 10.1038/ni.2640

Cai, X., Liu, Y., Zhou, X., Navaneethan, U., Shen, B., and Guo, B. (2012). An LC-ESI-MS method for the quantitative analysis of bile acids composition in fecal materials. *Biomed. Chromatogr.* 26, 101–108. doi: 10.1002/bmc.1633

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Cheng, P. H., Liang, J. B., Wu, Y. B., Wang, Y., Tufarelli, V., Laudadio, V., et al. (2017). *In vitro* fermentative capacity of swine large intestine: comparison between native Lantang and commercial Duroc breeds. *Anim. Sci. J.* 88, 1141–1148. doi: 10.1111/asj.12723

Clarke, G., Stilling, R. M., Kennedy, P. J., Stanton, C., Cryan, J. F., and Dinan, T. G. (2014). Minireview: gut microbiota: the neglected endocrine organ. *Mol. Endocrinol.* 28, 1221–1238. doi: 10.1210/me.2014-1108

Cresci, G. A., Thangaraju, M., Mellinger, J. D., Liu, K., and Ganapathy, V. (2010). Colonic gene expression in conventional and germ-free mice with a focus on the butyrate receptor GPR109A and the butyrate transporter SLC5A8. *J. Gastrointest. Surg.* 14, 449–461. doi: 10.1007/s11605-009-1045-x

Fiorucci, S., Mencarelli, A., Palladino, G., and Cipriani, S. (2009). Bile-acid-activated receptors: targeting TGR5 and farnesoid-X-receptor in lipid and glucose disorders. *Trends Pharmacol. Sci.* 30, 570–580. doi: 10.1016/j.tips.2009.08.001

Ganapathy, V., Thangaraju, M., Prasad, P. D., Martin, P. M., and Singh, N. (2013). Transporters and receptors for short-chain fatty acids as the molecular link between colonic bacteria and the host. *Curr. Opin. Pharmacol.* 13, 869–874. doi: 10.1016/j.coph.2013.08.006

Goodrich, J. K., Davenport, E. R., Waters, J. L., Clark, A. G., and Ley, R. E. (2016). Cross-species comparisons of host genetic associations with the microbiome. *Science* 352, 532–535. doi: 10.1126/science.aad9379

Ha, C. W., Lam, Y. Y., and Holmes, A. J. (2014). Mechanistic links between gut microbial community dynamics, microbial functions and metabolic health. *World J. Gastroenterol.* 20, 16498–16517. doi: 10.3748/wjg.v20.i44.16498

Hofmann, A. F., and Eckmann, L. (2006). How bile acids confer gut mucosal protection against bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4333–4334. doi: 10.1073/pnas.0600780103

Hu, J., Kyrou, I., Tan, B. K., Dimitriadis, G. K., Ramanjaneya, M., Tripathi, G., et al. (2016). Short-chain fatty acid acetate stimulates adipogenesis and mitochondrial biogenesis via GPR43 in brown adipocytes. *Endocrinology* 157, 1881–1894. doi: 10.1210/en.2015-1944

Islam, K. B., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., et al. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* 141, 1773–1781. doi: 10.1053/j.gastro.2011.07.046

Jiang, G. L., Liu, Y. Y., Oso, A. O., Li, F. N., Kong, X. F., Geng, M. M., et al. (2016). The differences of bacteria and bacteria metabolites in the colon between fatty and lean pigs. *J. Anim. Sci.* 94(Suppl. 3), 349–353. doi: 10.2527/jas.2015-9561

Kasubuchi, M., Hasegawa, S., Hiramatsu, T., Ichimura, A., and Kimura, I. (2015). Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. *Nutrients* 7, 2839–2849. doi: 10.3390/nu7042839

Lalles, J. P. (2016). Microbiota-host interplay at the gut epithelial level, health and nutrition. *J. Anim. Sci. Biotechnol.* 7:66. doi: 10.1186/s40104-016-0123-7

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotech.* 31, 814–821. doi: 10.1038/nbt.2676

Levy, M., Blacher, E., and Elinav, E. (2016). Microbiome, metabolites and host immunity. *Curr. Opin. Microbiol.* 35, 8–15. doi: 10.1016/j.mib.2016.10.003

Li, F., Jiang, C., Krausz, K. W., Li, Y., Albert, I., Hao, H., et al. (2013). Microbiome remodelling leads to inhibition of intestinal farnesoid X receptor signalling and decreased obesity. *Nat. Commun.* 4:2384. doi: 10.1038/ncomms3384

Li, Z., Wu, Z., Ren, G., Zhao, Y., and Liu, D. (2013). Expression patterns of insulin-like growth factor system members and their correlations with growth and carcass traits in Landrace and Lantang pigs during postnatal development. *Mol. Biol. Rep.* 40, 3569–3576. doi: 10.1007/s11033-012-2430-1

Malmuthuge, N., and Guan, L. L. (2016). Gut microbiome and omics: a new definition to ruminant production and health. *Anim. Front.* 6:17. doi: 10.2527/af.2016-0017

McCormack, U. M., Curião, T., Buzoianu, S. G., Prieto, M. L., Ryan, T., Varley, P., et al. (2017). Exploring a possible link between the intestinal microbiota and feed efficiency in pigs. *Appl. Environ. Microbiol.* 83. pii: e00380–e00317. doi: 10.1128/AEM.00380-17

Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., et al. (2012). Host-gut microbiota metabolic interactions. *Science* 336, 1262–1267. doi: 10.1126/science.1223813

Pajarillo, E. A., Chae, J. P., Balolong, M. P., Kim, H. B., Seo, K. S., and Kang, D. K. (2014). Pyrosequencing-based analysis of fecal microbial communities in three purebred pig lines. *J. Microbiol.* 52, 646–651. doi: 10.1007/s12275-014-4270-2

Pajarillo, E. A., Chae, J. P., Kim, H. B., Kim, I. H., and Kang, D. K. (2015). Barcoded pyrosequencing-based metagenomic analysis of the faecal microbiome of three purebred pig lines after cohabitation. *Appl. Microbiol. Biotechnol.* 99, 5647–5656. doi: 10.1007/s00253-015-6408-5

Park, S. J., Kim, J., Lee, J. S., Rhee, S. K., and Kim, H. (2014). Characterization of the fecal microbiome in different swine groups by high-throughput sequencing. *Anaerobe* 28, 157–162. doi: 10.1016/j.anaerobe.2014.06.002

Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494

Parseus, A., Sommer, N., Sommer, F., Caesar, R., Molinaro, A., Stahlman, M., et al. (2017). Microbiota-induced obesity requires farnesoid X receptor. *Gut* 66, 429–437. doi: 10.1136/gutjnl-2015-310283

Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J., Duncan, A. E., Kau, A. L., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341:1241214. doi: 10.1126/science.1241214

Ridlon, J. M., Kang, D. J., and Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* 47, 241–259. doi: 10.1194/jlr.R500013-JLR200

Rios-Covian, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., De Los Reyes-Gavilan, C. G., et al. (2016). Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* 7:185. doi: 10.3389/fmicb.2016.00185

Sivaprakasam, S., Prasad, P. D., and Singh, N. (2016). Benefits of short-chain fatty acids and their receptors in inflammation and carcinogenesis. *Pharmacol. Ther.* 164, 144–151. doi: 10.1016/j.pharmthera.2016.04.007

Sun, J., Zeng, B., Chen, Z., Yan, S., Huang, W., Sun, B., et al. (2017). Characterization of faecal microbial communities of dairy cows fed diets containing ensiled Moringaoleifera fodder. *Sci. Rep.* 7:41403. doi: 10.1038/srep41403

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031. doi: 10.1038/nature05414

Van den Abbeele, P., Belzer, C., Goossens, M., Kleerebezem, M., De Vos, W. M., Thas, O., et al. (2013). Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an *in vitro* gut model. *ISME J.* 7, 949–961. doi: 10.1038/ismej.2012.158

Wahlstrom, A., Sayin, S. I., Marschall, H. U., and Backhed, F. (2016). Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metab.* 24, 41–50.doi: 10.1016/j.cmet.2016.05.005

Wang, J. L., Zhang, T., Shen, X. T., Liu, J., Zhao, D. L., Sun, Y. W., et al. (2016). Serum metabolomics for early diagnosis of esophageal squamous cell carcinoma by UHPLC-QTOF/MS. *Metabolomics* 12:116. doi: 10.1007/s11306-016-1050-5

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Wheelock, A. M., and Wheelock, C. E. (2013). Trials and tribulations of 'omics data analysis: assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* 9, 2589–2596. doi: 10.1039/c3mb70194h

Xiao, Y., Li, K., Xiang, Y., Zhou, W., Gui, G., and Yang, H. (2017). The fecal microbiota composition of boar Duroc, Yorkshire, Landrace and Hampshire pigs. *Asian Australas J. Anim Sci.* 30, 1456–1463. doi: 10.5713/ajas.16.0746

Xiong, Y., Miyamoto, N., Shibata, K., Valasek, M. A., Motoike, T., Kedzierski, R. M., et al. (2004). Short-chain fatty acids stimulate leptin production in adipocytes through the G protein-coupled receptor GPR41. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1045–1050. doi: 10.1073/pnas.2637002100

Yan, S. J., Huang, J. F., Chen, Z. J., Jiang, Z. Y., Li, X., and Chen, Z. (2016). Metabolomics in gut microbiota: applications and challenges. *Sci. Bull.* 61, 1151–1153. doi: 10.1007/s11434-016-1142-7

Yang, L., Bian, G., Su, Y., and Zhu, W. (2014). Comparison of faecal microbial community of lantang, bama, erhualian, meishan, xiaomeishan, duroc, landrace, and yorkshire sows. *Asian Australas J. Anim. Sci.* 27, 898–906. doi: 10.5713/ajas.2013.13621

Zhang, L., Xie, C., Nichols, R. G., Chan, S. H., Jiang, C., Hao, R., et al. (2016). Farnesoid X receptor signaling shapes the gut microbiota and controls hepatic lipid metabolism. *mSystems* 1:e00070-16. doi: 10.1128/mSystems.00070-16

Zhu, C., Lv, H., Chen, Z., Wang, L., Wu, X., Zhang, W., et al. (2017). Dietary zinc oxide modulates antioxidant capacity, small intestine development, and jejunal gene expression in weaned piglets. *Biol. Trace Elem. Res.* 175, 331–338. doi: 10.1007/s12011-016-0767-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Yan, Zhu, Yu, Huang, Huang, Kong, Shi, Chen, Liu, Wang, Jiang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 March 2018
doi: 10.3389/fmicb.2018.00574

Alfalfa Intervention Alters Rumen Microbial Community Development in Hu Lambs During Early Life

Bin Yang[1,2,3], Jiaqing Le[1,3], Peng Wu[1,3], Jianxin Liu[1,3], Le L. Guan[2,3] and Jiakun Wang[1,3]**

[1] Institute of Dairy Science, College of Animal Sciences, Zhejiang University, Hangzhou, China, [2] Department of Agricultural, Food & Nutritional Science, University of Alberta, Edmonton, AB, Canada, [3] ZJU-UoA Joint Laboratory for Livestock Functional Genomics and Microbiology, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council (CONICET), Argentina

Reviewed by:
Shengguo Zhao, Institute of Animal Science (CAAS), China
Ilma Tapio, Natural Resources Institute Finland (Luke), Finland

***Correspondence:**
Le L. Guan
lguan@ualberta.ca
Jiakun Wang
jiakunwang@zju.edu.cn

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: 02 December 2017
Accepted: 13 March 2018
Published: 27 March 2018

Citation:
Yang B, Le J, Wu P, Liu J, Guan LL and Wang J (2018) Alfalfa Intervention Alters Rumen Microbial Community Development in Hu Lambs During Early Life. Front. Microbiol. 9:574. doi: 10.3389/fmicb.2018.00574

The pre-weaning period is crucial for rumen developmental plasticity, which can have a long-term impact on animal performance. Understanding the rumen microbiota during early life is important to elucidate its potential role in rumen development. In this study, the rumen microbiota of 10-day-old Hu lambs fed either milk replacer (B-10), milk replacer and starter (STA) or milk replacer and starter supplemented with alfalfa (S-ALF) in the pre- (d17, 24, and 38) and post-weaning periods (d45 and 66) were assessed to characterize rumen microbial colonization during early life and its response to fiber intervention. In the rumens of B-10 lambs, 498 operational taxonomic units belonging to 33 predominant genera were observed, and the top six predicted functions included "Membrane transport," "carbohydrate metabolism," "amino acid metabolism," "replication and repair," "translation," and "energy metabolism." *Prevotella*, *Succinivibrio*, *Bifidobacterium*, and *Butyrivibrio* abundances were increased at d38 for both STA and S-ALF groups compared to the B-10 group, whereas fibrolytic bacteria of the taxa *Lachnospiraceae* and *Treponema* were only increased in the S-ALF group at d38. A number of saccharolytic bacteria (*Bacteroidaceae*), organic acid-producing bacteria (*Coprococcus* and *Actinomyces*), proteolytic and amino acid fermenters (*Fusobacterium*) and fibrolytic bacteria (unclassified *Ruminococcaceae*) were significantly decreased in the STA lambs but not in the S-ALF lambs at d38. After weaning and exposed to alfalfa, the rumen microbial composition in the STA group started to appear similar to that of the S-ALF lambs. The relative abundance of unclassified *Clostridiales* was higher in S-ALF lambs than STA lambs after weaning. Spearman's correlation analysis showed positive relationships between unclassified *Lachnospiraceae*, unclassified *Clostridiales*, *Treponema*, unclassified *Bacteroidales*, *Coprococcus* and crude protein intake, neutral detergent fiber intake, and plasma β-hydroxybutyrate. The unclassified *Lachnospiraceae* and *Treponema* were also positively correlated with average daily gain. Our results revealed that alfalfa stimulated

changes in rumen microbiota during the pre- and post-weaning periods and was consistent with rumen development for better feed intake and animal performance before and after weaning. The findings of this study provide clues for strategies to improve rumen function through manipulation of the rumen microbiota during early life.

Keywords: Hu lamb, rumen microbiota, starter, alfalfa intervention, amplicon sequencing

INTRODUCTION

Early life, especially the pre-weaning period, is a critical period for the developmental plasticity of mammals and can have a long-term impact on various biological functions (Fisher et al., 2012; Bartol et al., 2013; Soberon and Van Amburgh, 2013). Within the first few weeks after birth, young ruminants face a weaning transition and dietary changes from milk or milk replacer to a solid diet. With the change in diet, the gastrointestinal tissues must transition from metabolizing glucose from milk to short-chain fatty acids from a solid diet, especially in the rumen (Baldwin et al., 2004). The weaning transition results in tremendous gastrointestinal and metabolic ramifications for the calf/lamb growth rate (Budzynska and Weary, 2008; de Passillé et al., 2011).

A starter diet containing highly fermentable carbohydrate has been widely used to feed young pre-weaned ruminants due to its ability to promote rumen development by enhancing rumen fermentation, primarily volatile fatty acid (VFA) production (Baldwin et al., 2004; Heinrichs, 2005; Suárez et al., 2006). Past research has revealed that the physical characteristics of feed, such as the particle sizes of roughages, can contribute to ruminal muscular development and size expansion (Tamate et al., 1962). A recent study observed that alfalfa supplementation to starter diets during the pre-weaning period increased rumen papillae length and rumen weight, decreased the incidence of feed plaques, and consequently led to increased feed intake, average daily gain (ADG), and carcass weight during the pre- and post-weaning periods (Yang et al., 2015). In addition, microbial colonization can also affect rumen development and function during early life. Using next generation DNA sequencing techniques, Li et al. (2012) and Jami et al. (2013) observed that prior to weaning, the ruminal microbiota has a similar functional capacity as that of a mature ruminant. Rey et al. (2012) confirmed these findings by measuring enzyme activities in the rumens of dairy calves from birth through weaning. Jami et al. (2013) deduced that the appearance of microbial populations during early life is not dependent on nutrient digestion with limited functional capacity but that it may play a role in long-term imprinting of the microbial community. Such speculation was supported by observations of reduced methane emissions in adult lambs after altering the methanogen community during the pre-weaning phase (Abecia et al., 2014). Based on these findings, we hypothesized that the positive effects of alfalfa supplementation to starter diets on both pre- and post-weaned lambs could be due to the impact on rumen microbial colonization. Therefore, in this study, we assessed the rumen microbiota of Hu lambs fed either a starter or a starter supplemented with alfalfa from the pre- to post-weaning period with the aim of characterizing the effects of alfalfa intervention on rumen microbial colonization during early life.

MATERIALS AND METHODS

Animal Study and Sample Collection

All the experimental protocols performed in this study were approved by the Animal Care Committee of Zhejiang University (Hangzhou, China), and the experimental procedures used in this study were in accordance with the recommendations of the University's guidelines for animal research.

In a previously published animal study (Yang et al., 2015), rumen samples were collected from Hu lambs without separating the solid and liquid fractions. Briefly, 66 healthy male Hu lambs at d5 [body weight (BW) = 3.69 ± 0.67 kg (mean ± SD)] were purchased and housed at the University research facility and fed milk replacer for 5 adaptation days before the feeding trial. At d10, six lambs were sacrificed as the baseline group (B-10), and the other 60 animals were randomly assigned to one of two diets (STA or S-ALF) and sacrificed at the age of d17, 24, 38, 45, or 66. Six lambs were assigned to each group for each sampling age. From d10 to d38 (pre-weaning), the lambs in the STA group were fed milk replacer and *ad libitum* starter pellets, whereas the lambs of the S-ALF group were provided the same starter diet with supplemental *ad libitum* chopped alfalfa. After weaning (from d38 to d66), all the lambs were fed 300 g/d of a concentrate mixture (Supplementary Table S1) and *ad libitum* alfalfa. BW was measured on two consecutive days before morning feeding in the beginning of the experiment (d10, initial BW), before sacrifice (end BW) and every week to calculate the ADG. Daily feed and ort samples were collected for chemical analysis to determine the intake of crude protein (CP) and neutral detergent fiber (NDF). Lambs were sacrificed before morning feeding, with plasma obtained before sacrifice and rumen tissues collected after sacrifice to measure the concentration of β-hydroxybutyrate (BHBA) and ruminal papillae length and width. Detailed information was published in Yang et al. (2015), with the exception that the data collected on the week of animal sacrifice for bacterial analysis were used in the present study.

The rumen content samples were collected immediately after sacrifice and stored at −20°C until further analysis. Only liquids were obtained from the rumen of the B-10 lambs, whereas mixed liquid and solid contents were obtained from the rumen of the STA and S-ALF lambs. After filtering out the lambs without solid feed intake (only milk replacer in the rumen) or lambs for which

no rumen content was present after sacrifice, with the exception of the B-10 lambs (Yang et al., 2015), 55 valid rumen samples were acquired (6 samples from the B-10 group; 25 samples from the STA group, including 4, 3, 6, 6, and 6 samples at d17, 24, 38, 45, and 66, respectively; and 24 samples from the S-ALF group, including 4, 4, 5, 6, and 5 samples at d17, 24, 38, 45, and 66, respectively).

Total DNA Extraction, Illumina Sequencing, and Data Processing

Total DNA from the rumen content samples was extracted using the cetyltrimethylammonium bromide method (Brookman and Nicholson, 2005) with a bead-beater (Biospec Products; Bartlesville, OK, United States) as described by Gagen et al. (2010). The amplicon library of the V4 hypervariable region of the 16S rRNA gene was prepared from each of the DNA samples using the primer set 515F/806R and Phusion® High-Fidelity PCR Master Mix (New England Biolabs, Ipswich, MA, United States) as described by Caporaso et al. (2011). Each forward and reverse primer had a 6-bp error-correcting barcode at the 5′ terminus that was unique to each DNA sample. The amplicon libraries for all samples were pooled at an equimolar ratio and sequenced on an Illumina HiSeq platform by Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China) to generate 2 × 250 bp paired-end reads.

The paired-end reads were joined to form single sequences using FLASH (Magoč and Salzberg, 2011) based on overlapping regions. The sequences were demultiplexed and assigned to each sample according to the individual unique barcode using Quantitative Insight into Microbial Ecology (QIIME, Caporaso et al., 2010). Sequences with a quality score of <20 and a length of >300 bp or <200 bp were discarded. Possible chimeric sequences were identified and removed using the usearch61 algorithm in USEARCH 6.1 (Edgar, 2010) with the Gold database[1]. Operational taxonomic units (OTUs) were clustered at a 97% identity threshold, and taxa were assigned using the core set in the Greengenes 13.8 database (DeSantis et al., 2006) using the UCLUST algorithm (Edgar, 2010). The alpha diversity of the ruminal bacteria was estimated using the number of OTUs, Chao1, Shannon indices, and Good's coverage implemented in QIIME (Caporaso et al., 2010). Analysis of similarity (ANOSIM) was used to test whether a significant difference existed between two groups of samples. An R-value > 0.75 with a P-value < 0.05 denotes groups that are completely different from one another; $0.5 < R$-value < 0.75 with a P-value < 0.05 denotes groups that are different from one another; $0.3 < R$-value < 0.5 with P-value < 0.05 denotes groups that tend to be different from one another; and an R-value < 0.3 denotes groups that are not different from one another.

Predicted Microbial Functions Using PICRUSt

PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States, Langille et al., 2013) was used to predict the functional capabilities of microbial communities based on the 16S rRNA gene data against a Greengenes reference taxonomy (Greengenes 13.8). Briefly, after the abundance of each OTU was normalized by marker gene copy number, the molecular functions were predicted by the KEGG pathways.

Statistical Analysis

All the data were subjected to a normal distribution test using SPSS 20.0 (SPSS, Inc., Chicago, IL, United States). Normally distributed data included alpha diversity indices (number of OTUs, Chao1, and Shannon index) and the phenotypic data (CP and NDF intake, initial and end BW, ADG, plasma BHBA concentration, and rumen papillae length and width). For the alpha diversity indices, an orthogonal polynomial regression analysis was performed to analyze the linear and quadratic effects of age from d10 to d38 and from d10 to d66 with Tukey's multiple comparison test of the mean values performed among ages; one-way ANOVA was performed to analyze differences between the STA and S-ALF groups. For the phenotypic data one-way ANOVA was performed to analyze the age effect from d17 to d66 and the difference between the STA and S-ALF groups at different ages.

The non-normally distributed data included the relative abundance of rumen bacteria and the predicted KEGG pathway relative abundance. For the relative abundances of rumen bacteria, all the taxa analyzed in the present study were identified in at least 4 lambs at each assayed age for each feeding group (4 out of 6 lambs; except for d24 of STA group, which were identified in 3 lambs), with an average relative abundance of ≥0.5% in at least one age group (data were shown in Supplementary Table S2). This result meant that only the "core" bacteria that were observed in at least 60% of samples were considered in the relative abundance analysis. Taxa identified in each sample within each group but with an abundance <0.5% were not considered "core" bacteria. To show the changes in bacterial relative abundances, bacterial data were presented as $\log_2$ (fold change of d17-38 to B-10) during the pre-weaning period, and $\log_2$ (fold change of d45-66 to d38) during the post-weaning period, with taxa subjected to LEfSe analysis (Segata et al., 2011). A significant change was observed with a LDA (Linear Discriminant Analysis) score > 2.0 calculated by LEfSe. For the KEGG pathway relative abundance generated by PICRUSt, a Kruskal–Wallis signed rank test was performed in SPSS 20.0. A significant change was observed with $P \leq 0.05$.

Spearman's rank correlations between the ruminal bacteria (relative abundance) and CP intake, NDF intake, ADG, concentration of plasma BHBA, rumen papillae length and width were analyzed using SPSS 20.0, and only the significant correlations ($P \leq 0.05$) of the changes in bacteria were plotted by R software (version 3.3.0) and the package "corrplot".[2]

Accession Number(s)

The paired-end sequence data were deposited and are available in the European Nucleotide Archive with the accession numbers

[1] https://drive5.com/uchime/uchime_download.html

[2] https://CRAN.R-project.org/package=corrplot

ERS1929787-ERS1929792, ERS1485434-1485443, ERS1485445-1485470, ERS1485472-1485482, ERS1509216, and ERS1509217.

RESULTS

Changes in Rumen Bacterial Diversity During Early Life of Hu Lambs in the STA and S-ALF Groups

A total of 1,458,338 qualified sequences were obtained from 55 rumen samples with an average of 32,734 ± 10,713 sequences per sample, and 1,682 operational taxonomic units (OTUs, 579 ± 85 OTUs per sample) were detected based on 97% similarity. With a subsample of 23,600 sequences (the minimum number detected) for each sample, the Good's coverage (>0.98) revealed that our data provided sufficient sequencing depth to accurately describe the rumen bacterial composition of the Hu lambs used in this study.

In the rumens of the B-10 Hu lambs, 498 OTUs were identified with a Shannon index of 5.29 (**Table 1**). When considering the whole experimental period from d10 to d66, the number of OTUs and the Shannon index increased linearly for lambs in both the STA and S-ALF groups ($P < 0.05$, **Table 1**), and no significant difference ($P > 0.05$) was observed between lambs in the two groups for each age. However, the alpha diversity patterns for the lambs in the two groups were different during the pre- and post-weaning periods. The number of OTUs increased quadratically ($P < 0.05$) from d10 to d38 for lambs in the STA group, and the highest number of OTUs appeared on d17. After weaning, the number of OTUs and the Shannon index increased significantly ($P < 0.05$) when compared with the values observed on d38 for lambs of the STA group, whereas these two indices were not significantly higher on d45 than d38 for lambs in the S-ALF group (**Table 1**). At d66, the number of OTUs and the Shannon index for lambs in the STA group increased to 637 ± 44 and 6.68 ± 0.29, respectively, and these two indices for S-ALF lambs increased to 675 ± 77 and 7.11 ± 0.69 (**Table 1**).

Succession of Rumen Bacterial Communities During the Early Life of Hu Lambs in the STA and S-ALF Groups

Similar to the observed changing alpha diversity patterns, the analysis of similarity (ANOSIM) results showed that the rumen microbial communities were similar among the B-10, d17 and d24 lambs before weaning and between d45 and d66 lambs after weaning, and there was no significant difference between lambs of the two groups at each age ($R < 0.3$ with $P > 0.05$, **Table 2**). For the STA group, the rumen microbial communities in the d38 animals were different from those observed in the B-10 and STA d17 and d24 lambs ($0.5 < R < 0.75$ with $P < 0.05$), and the microbial communities were significantly shifted after weaning. Compared to the d38 STA lambs, the *R*-value of ANOSIM was 0.887 ($P < 0.05$) for d45 and 1.000 ($P < 0.05$) for d66 STA lambs (**Table 2**). For the S-ALF group, the microbial communities in d38 lambs tended to be different from those in the B-10 and STA d17 and d24 lambs ($0.3 < R < 0.5$ with $P < 0.05$), and this tendency was observed at d45 after weaning (**Table 2**). Compared to the d38 S-ALF lambs, the *R*-value of ANOSIM was 0.403 ($P < 0.05$) for d45 and 0.648 ($P < 0.05$) for d66 S-ALF lambs (**Table 2**).

Divergence of the Rumen Bacterial Communities During the Pre-weaning Period of Hu Lambs in the STA and S-ALF Groups

Rumen Microbial Composition of Milk Replacer-Fed Hu Lambs

Seven predominant phyla were identified in the rumens of the B-10 lambs. Among them, *Bacteroidetes*, *Firmicutes* and

TABLE 1 | Alpha diversity (mean ± SD) of rumen bacterial communities in Hu lambs with (S-ALF) or without (STA) alfalfa intervention at different ages.

	STA			S-ALF			STA vs. S-ALF[2]		
Age (d)	**OTUs**	**Chao1**	**Shannon**	**OTUs**	**Chao1**	**Shannon**	**OTUs**	**Chao1**	**Shannon**
10	498 ± 60^{b}	1028 ± 98	5.29 ± 0.67^{bc}	498 ± 60^{b}	1028 ± 98	5.29 ± 0.67^{b}	–	–	–
17	566 ± 57^{ab}	1016 ± 110	5.33 ± 0.60^{bc}	558 ± 72^{ab}	1007 ± 75	5.50 ± 0.89^{ab}	0.868	0.907	0.759
24	554 ± 45^{ab}	1044 ± 81	5.33 ± 0.20^{bc}	625 ± 127^{ab}	1070 ± 66	5.94 ± 1.38^{ab}	0.407	0.660	0.496
38	487 ± 36^{b}	946 ± 98	4.86 ± 0.31^{cd}	566 ± 100^{ab}	1007 ± 108	5.61 ± 0.89^{ab}	0.102	0.354	0.086
45	586 ± 73^{a}	1038 ± 95	6.14 ± 0.92^{ab}	626 ± 31^{ab}	1062 ± 42	6.65 ± 0.54^{ab}	0.244	0.574	0.263
66	637 ± 44^{a}	1052 ± 67	6.68 ± 0.29^{a}	675 ± 77^{a}	1135 ± 118	7.11 ± 0.69^{a}	0.331	0.177	0.190
Age effect[1]									
From d10 to 38									
Linear	0.641	0.260	0.201	0.144	0.994	0.469	–	–	–
Quadratic	0.014	0.380	0.314	0.182	0.628	0.551	–	–	–
From d10 to 66									
Linear	0.002	0.795	<0.001	0.002	0.063	0.001	–	–	–
Quadratic	0.118	0.316	0.004	0.805	0.242	0.293	–	–	–

$^{a-c}$Means within a column with different superscripts differ (P-value ≤ 0.05). [1]Linearly and quadraticly effect of age during the pre-weaning period (from d10 to 38), and the whole period (from d10 to 66). [2]P-value of the comparison between STA and S-ALF group.

TABLE 2 | *R*- and *P*-values of pairwise comparison between the different ages in the groups with (S-ALF) or without (STA) alfalfa intervention performed using analysis of similarity (ANOSIM)[1].

R-values	B-10	STA-17	STA-24	STA-38	STA-45	STA-66	S-ALF-17	S-ALF-24	S-ALF-38	S-ALF-45	S-ALF-66
B-10	0										
STA-17	0.298	0									
STA-24	0.173	−0.093	0								
STA-38	0.591*	0.385*	0.531*	0							
STA-45	0.550*	0.504*	0.543*	0.887*	0						
STA-66	0.624*	0.794*	0.815*	1.000*	0.150	0					
S-ALF-17	0.024	0.031	0.074	0.738*	0.468*	0.706	0				
S-ALF-24	0.302	0.125	−0.130	0.690*	0.361*	0.496	0.208	0			
S-ALF-38	0.440*	0.094	0.405	0.035	0.523*	0.688	0.462*	0.394*	0		
S-ALF-45	0.559*	0.464*	0.469*	0.722*	0.102	0.028	0.488*	0.282*	0.403*	0	
S-ALF-66	0.419*	0.587*	0.579*	0.976*	0.133	0.091	0.388*	0.106	0.648*	0.021	0

[1]*R-value > 0.75 with P-value < 0.05 denote groups completely different from one another; 0.5 < R-value < 0.75 with P-value < 0.05 denote groups different from one another; 0.3 < R-value < 0.5 with P-value < 0.05 denote groups tend to be different from one another; R-value < 0.3 denote groups not different from one another.* **Represents the pairwise comparison with P-value < 0.05.*

Proteobacteria were the dominant phyla and accounted for 41.5, 35.5, and 17.8% of the total sequences, respectively, followed by *Fusobacteria* (1.7%), *Actinobacteria* (1.4%), *Verrucomicrobia* (1.1%) and *Chloroflexi* (0.6%) (Supplementary Table S3). Thirty-three predominant genera were identified in the rumens of the B-10 lambs, with *Bacteroides* (20.9%) being the most abundant; the relative abundances of unclassified *BS11*, *Prevotella*, and *Dialister* were between 5 and 10%; unclassified *Lachnospiraceae*, unclassified *Clostridiales*, *Eikenella*, *Sharpea*, *Porphyromonas*, unclassified *Pasteurellaceae*, unclassified *Ruminococcaceae*, *Bibersteinia*, *Fusobacterium*, *Mitsuokella*, *Coprococcus*, *Megasphaera*, *Oscillospira*, unclassified *Enterobacteriaceae*, *Butyrivibrio*, *Streptococcus*, *Akkermansia*, unclassified *Mogibacteriaceae*, and *Lactobacillus* were between 1 and 5%; and unclassified *Bacteroidales*, *Moraxella*, unclassified *S24-7*, *Ruminococcus*, *CF231*, unclassified *Coriobacteriaceae*, *Sutterella*, *SHD-231*, unclassified *Aeromonadaceae*, and unclassified *Veillonellaceae* were less than 1% (Supplementary Table S4).

Microbial Compositional Changes in the Rumen of STA-Fed Lambs

Figure 1 shows the taxa that significantly ($P \leq 0.05$) changed in abundance during the pre-weaning period. Compared to the B-10 group, 15 core families and 15 core genera were altered in the STA group on d38 (**Figures 1A,B**). The families *Coriobacteriaceae*, *S24-7*, *Prevotellaceae*, *Bifidobacteraceae,* and *Succinivibrionaceae* and the genera *Prevotella*, *Bifidobacterium*, *Succinivibrio*, *Butyrivibrio*, unclassified genera within *Paraprevotellaceae* and *Coriobacteriaceae* increased significantly. In contrast, the families *Shewanellaceae*, *Pasteurellaceae*, *Alcaligenaceae*, *Fusobacteriaceae*, *Moraxellaceae*, *Bacteroidaceae*, *Neisseriaceae*, *Enterobacteriaceae*, *Desulfovibrionaceae*, and *Actinomycetaceae* and the genera *Moraxella*, *Bibersteinia*, *Fusobacterium*, *Coprococcus*, *Bacteroides*, *Actinomyces*, unclassified genera within *Pasteurellaceae*, *Enterobacteriaceae*, and *Ruminococcaceae* decreased significantly.

Although fluctuations in the bacterial relative abundances between the B-10 group and the STA lambs on d38, the 15 families and 15 genera that exhibited significant changes in the d38 animals compared to the B-10 group were observed to continuously increase or decrease from B-10 to d38 in the STA group (**Figures 1C,D**).

Microbial Compositional Changes in the Rumen of the S-ALF-Fed Lambs

Compared to the B-10 lambs, 15 core families and 10 core genera were altered in the S-ALF group on d38 (**Figures 1A,B**). In addition to the changes in taxa observed in the d38 STA lambs, the families *Spirochaetaceae*, *Lachnospiraceae*, and *p-2534-18B5* and the genus *Treponema* were additionally increased in the d38 S-ALF lambs. In contrast, the families *Fusobacteriaceae*, *Bacteroidaceae*, and *Actinomycetaceae* and the genera *Moraxella*, *Fusobacterium*, *Coprococcus*, *Bacteroides*, *Actinomyces*, unclassified genera within *Pasteurellaceae* and *Ruminococcaceae* were not significantly decreased at d38 in the S-ALF lambs. The relative abundance of *Eikenella* was significantly decreased at d38 in the S-ALF lambs.

These significantly changes in families and genera on in the S-ALF group on d38 compared to the B-10 group were continuously increased or decreased from B-10 to d38 (**Figures 1C,D**), except for families *Lachnospiraceae* and *p-2534-18B5*, with the family *Lachnospiraceae* increasing since d24 and the family *p-2534-18B5* was decreased at d24 (**Figure 1D**).

Comparing the relative abundances of taxa between the two groups of lambs, unclassified *Lachnospiraceae*, unclassified *Ruminococcaceae*, and *Oscillospira* were significantly higher in the rumen of the d38 S-ALF lambs than that in d38 STA lambs (Supplementary Figure S1).

Effect of Pre-weaned Dietary Intervention on Rumen Microbial Adaptation to Weaning and Diet Transition

Although the rumen bacterial communities in the d45 and d66 lambs did not differ within or between the STA or S-ALF groups (**Table 2**), the succession of the rumen bacterial communities

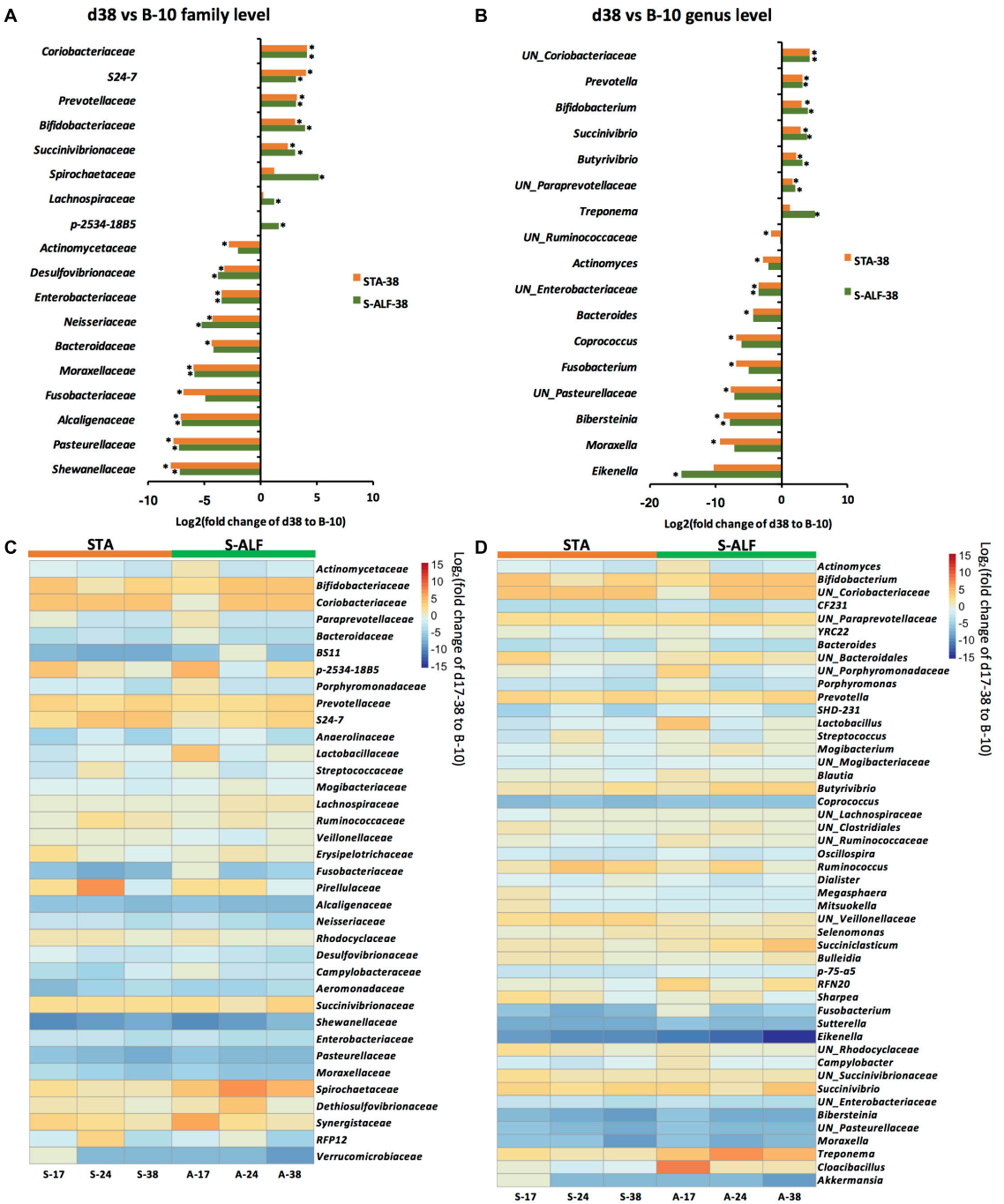


FIGURE 1 | Succession of rumen bacterial communities during the pre-weaning period in the groups with (S-ALF) or without (STA) alfalfa intervention. **(A)** Family level changes between the 10-day baseline group (B-10) and the STA and S-ALF groups at end of the pre-weaning period (d38). **(B)** Genus level changes between B-10 and the d38 lambs. **(C)** Bacterial changes at the family level from the d10 to d38 lambs. **(D)** Bacterial changes at the genus level from d10 to d38. * in **(A)** and **(B)** represents significant difference compared to the B-10 group. UN denotes an unclassified genus.

from d38 to the post-weaning period was different. **Figure 2** shows the significant changes in the families and genera of the d45 and d66 vs. d38 lambs for the STA and S-ALF groups. Weaning and diet transition significantly increased the relative abundances of *Paraprevotellaceae*, *Anaerolinaceae*, *Mogibacteriaceae* and *Desulfovibrionaceae* in both the STA and S-ALF groups (**Figure 2A**). Among these taxa, the significant increases in *Anaerolinaceae* and *Mogibacteriaceae* were observed in the S-ALF group since d45, while the significant increases of the 4 families in the STA group were observed at d66

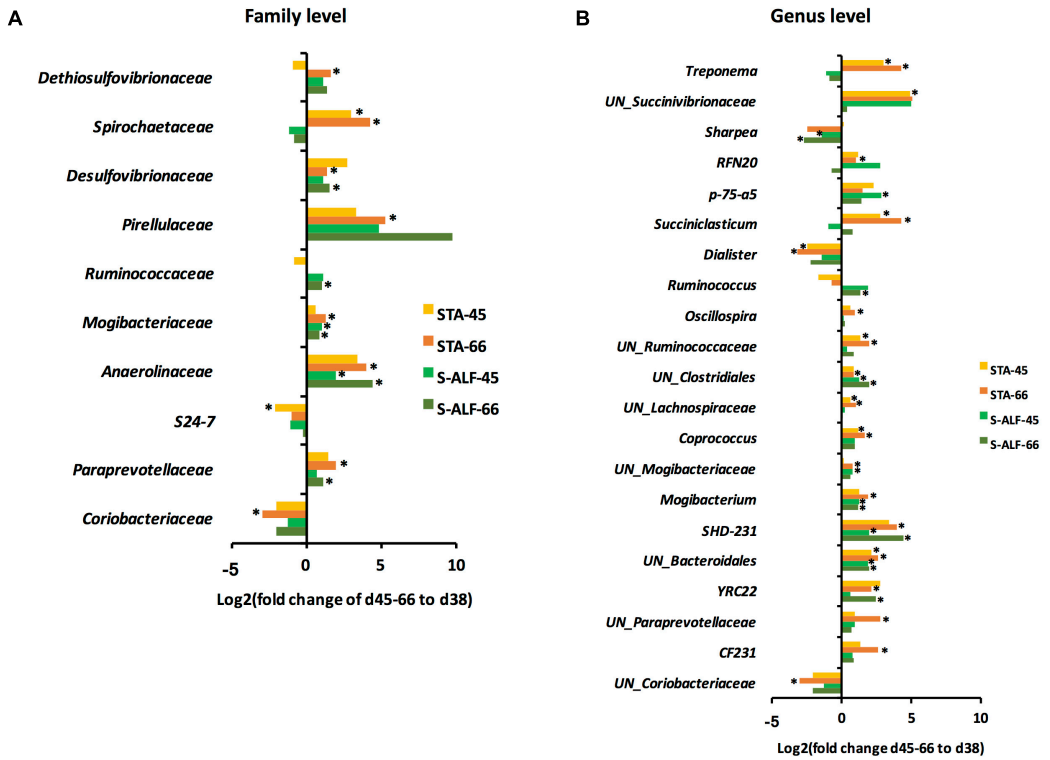


FIGURE 2 | Changes in the rumen bacterial communities during the post-weaning period in lambs with (S-ALF) or without (STA) alfalfa intervention. **(A)** Family level changes between post-weaning (d45 and 66) and d38 lambs. **(B)** Genus level changes between post-weaning (d45 and 66) and d38 lambs. *Represents significant difference when compared with d38 age group. UN denotes an unclassified genus.

(**Figure 2A**). Additionally, *Spirochaetaceae*, which was increased in the S-ALF lambs before weaning (**Figure 1A**), was increased in the STA lambs since d45 (**Figure 2A**). The *Coriobacteriaceae* was increased in both the STA and S-ALF groups before weaning (**Figure 1A**), but decreased after weaning, and this decrease was significant in the STA group on d66 (**Figure 2A**). The *Ruminococcaceae* was only significantly increased in the S-ALF group on d66 (**Figure 2A**). The *Dethiosulfovibrionaceae* and *Pirellulaceae* were only significantly increased in the STA groups on d66 (**Figure 2A**).

Weaning and diet transition significantly increased the *YRC22*, *SHD-231*, *Mogibacterium*, and unclassified genera within the taxa *Mogibacteriaceae*, *Bacteroidales* and *Clostridiales* in both the STA and S-ALF groups (**Figure 2B**). The relative abundances of unclassified *Lachnospiraceae*, unclassified *Ruminococcaceae*, and *Treponema*, which were increased in the S-ALF lambs before weaning (**Figure 1B**), were increased in the STA lambs since d45 (**Figure 2B**). The relative abundance of *Coprococcus*, which decreased with and without significance in the STA and S-ALF lambs at d38 (**Figure 1B**), increased significantly in the STA lambs since d45 (**Figure 2B**). The significantly increased *Ruminococcus* was only observed in S-ALF group on d66 (**Figure 2B**).

Comparing the relative abundances of taxa between the two groups of lambs, the unclassified *Mogibacteriaceae* was higher in the S-ALF lambs at d45; and the *SHD-231*, *Bibersteinia*, *Actinomyces*, and unclassified *Clostridiales* were higher, while the *Prevotella*, unclassified *Paraprevotellaceae*, and *Campylobacter* were lower in the S-ALF lambs at d66 (Supplementary Figure S1).

Divergence of Predicted Rumen Microbial Functions in the STA and S-ALF Groups

"Membrane transport" (11.1%), "carbohydrate metabolism" (10.5%), "amino acid metabolism" (10.0%), "replication and repair" (9.4%), "translation" (6.2%), and "energy metabolism" (5.9%) were identified as the top six predicted functions for the rumen microbiota in the B-10 lambs, which were also the top predicted functions for rumen microbiota in the STA and S-ALF lambs across all ages. In the STA lambs, the functions "nucleotide metabolism" and "xenobiotic biodegradation and metabolism" increased ($P < 0.05$), whereas "signal transduction," "glycan biosynthesis and metabolism," and "metabolism of cofactors and vitamins" decreased ($P < 0.05$) in d38 lambs compared to the B-10 group (**Figure 3A**). For the STA lambs, after weaning, "cell growth and death," "cell motility," "signal transduction," "amino acid metabolism," "biosynthesis of other secondary metabolites," "energy metabolism," "glycan biosynthesis and metabolism," and "metabolism of cofactors and vitamins" increased ($P < 0.05$), whereas "membrane transport," "signaling molecules and interactions," "translation," "carbohydrate metabolism," "nucleotide metabolism," and "xenobiotic biodegradation

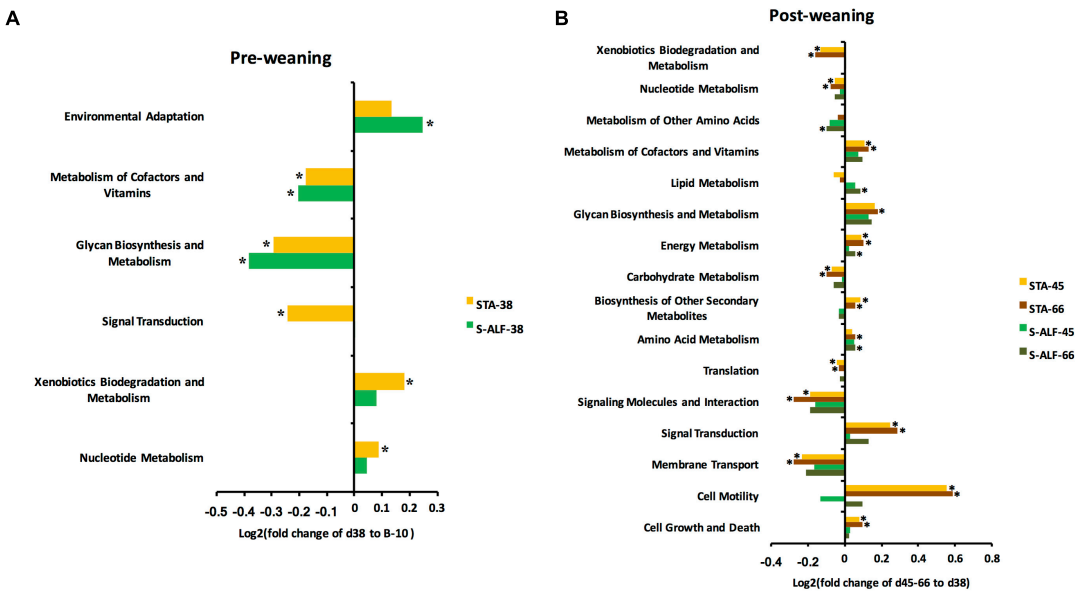


FIGURE 3 | Changes in predicted metagenomic functions of rumen bacteria in lambs with (S-ALF) or without (STA) alfalfa intervention during the pre- **(A)** and post-weaning period **(B)**. *Represents a significant difference ($P \leq 0.05$) when compared to B-10 during the pre-weaning period and when compared to the d38 lambs during the post-weaning period.

and metabolism" decreased ($P < 0.05$) compared to the d38 animals (**Figure 3B**).

For the S-ALF group, "environmental adaptation" increased while "glycan biosynthesis and metabolism," and "metabolism of cofactors and vitamins" decreased at d38 compared to the B-10 group ($P < 0.05$, **Figure 3A**). After weaning, the changes in the S-ALF were less pronounced when compared to the STA lambs. Only "amino acid metabolism," "energy metabolism," and "lipid metabolism" increased ($P < 0.05$), and the "metabolism of other amino acids" decreased ($P < 0.05$) in d66 lambs compared to d38 lambs (**Figure 3B**).

Relationship Between Bacterial Community and Phenotypic Variables

The CP and NDF intake, initial and end BW, ADG, plasma BHBA concentration, rumen papillae length and width were obtained from our previous study (Yang et al., 2015) and re-analyzed with data of 55 lambs for bacteria analysis only in the week of sacrifice (**Table 3**). Correlation analysis showed that the relative abundances of *CF231*, *YRC22*, *SHD-231*, *Mogibacterium*, *Butyrivibrio*, *Coprococcus*, *Succiniclusticum*, *p-75-a5*, *Treponema*, and unclassified genera within the *Bacteroidales*, *Paraprevotellaceae*, *Clostridiales*, *Mogibacteriaceae*, *Lachnospiraceae*, and *Ruminococcaceae* taxa were positively correlated with CP and NDF intake, while the relative abundances of *Sharpea* and unclassified *Pasteurellaceae* were negatively correlated with CP and NDF intake; *Eikenella* was negatively correlated with CP intake, and *Actinomyces* and *Dialister* were negatively correlated with NDF intake ($P < 0.05$, **Figure 4**). Significant correlations were also observed between these bacteria and rumen developmental parameters of plasma BHBA and ruminal papillae length or width, except for *YRC22* and unclassified *Ruminococcaceae*. In addition, *Bifidobacterium* and *Moraxella* were negatively correlated with plasma BHBA, and unclassified *Succinivibrionaceae* was negatively correlated with ruminal papillae width ($P < 0.05$, **Figure 4**). The relative abundances of the genera *YRC22*, *Butyrivibrio*, *Succiniclasticum*, *Treponema*, and unclassified genera within the families *Paraprevotellaceae* and *Lachnospiraceae* showed positive correlations with ADG, while the genera *Sharpea*, *Eikenella*, and unclassified genera within the family *Pasteurellaceae* showed negative correlations with ADG ($P < 0.05$, **Figure 4**).

DISCUSSION

The structure of the rumen microbiome during early life has recently attracted attention because of its potential relationship with rumen development (Li et al., 2012; Jami et al., 2013) and its long-term impact on an animal's performance (Cahenzli et al., 2013; Abecia et al., 2014). Although the appearance of the microbial populations precedes rumen development, it has been suggested that the development of the rumen and its microbiota begins with the intake of solid feed (Rey et al., 2014). In the rumens of the B-10 lambs, the 498 observed OTUs belonged to 33 predominant genera and the top six predicted functions included "Membrane transport," "carbohydrate metabolism," "amino acid metabolism," "replication and repair," "translation," and "energy metabolism." These findings confirmed that the appearance of the metabolically related microbial populations precedes rumen development and is not dependent on solid feed intake as was previously suggested (Jami et al., 2013).

TABLE 3 | Phenotypic variables in Hu lambs of the groups with (S-ALF) or without (STA) alfalfa intervention at different ages.

		STA (d)						S-ALF (d)					
Items[1]	B-10	17	24	38	45	66	SEM[2]	*P*-value[2]	17	24	38	45	66	SEM[3]	*P*-value[3]
CP intake, g	19.57	25.30	20.96	81.80	54.29	126.85	8.630	< 0.001	24.82	24.34	107.26*	102.25*	154.59*	10.504	< 0.001
NDF intake, g	3.85	6.33	6.37	54.03	85.33	197.08	15.242	< 0.001	6.35	11.49	80.49*	162.77*	244.49*	19.211	< 0.001
Initial BW, kg	3.10	3.64	3.68	3.62	3.99	3.90	0.151	0.913	3.65	3.08	3.88	3.79	4.16	0.134	0.140
End BW, kg	3.10	3.80	3.95	6.18	7.05	12.18	0.679	< 0.001	3.80	3.41	7.38	8.76	15.21*	0.885	< 0.001
ADG, kg	–	0.054	0.005	0.182	0.034	0.280	0.026	< 0.001	0.050	0.032	0.251*	0.183*	0.268	0.022	< 0.001
Rumen papillae															
Length, mm	0.383	0.508	0.484	1.474	1.028	1.329	0.086	< 0.001	0.431	0.587*	1.681	1.311*	1.626*	0.108	< 0.001
Width, mm	0.171	0.209	0.226	0.268	0.207	0.234	0.011	0.311	0.208	0.233	0.323*	0.209	0.251	0.013	0.006
BHBA, μmol/L	52.8	344.3	236.0	241.2	400.4	471.4	30.036	0.032	142.8	133.7	339.7	789.8*	560.0	65.031	< 0.001

*[1]CP, crude protein; NDF, neutral detergent fiber; Initial BW, body weight at d10; End BW, body weight before sacrifice; ADG, average daily gain; BHBA, β-hydroxybutyrate. [2]SEM and P-value of the STA group from d17 to 66. [3]SEM and P-value of the S-ALF group from d17 to 66. *Indicates that the item was significantly different in the S-ALF group (P ≤ 0.05) when compared to the same age of the STA group.*

FIGURE 4 | Spearman's correlation between the rumen bacterial communities (genus level) and phenotypic variables. The color and dot size represent the correlation coefficient. Red represents a positive correlation, and blue represents a negative correlation; a larger dot size represents a stronger correlation, and a smaller dot size represents a weaker correlation. CP, crude protein; NDF, neutral detergent fiber; ADG, average daily gain; BHBA, β-hydroxybutyrate.

The divergence of the rumen bacterial communities during the pre-weaning period of Hu lambs in the STA and S-ALF groups (**Figure 1** and **Table 2**), together with the significant correlations between bacteria and CP and NDF intake (**Figure 4**), further confirmed that changes in the rumen microbiota could in response to the intake of solid feed. Li et al. (2012) also showed that solid feed intake distinctly altered the rumen microbial composition.

Only 10–15 genera exhibited significant changes on d38 after solid feed intake (**Figure 1**). How can these microbes play such an important function in a redundant ruminal microbial environment? It was previously noted that significant microbial compositional changes may not lead to a functional shift because many microbes share the same metabolic pathways. Li et al. (2012) observed that all of the functional classes between two age groups (d14 and d42 of calves) were similar, suggesting that although their phylogenetic composition greatly fluctuated, the rumen microbial communities of pre-ruminant calves maintained a stable function and metabolic potentials. In our current study, the abundances of the genera *Prevotella* and *Butyrivibrio* increased the most, with increases in *Succinivibrio*

and *Bifidobacterium* also observed at d38 when compared to the B-10 lambs. The members of the *Prevotella* genus are highly amylolytic and proteolytic (Matsui et al., 2000; Xu and Gordon, 2003). The genus *Butyrivibrio* represents the primary butyrate producers in the rumen and are considered effective hemicellulose degraders (Diez-Gonzalez et al., 1999; Paillard et al., 2007), and several species in *Butyrivibrio* are also responsible for their high proteolytic activity (Cotta and Hespell, 1986; Attwood and Reilly, 1995; Sales et al., 2000). *Succinivibrio* and *Bifidobacterium* are saccharolytic bacteria and can produce acetate and lactate (Bryant, 2015; Biavati and Mattarelli, 2015). The higher abundances of these genera suggests the potential increase in carbohydrate and protein metabolism in the rumen with increased starter intake. In addition, the increased concentration of ruminal butyrate before weaning was consistent with the relative abundance of *Butyrivibrio* (Yang et al., 2015). However, the predicted functions of the carbohydrate and amino acid metabolism by PICRUSt were not significantly changed during the pre-weaning period. Although PICRUSt has been demonstrated to be a useful tool to predict the function of microbiota from various environments based on 16S rRNA gene sequences (Langille et al., 2013), this tool was developed based on Greengenes (DeSantis et al., 2006) and IMG (the integrated microbial genomes database and comparative analysis system, Markowitz et al., 2011). Due to the nature of the rumen microbiota, many of the functions of unclassified bacteria may be underestimated. To further understand the impact of increased bacterial taxa on rumen function, functional analysis of the rumen microbiome using metagenomic and/or metabolomic analyses should be integrated.

The fibrolytic bacteria of the taxa *Lachnospiraceae* (Biddle et al., 2013) and *Treponema* (Ziołecki, 1979) were only increased in the S-ALF group at d38. The relative abundances of some saccharolytic bacteria (such as family *Bacteroidaceae*) (Song et al., 2015), short-chain fatty acid producers (such as *Coprococcus* and *Actinomyces*) (Buchanan and Pine, 1962; Tsai et al., 1976), proteolytic fermenters (such as *Fusobacterium*) (Takahashi, 2003), and fibrolytic bacteria of unclassified *Ruminococcaceae* (Brulc et al., 2009; Biddle et al., 2013; Nyonyo et al., 2014) were significantly decreased in the STA lambs but not in the S-ALF lambs before weaning (**Figures 1B,D**). These results occurred because of the bigger variation among individual S-ALF lambs than was observed in the STA group, likely as an effect of alfalfa intervention. After weaning and dietary transition, the abundances of *Treponema*, unclassified *Lachnospiraceae*, and unclassified *Ruminococcaceae*, which had higher relative abundances in the S-ALF lambs before weaning, began to increase in the STA group (**Figure 2B**). It was previously reported that *Treponema* were closely associated with pectin-rich treatments due to the ability of species of this genus to degrade pectin (Liu et al., 2015). Alfalfa contains a pectin content of 10.5–14.2% (Mertens, 2003), which is more than grass and corn stover (Waite and Gorrod, 1959; Mullen et al., 2010). Therefore, in the present study, the increase in *Treponema* was a response to alfalfa intake, suggesting that the intervention of alfalfa had an effect on the rumen microbial composition. After weaning, when both groups were exposed to alfalfa, the microbial composition in the STA group began to approach that of the S-ALF lambs. Furthermore, after weaning, the relative abundances of unclassified *Clostridiales* were higher in S-ALF lambs than in the STA lambs, and *Ruminococcus* was only increased in S-ALF lambs. The order *Clostridiales* includes many polysaccharolytic bacteria that contribute to the production of VFAs in the gut (Chinda et al., 2004). Some *Ruminococcus* strains are cellulolytic fiber-degrading bacteria (Ezaki, 2015) with cellulosome systems (Ben David et al., 2015). Members of this genus are all organic acid-producing bacteria relating to fiber digestion. These results suggested the positive effects of alfalfa supplementation on microbial changes and might improve rumen digestion. The microbial changes that occurred in response to weaning transition and dietary changes have been previously reported (Li et al., 2012; Jami et al., 2013; Oikonomou et al., 2013; Meale et al., 2016), but these studies used dairy calves. Studies by Oikonomou et al. (2013) and Meale et al. (2016) focused on fecal microbiota, and the other two studies only reported rumen microbial changes at the phylum level, and none of these studies compared post-weaning changes after dietary interventions. Therefore, our study is a more comprehensive assessment of rumen microbial colonization, taking into account the adaptation of the microbiota to weaning transition and dietary changes and how the dietary intervention can potentially manipulate this process.

Butyrate was shown to be an important regulator and stimulator of development of the rumen (Górka et al., 2011) and small intestine (Guilloteau et al., 2009) in calves. Members of the genus *Butyrivibrio* represents the primary butyrate producers in the rumen (Bryant, 1986) and a positive relationship between *Butyrivibrio* and papillae length in the rumen was observed in our correlation analysis (**Figure 4**). Plasma BHBA is a parameter that is associated with the physical development of the rumen (Quigley et al., 1991; Khan et al., 2011). The positive correlation between the unclassified *Lachnospiraceae* and unclassified *Clostridiales* and BHBA, the higher abundance of unclassified *Lachnospiraceae* in the S-ALF lambs before weaning, and the significant increase of unclassified *Clostridiales* in S-ALF lambs that occurred shortly after weaning suggested the positive effects of alfalfa supplementation on microbial changes, which may promote rumen physical development. Acetate is a VFA that provide energy for the host through its conversion to ketone bodies (Pennington, 1952). Unclassified *Lachnospiraceae* and unclassified *Clostridiales* are the major producers of acetate (Ezaki, 2015). The concentration of acetate in the rumen did not differ between the STA and S-ALF lambs at each assayed age (Yang et al., 2015), while a higher ($P \leq 0.05$) BHBA concentration in the plasma was observed in the S-ALF lambs after weaning (**Table 3**). Therefore, higher amounts of acetate, which is quickly absorbed by the rumen, were not detected in the rumens of the S-ALF lambs compared with those of the STA lambs. Previous studies on the positive effect of fiber supplementation on the performance of young ruminants (Terré et al., 2013; Mirzaei et al., 2015) were primarily focused on the physical stimulation and/or chemical nutrition of the forage (Beiranvand et al., 2014; Terré et al., 2015), whereas our current study is the first to

investigate the role of the rumen microbiota. Taken together, the results of previous studies and our current study suggest that the supplementation of fibers can influence rumen development through physical, chemical and microbial mechanisms.

The abundances of *Eikenella* and *Campylobacter* were lower in the rumens of the S-ALF lambs before and after weaning, respectively. Bacteria belonging to these genera have been associated with a variety of veterinary diseases (Behling et al., 1979; Humphrey and Beckett, 1987). The growth or persistence of *Campylobacter jejuni* in the rumen was supported by the data of Stanley et al. (1998a,b), especially in the rumens of young animals (Stanley and Jones, 2003). The observed decrease in *Eikenella* and *Campylobacter* indicated the potential impact of the alfalfa intervention on reducing the presence these genera, which could reduce the incidence of diseases and their shedding into the environment.

In summary, alfalfa supplementation with starter stimulates the proliferation of fibrolytic bacteria, including unclassified *Lachnospiraceae* and *Treponema*, and promoted the presence of some saccharolytic bacteria and short-chain fatty acid producers prior to weaning. Positive relationships were observed between unclassified *Lachnospiraceae*, *Treponema* and nutrient intake, ADG, and plasma BHBA, although the causal relationship between the host and microbiota is still unclear. While limitations of our study included that the rumen samples were collected by slaughter and comparisons of the ages and diet effects were not performed from the same lambs, the significantly changes in taxa between the B-10 group and the d38 STA and S-ALF lambs might allow the causal effect relationship of the host and microbiota to be determined. The microbial transplantation of d38 STA or S-ALF microbiotas into B-10 lambs or treating the rumen epithelial cell culture with ultrafiltered rumen fluid from lambs from these groups should provide more direct evidence on whether the identified microbial taxa changes could influence rumen development. Regardless, our findings suggest that after colostrum intake, a milk replacer and *ad libitum* starter pellets supplemented with alfalfa are recommended for the early weaning system to improve young ruminant health and performance.

AUTHOR CONTRIBUTIONS

All of the authors contributed intellectual input and assisted with this study and manuscript. JW and BY designed the study and collected content samples. BY and JQL extracted the rumen content DNA. BY and PW contributed to the data analysis; and BY, JXL, LG, and JW prepared the manuscript. All of the authors have read and approved the manuscript.

FUNDING

This experiment was funded by the National Key Research and Development Program of China (2017YFD0500502), the National Natural Science Foundation of China (31572431), the Fundamental Research Funds for the Central Universities (2014QNA6027), and the China Scholarship Council.

ACKNOWLEDGMENTS

We thank the members of the Institute of Dairy Science and Animal Science Experimental Teaching Center of Zhejiang University for their assistance. Special thanks to Drs. O. Wang, F. Li, and N. Malmuthuge from the University of Alberta for their help with the data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00574/full#supplementary-material

REFERENCES

Abecia, L., Waddams, K. E., Martínez-Fernandez, G., Martín-García, A. I., Ramos-Morales, E., Newbold, C. J., et al. (2014). An antimethanogenic nutritional intervention in early life of ruminants modifies ruminal colonization by Archaea. *Archaea* 2014:841463. doi: 10.1155/2014/841463

Attwood, G. T., and Reilly, K. (1995). Identification of proteolytic rumen bacteria isolated from New Zealand cattle. *J. Appl. Microbiol.* 79, 22–29. doi: 10.1111/j.1365-2672.1995.tb03119.x

Baldwin, R. L., McLeod, K. R., Klotz, J. L., and Heitmann, R. N. (2004). Rumen development, intestinal growth and hepatic metabolism in the pre- and postweaning ruminant. *J. Dairy Sci.* 87, E55–E65. doi: 10.3168/jds.S0022-0302(04)70061-2

Bartol, F. F., Wiley, A. A., Miller, D. J., Silva, A. J., Roberts, K. E., Davolt, M. L., et al. (2013). Lactation biology symposium: lactocrine signaling and developmental programming. *J. Anim. Sci.* 91, 696–705. doi: 10.2527/jas.2012-5764

Behling, U. H., Pham, P. H., and Nowotny, A. (1979). Biological activity of the slime and endotoxin of the periodontopathic organism *Eikenella corrodens*. *Infect. Immun.* 26, 580–584.

Beiranvand, H., Ghorbani, G. R., Khorvash, M., Nabipour, A., Dehghan-Banadaky, M., Homayouni, A., et al. (2014). Interactions of alfalfa hay and sodium propionate on dairy calf performance and rumen development. *J. Dairy Sci.* 97, 2270–2280. doi: 10.3168/jds.2012-6332

Ben David, Y., Dassa, B., Borovok, I., Lamed, R., Koropatkin, N. M., Martens, E. C., et al. (2015). Ruminococcal cellulosome systems from rumen to human. *Environ. Microbiol.* 17, 3407–3426. doi: 10.1111/1462-2920.12868

Biavati, B., and Mattarelli, P. (2015). "*Bifidobacterium*," in *Bergey's Manual of Systematics of Archaea and Bacteria,* ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons). doi: 10.1002/9781118960608.gbm00019

Biddle, A., Stewart, L., Blanchard, J., and Leschine, S. (2013). Untangling the genetic basis of fibrolytic specialization by *Lachnospiraceae* and *Ruminococcaceae* in diverse gut communities. *Diversity* 5, 627–640. doi: 10.3390/d5030627

Brookman, J. L., and Nicholson, M. J. (2005). "Anaerobic fungal populations," in *Methods in Gut Microbial Ecology for Ruminants*, eds H. P. S. Makkar and C. S. McSweeney (Berlin: Springer), 139–150.

Brulc, J. M., Antonopoulos, D. A., Miller, M. E., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1948–1953. doi: 10.1073/pnas.0806191105

Bryant, M. P. (1986). "Genus IV. *Butyrivibrio* Bryant and small 1956, 18, emend. Moore, Johnson and Holdeman 1976, 241," in *Bergey's Manual of Systematic Bacteriology*, Vol. 2, eds P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (Baltimore, MD: Williams & Wilkins), 1376–1379.

Bryant, M. P. (2015). "*Succinivibrio*," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons). doi: 10.1002/9781118960608.gbm01087

Buchanan, B. B., and Pine, L. (1962). Characterization of a propionic acid producing *Actinomycete*, *Actinomyces propionicus, sp. nov*. *Microbiology* 28, 305–323. doi: 10.1099/00221287-28-2-305

Budzynska, M., and Weary, D. M. (2008). Weaning distress in dairy calves: effects of alternative weaning procedures. *Appl. Anim. Behav. Sci.* 112, 33–39. doi: 10.1016/j.applanim.2007.08.004

Cahenzli, J., Köller, Y., Wyss, M., Geuking, M. B., and McCoy, K. D. (2013). Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe* 14, 559–570. doi: 10.1016/j.chom.2013.10.004

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4516–4522. doi: 10.1073/pnas.1000080107

Chinda, D., Nakaji, S., Fukuda, S., Sakamoto, J., Shimoyama, T., Nakamura, T., et al. (2004). The fermentation of different dietary fibers is associated with fecal clostridia levels in men. *J. Nutr.* 134, 1881–1886. doi: 10.1093/jn/134.8.1881

Cotta, M. A., and Hespell, R. B. (1986). Proteolytic activity of the ruminal bacterium *Butyrivibrio fibrisolvens*. *Appl. Environ. Microbiol.* 52, 51–58.

de Passillé, A. M., Borderas, T. F., and Rushen, J. (2011). Weaning age of calves fed a high milk allowance by automated feeders: effects on feed, water, and energy intake, behavioral signs of hunger, and weight gains. *J. Dairy Sci.* 94, 1401–1408. doi: 10.3168/jds.2010-3441

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

Diez-Gonzalez, F., Bond, D. R., Jennings, E., and Russell, J. B. (1999). Alternative schemes of butyrate production in *Butyrivibrio fibrisolvens* and their relationship to acetate utilization, lactate production, and phylogeny. *Arch. Microbiol.* 171, 324–330. doi: 10.1007/s002030050717

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Ezaki, T. (2015). "*Ruminococcus*," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons).

Fisher, R. E., Steele, M., and Karrow, N. A. (2012). Fetal programming of the neuroendocrine-immune system and metabolic disease. *J. Pregnancy* 2012:792934. doi: 10.1155/2012/792934

Gagen, E. J., Denman, S. E., Padmanabha, J., Zadbuke, S., Al Jassim, R., Morrison, M., et al. (2010). Functional gene analysis suggests different acetogen populations in the bovine rumen and tammar wallaby forestomach. *Appl. Environ. Microbiol.* 76, 7785–7795. doi: 10.1128/AEM.01679-10

Górka, P., Kowalski, Z. M., Pietrzak, P., Kotunia, A., Jagusiak, W., Holst, J. J., et al. (2011). Effect of method of delivery of sodium butyrate on rumen development in newborn calves. *J. Dairy Sci.* 94, 5578–5588. doi: 10.3168/jds.2011-4166

Guilloteau, P., Zabielski, R., David, J. C., Blum, J. W., Morisset, J. A., Biernat, M., et al. (2009). Sodium-butyrate as a growth promoter in milk replacer formula for young calves. *J. Dairy Sci.* 92, 1038–1049. doi: 10.3168/jds.2008-1213

Heinrichs, J. (2005). Rumen development in the dairy calf. *Adv. Dairy Technol.* 17, 179–187.

Humphrey, T. J., and Beckett, P. (1987). Campylobacter jejuni in dairy cows and raw milk. *Epidemiol. Infect.* 98, 263–269. doi: 10.1017/S0950268800062014

Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* 7, 1069–1079. doi: 10.1038/ismej.2013.2

Khan, M. A., Weary, D. M., and Von Keyserlingk, M. A. G. (2011). Hay intake improves performance, and rumen development of calves fed higher quantities of milk. *J. Dairy Sci.* 94, 3547–3553. doi: 10.3168/jds.2010-3871

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 8, 1–10. doi: 10.1038/nbt.2676

Li, R. W., Connor, E. E., Li, C., Baldwin, V. I., Ransom, L., and Sparks, M. E. (2012). Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environ. Microbiol.* 14, 129–139. doi: 10.1111/j.1462-2920.2011.02543.x

Liu, J., Pu, Y. Y., Xie, Q., Wang, J. K., and Liu, J. X. (2015). Pectin induces an in vitro rumen microbial population shift attributed to the pectinolytic *Treponema* group. *Curr. Microbiol.* 70, 67–74. doi: 10.1007/s00284-014-0672-y

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2011). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/gkr1044

Matsui, H., Ogata, K., Tajima, K., Nakamura, M., Nagamine, T., Aminov, R. I., et al. (2000). Phenotypic characterization of polysaccharidases produced by four Prevotella type strains. *Curr. Microbiol.* 41, 45–49. doi: 10.1007/s00284001089

Meale, S. J., Li, S., Azevedo, P., Derakhshani, H., Plaizier, J. C., Khafipour, E., et al. (2016). Development of ruminal and fecal microbiomes are affected by weaning but not weaning strategy in dairy calves. *Front. Microbiol.* 7:582. doi: 10.3389/fmicb.2016.00582

Mertens, D. (2003). "Nutritional implications of fiber and carbohydrate characteristics of corn silage and alfalfa hay," in *Proceedings of the California Animal Nutrition Conference*, Fresno, CA, 94–107.

Mirzaei, M., Khorvash, M., Ghorbani, G. R., Kazemi-Bonchenari, M., Riasi, A., Nabipour, A., et al. (2015). Effects of supplementation level and particle size of alfalfa hay on growth characteristics and rumen development in dairy calves. *J. Anim. Physiol. Anim. Nutr.* 99, 553–564. doi: 10.1111/jpn.12229

Mullen, C. A., Boateng, A. A., Goldberg, N. M., Lima, I. M., Laird, D. A., and Hicks, K. B. (2010). Bio-oil and bio-char production from corn cobs and stover by fast pyrolysis. *Biomass Bioenergy* 34, 67–74. doi: 10.1016/j.biombioe.2009.09.012

Nyonyo, T., Shinkai, T., and Mitsumori, M. (2014). Improved culturability of cellulolytic rumen bacteria and phylogenetic diversity of culturable cellulolytic and xylanolytic bacteria newly isolated from the bovine rumen. *FEMS Microbiol. Ecol.* 88, 528–537. doi: 10.1111/1574-6941.12318

Oikonomou, G., Teixeira, A. G., Foditsch, C., Bicalho, M. L., Machado, V. S., and Bicalho, R. C. (2013). Fecal microbial diversity in pre-weaned dairy calves as described by pyrosequencing of metagenomic 16S rDNA. Associations of Faecalibacterium species with health and growth. *PLoS One* 8:e63157. doi: 10.1371/journal.pone.0063157

Paillard, D., McKain, N., Chaudhary, L. C., Walker, N. D., Pizette, F., Koppova, I., et al. (2007). Relation between phylogenetic position, lipid metabolism and butyrate production by different *Butyrivibrio*-like bacteria from the rumen. *Antonie Van Leeuwenhoek* 91, 417–422. doi: 10.1007/s10482-006-9121-7

Pennington, R. J. (1952). The metabolism of short-chain fatty acids in the sheep. 1. Fatty acid utilization and ketone body production by rumen epithelium and other tissues. *Biochem. J.* 51, 251–258. doi: 10.1042/bj0510251

Quigley, J. D. III, Caldwell, L. A., Sinks, G. D., and Heitmann, R. N. (1991). Changes in blood glucose, nonesterified fatty acids, and ketones in response to weaning and feed intake in young calves. *J. Dairy Sci.* 74, 250–257. doi: 10.3168/jds.S0022-0302(91)78167-8

Rey, M., Enjalbert, F., Combes, S., Cauquil, L., Bouchez, O., and Monteils, V. (2014). Establishment of ruminal bacterial community in dairy calves from birth to weaning is sequential. *J. Appl. Microbiol.* 116, 245–257. doi: 10.1111/jam.12405

Rey, M., Enjalbert, F., and Monteils, V. (2012). Establishment of ruminal enzyme activities and fermentation capacity in dairy calves from birth through weaning. *J. Dairy Sci.* 95, 1500–1512. doi: 10.3168/jds.2011-4902

Sales, M., Lucas, F., and Blanchart, G. (2000). Effects of ammonia and amino acids on the growth and proteolytic activity of three species of rumen bacteria: *Prevotella albensis*, *Butyrivibrio fibrisolvens*, and *Streptococcus bovis*. *Curr. Microbiol.* 40, 380–386. doi: 10.1007/s002840010074

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60

Soberon, F., and Van Amburgh, M. E. (2013). Lactation biology symposium: the effect of nutrient intake from milk or milk replacer of preweaned dairy calves on lactation milk yield as adults: a meta-analysis of current data. *J. Anim. Sci.* 91, 706–712. doi: 10.2527/jas.2012-5834

Song, Y., Liu, C., and Finegold, S. M. (2015). "*Bacteroides*," in *Bergey's Manual of Systematics of Archaea and Bacteria,* ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons). doi: 10.1002/9781118960608.gbm00238

Stanley, K., and Jones, K. (2003). Cattle and sheep farms as reservoirs of *Campylobacter*. *J. Appl. Microbiol.* 94, 104–113. doi: 10.1046/j.1365-2672.94.s1.12.x

Stanley, K. N., Wallace, J. S., Currie, J. E., Diggle, P. J., and Jones, K. (1998a). Seasonal variation of thermophilic campylobacters in lambs at slaughter. *J. Appl. Microbiol.* 84, 1111–1116.

Stanley, K. N., Wallace, J. S., Currie, J. E., Diggle, P. J., and Jones, K. (1998b). The seasonal variation of thermophilic campylobacters in beef cattle, dairy cattle and calves. *J. Appl. Microbiol.* 85, 472–480.

Suárez, B. J., Van Reenen, C. G., Beldman, G., van Delen, J., Dijkstra, J., and Gerrits, W. J. (2006). Effects of supplementing concentrates differing in carbohydrate composition in veal calf diets: I. Animal performance and rumen fermentation characteristics. *J. Dairy Sci.* 89, 4365–4375. doi: 10.3168/jds.S0022-0302(06)72483-3

Takahashi, N. (2003). Acid-neutralizing activity during amino acid fermentation by *Porphyromonas gingivalis*, *Prevotella intermedia* and *Fusobacterium nucleatum*. *Mol. Oral Microbiol.* 18, 109–113. doi: 10.1034/j.1399-302X.2003.00054.x

Tamate, H., McGilliard, A. D., Jacobson, N. L., and Getty, R. (1962). Effect of various dietaries on the anatomical development of the stomach in the Calf. *J. Dairy Sci.* 45, 408–420. doi: 10.3168/jds.S0022-0302(62)89406-5

Terré, M., Castells, L., Khan, M. A., and Bach, A. (2015). Interaction between the physical form of the starter feed and straw provision on growth performance of Holstein calves. *J. Dairy Sci.* 98, 1101–1109. doi: 10.3168/jds.2014-8151

Terré, M., Pedrals, E., Dalmau, A., and Bach, A. (2013). What do preweaned and weaned calves need in the diet: a high fiber content or a forage source? *J. Dairy Sci.* 96, 5217–5225. doi: 10.3168/jds.2012-6304

Tsai, C. G., Gates, D. M., Ingledew, W. M., and Jones, G. A. (1976). Products of anaerobic phloroglucinol degradation by Coprococcus sp. Pe15. *Can. J. Microbiol.* 22, 159–164.

Waite, R., and Gorrod, A. R. N. (1959). The comprehensive analysis of grasses. *J. Sci. Food Agric.* 10, 317–326. doi: 10.1002/jsfa.2740100604

Xu, J., and Gordon, J. I. (2003). Honor thy symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10452–10459. doi: 10.1073/pnas.1734063100

Yang, B., He, B., Wang, S. S., Liu, J. X., and Wang, J. K. (2015). Early supplementation of starter pellets with alfalfa improves the performance of pre- and postweaning Hu lambs. *J. Anim. Sci.* 93, 4984–4994. doi: 10.2527/jas.2015-9266

Ziołecki, A. (1979). Isolation and characterization of large treponemes from the bovine rumen. *Appl. Environ. Microbiol.* 37, 131–135.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Yang, Le, Wu, Liu, Guan and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 August 2018
doi: 10.3389/fmicb.2018.01846

Exploring the Spatial-Temporal Microbiota of Compound Stomachs in a Pre-weaned Goat Model

*Yu Lei†, Ke Zhang†, Mengmeng Guo, Guanwei Li, Chao Li, Bibo Li, Yuxin Yang, Yulin Chen and Xiaolong Wang**

College of Animal Science and Technology, Northwest A&F University, Xianyang, China

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Amlan Kumar Patra, West Bengal University of Animal & Fishery Sciences, India
Mick Watson, University of Edinburgh, United Kingdom

***Correspondence:**
Xiaolong Wang
xiaolongwang@nwafu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *19 January 2018*
Accepted: *24 July 2018*
Published: *15 August 2018*

Citation:
Lei Y, Zhang K, Guo M, Li G, Li C, Li B, Yang Y, Chen Y and Wang X (2018) Exploring the Spatial-Temporal Microbiota of Compound Stomachs in a Pre-weaned Goat Model. Front. Microbiol. 9:1846. doi: 10.3389/fmicb.2018.01846

Ruminant animals possess a characteristic four-compartment stomach (rumen, reticulum, omasum, and abomasum) that is specialized for pre-intestinal digestion of plant materials. Of these four compartments, the rumen is the largest. The rumen's diverse microbial community has been well studied. However, the current understanding of microbial profiles in the reticulum, omasum and abomasum are lacking. In the present study, fluid samples from the reticulum, omasum, and abomasum of goats at 3, 7, 14, 21, 28, 42, and 56 days after birth, as well as the negative controls (NC) used for microbial DNA extraction, were subjected to 16S rRNA sequencing. By filtering operational taxonomic units (OTUs) in NC, distinct temporal distributions of microbes were observed in the different compartments, we showed that the OTUs in control samples had a large effect to the samples with low microbial density. In addition, Proteobacteria gradually decreased with age from days 3 to 56 in all three compartments, and the relative abundance of Bacteroidetes increased from 24.15% (Day 3) to 52.03% (Day 56) in abomasum. Network analysis revealed that *Prevotellaceae_UGG-03* and *Rikenellaceae_RC9* were positively correlated with *Prevotella_1*, lending support to the well understood fact that cellulose is well digested in compound stomachs prior to the rumen. Pathway analysis revealed that gene expression in abomasum at Day 3 were primarily related to Glycolysis/Gluconeogenesis and Pyruvate metabolism, suggesting that colostrum digestion is the dominant function of the abomasum at an early age. These findings combined with other recent rumen microbiota data show that the microbiome landscape represents three distinct stages in ruminant stomachs. The first stage is to gain access to external microorganisms at Day 0–14, the secondary stage is for microbial transition at Day 14–28, and the third stage is for exogenous and endogenous microbial colonization beyond Day 28 of age. Our results provide insight into microbiota dynamics in ruminant stomachs, and will facilitate efforts for the maintenance of gastrointestinal balance and intervention with starter diets in juvenile ruminants during early development.

Keywords: taxonomic diversity, ruminant, microbial community, rumen microbiology, gut microbiota

INTRODUCTION

Microbiota of mammalian gastrointestinal tracts (GIT) are complex ecosystems which are composed of diverse microbial populations, including bacteria, archaea, ciliate protozoa, and anaerobic fungi (Falony et al., 2016). Maternal animals are subjected to three major physiological stresses, including gestation, parturition, and nursing. Of them, stresses of nursing occur during the pre-weaning stages (e.g., 60 days after birth in small ruminants such as goats and sheep) and is the most important physiological stage for kids. After birth, the main colostrum or mature milk substitute is liquid feed administered along the esophageal ditch directly into the abomasum and intestines of ruminants, which is similar to other monogastric animals (Khan et al., 2011). Studies have demonstrated that colostrum and mature milk are an important source of gut bacterial colonization, accounting for the introduction of >200 species (Toscano et al., 2017). This has a strong impact on the newborn's health for several months after birth (Fernández et al., 2013; Toscano et al., 2017). As solid feed intake increases, the rumen transitions to its mature role in making important physiological nutrients bioavailable. At this stage, the kids have transitioned from non-ruminant to ruminant, and the nutrient digestion and metabolism pathways undergo a qualitative change (Rey et al., 2012). In addition, the rumen and the other stomach compartments (reticulum, omasum, and abomasum) serve as important sites of colonization for many commensal microorganisms (Peng et al., 2015). Although studies of ruminal microorganisms have resulted in an improved understanding of the composition and colonization of rumen microbial communities (Jami et al., 2013; Mao et al., 2015; Morgavi et al., 2015; Wang et al., 2016), the dynamics of microbiota development in the three other stomach compartments has been severely understudied.

Digestion in ruminants is characterized based on each of the four stomach compartments. Three of them are known as the forestomaches, which consist of the rumen, reticulum, and omasum (Jonsson, 2011). The forestomaches contain large numbers of microorganisms that aid in the anaerobic degradation of nutrients. The forestomach microorganisms are capable of breaking bonds in cellulose and hemicellulose, which are major energy sources for ruminants. Once the partially digested feed has passed through the forestomaches, it enters the abomasum, which is similar to the stomach in monogastric animals. Up until approximately 3 weeks of age, the rumen is largely non-functional. Development of the rumen undergoes a process of development spanning roughly the first 6 months of life, at which point it becomes fully functional (Trenkle, 2001). In contrast, the lower gastrointestinal tract can develop a functional microbiome within a few weeks (Morowitz et al., 2010). In addition, fermentation in forestomaches produces a variety of end products which are subsequently absorbed in the digestive tract. Fermentation of a fiber-rich diet is a time-consuming process (Animut and Goetsch, 2008), whereas in hay-fed goats, the rumen plays a major role in the digestion of a fiber-rich diet, but in milk-fed kids, digestion occurs mainly in the abomasum.

The temporal sequence of microbial establishment in ruminant stomachs at early developmental stages holds great promise for the economical rearing of replacements, as well as for the host's well-being. We have previously investigated microbial communities in the rumen of weaned (Han et al., 2015) and pre-weaned goats (Zhang et al., 2018, unpublished). We observed that the colonization of microorganisms is highly correlated to the function in the rumen. The composition of the primary bacterial communities is determined and acquired shortly after birth not only in the rumen, but also in the other stomach compartments of ruminants. The spatial-temporal patterns occurring in these communities remain largely unknown. In this study, 16S rRNA sequencing was used to investigate the colonization of reticulum, omasum, and abomasum microbes in pre-weaned goats as a ruminant model. We sought to determine which developmental stages have significant effects on microbiota communities in the different stomach compartments. This study provides new insights into bacteria communities in pre-weaned goats, which may be useful in designing strategies to promote colonization of target communities.

MATERIALS AND METHODS

Animal Handling and Sample Collection

All sampling of animals was approved by the Institutional Animal Care and Use Committee of the Northwest A&F University under permit number 2014ZX08008002. All surgeries were performed while animals were anesthetized with xylazine chlorhydrate. All efforts were made to minimize animal suffering.

Two months prior to sample collection, pregnant does were raised at the experimental facilities of the Shaanbei Cashmere Goat Farm (Hengshan, Shaanxi). After delivery, the single kids were housed together with their mothers for nursing. The mother's milk was the sole food until Day 25. Between Day 25 and Day 56, kids were provided granule feed and high-quality alfalfa twice daily at 09:00 and 18:00 (not limit the feed intake of the kid during feeding process). The ingredients and nutrient composition of the diets are summarized in **Supplementary Table S1**. Fresh water was provided for *ad libitum* consumption throughout the experimental period. Three kids were slaughtered at each experimental time point (Day 3, Day 7, Day 14, Day 21, Day 28, Day 42, Day 56). The reticulum, omasum, and abomasum contents were collected into 20 mL cryopreservation tubes. The sampled liquid was and immediately stored at −80°C for further analysis as previously reported (Zhang et al., 2018, unpublished).

DNA Extraction, PCR Amplification, and Sequencing

Microbial DNA was extracted from reticulum, omasum, and abomasum fluid samples using the Fast-DNA® Spin kit for soil (Bio101, Vista, CA, United States) according to manufacturer's instructions. During the study, we strictly controlled the test samples and negative controls (NC) were set during the DNA extraction and amplification process. To ensure the fluid samples were not contaminated by reagents used for DNA extraction, two commercial kits to isolate DNA from fluid and NCs

(ddH_2O and reagents used for DNA extraction) were used (**Supplementary Figure S1**), these extracted DNA were further amplified (**Supplementary Figure S1**). These results collectively indicated we found that the stomach fluid DNA samples were not contaminated during the DNA extraction process, thereby confirming the reliability of samples used in this study (**Supplementary Figure S1**). We were also unable to extract DNA from the Day 0 samples in reticulum due to inadequate fluids. Additionally, the NC, DNA-free water and buffer used for DNA extraction, were subjected to paired-end sequencing (2 × 250/bp) using the Illumina HiSeq2500 platform. The R package "decontam"[1] was used to identify contaminants in the metagenomic sequencing data (Davis et al., 2017). In addition, the V3-V4 hypervariable region of the bacteria were amplified by PCR using the following thermocycling protocol: 95°C for 2 min, followed by 30 cycles at 95°C for 20 s, 55°C for 30 s, and 72°C for 30 s, followed by a final extension at 72°C for 5 min. The primers used to amplify the sequences were as follows: V341F (5′-CCTAYGGGRBGCASCAG-3′) and V806R, (5′-GGACTACHVGGGTWTCTAAT-3′). A molecular barcode in the form of an 8-base sequence unique to each sample was added for subsequent identification and processing. PCR reactions were performed in triplicate 20 μL reactions containing 5 μL of 5× FastPfu Buffer, 2 μL of 2.5 mM dNTPs, 0.8 μL of each primer (5 μM), 0.4 μL of FastPfu Polymerase, and 10 ng of template DNA. The resulting amplicons were resolved on 2% agarose gels. The resulting bands were extracted and purified from the gels using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, United States) according to the manufacturer's instructions. The resulting products were quantified using the QuantiFluor^{TM} -ST kit (Promega, United States). Sample libraries were pooled in equimolar ratios, and were subjected to paired-end sequencing (2 × 250/bp) using the Illumina HiSeq platform according to a standard protocol at the Mega Genomics Company Limited, Beijing, China. Amplicon sequences generated in this study are available at the NCBI-SRA under accession: SRP090491.

Data Analysis

Raw fastq files were de-multiplexed, and quality-filtered using QIIME (version 1.9.1) (Caporaso et al., 2010) with the following criteria: 300 bp reads were truncated at any site with an average quality score <20 over a 50 bp sliding window; truncated reads shorter than 50 bp were discarded; exact barcode matching; two nucleotide mismatch in primer matching; reads containing ambiguous characters were removed; sequences with at least a 10 bp overlap were assembled according to their overlap sequence. Reads which could not be assembled were discarded. Based on the overlapping sequences, paired-reads were merged in to a single read. The merged reads were then used for operational taxonomic unit (OTU) clustering, taxonomic classification, and community diversity assessment. The resulting microbial communities were used for comparison of similarity or dissimilarity between different sample groups, analysis of the relationships between microbial communities and environmental factors, phylogenetic analysis, as well as other statistical analyses. In this study, total 4,481,828 raw tags were obtained, among them 4,106,574 effective tags were used for further analysis.

For phylogenetic analysis, OTUs were clustered with a cutoff of at least 97% similarity using UPARSE (version 7.1[2]) (Blaalid et al., 2013; Edgar, 2013). Chimeric sequences were identified and removed using UCHIME. In ecology, alpha diversity (α-diversity) is the mean species diversity in sites or habitats on a local scale. These OTUs were used for the determination of α-diversity (Shannon and Simpson) and richness (Ace and Chao) (Schloss et al., 2011; Quast et al., 2012; Wang et al., 2012; Amato et al., 2013). Principal component analysis (PCA), PLS-DA (Partial Least Squares Discriminant Analysis) and species correlation network were analyzed on the free online platform of Major bio I-Sanger Cloud Platform[3].

The taxonomy of each 16S rRNA gene sequence was analyzed using RDP Classifier[4] against the Silva (Cole et al., 2008) (SSU123) 16S rRNA database using a confidence threshold of 0.7. Functional gene annotations were predicted using FGR (Release7.3[5]) (Cole et al., 2013). Based on the relative abundance of the microorganisms, those in high abundance or those that were artificially selected in the sample comparison between two or more groups were subjected to a Welch's test to determine statistical significance using the R "gplots package." Significant differences between taxa were assessed by Tukey' test comparison procedure using *SPSS* version 22.0 for Windows.

Prediction of COG Function Classification and Functional KEGG Pathways

PICRUSt (Langille et al., 2013) was used to normalize the OTU table by the 16S rRNA copy number predictions. Consequently, the OTU abundances more accurately reflects that of the organisms in the population. Metagenomic predictions were then made by looking up the re-calculated genome content for each OTU, and then multiplying the normalized OTU abundance by each KEGG ontology (KO) abundance in the genome, and summing the KO abundances for each sample. The resulting predictions yielded a table of KO abundances for each metronome sample in the OTU table (White et al., 2009; Segata et al., 2011).

RESULTS

Diversity, Richness, and Composition of the Bacterial Communities in Compound Stomachs

A total of 67 samples, representing the reticulum, omasum, and abomasum from 21 goats (n = 3) and NC including

[1]https://github.com/benjjneb/decontam

[2]http://drive5.com/uparse/

[3]www.i-sanger.com

[4]http://rdp.cme.msu.edu/

[5]http://fungene.cme.msu.edu/

TABLE 1 | Alpha diversities within each age group in bacteria, richness, and evenness were measured by the Chao1 evenness index.

Time	Reticulum	Omasum	Abomasum
Day 3	433.78^{a}	362.28^{a}	321.26^{a}
Day 7	525.27^{ab}	528.28^{ab}	392.21^{a}
Day 14	805.07^{bc}	659.58^{bc}	503.03^{a}
Day 21	859.2^{d}	871.67^{de}	869.06^{b}
Day 28	803.96^{d}	982.83^{e}	785.42^{b}
Day 42	637.20^{cd}	737.06^{cd}	724.02^{b}
Day 56	724.38^{d}	822.95^{cde}	845.39^{b}

Tukey' test for Chao index of OTU level in reticulum, omasum, and abomasum. a, b, c, d, e Means with different superscripts differ significantly each column.

DNA-free water and buffer (n = 4) were used for DNA extraction. Sequencing of the16S rRNA samples from these stomach liquids generated 4,181,452 clean tags with an average of 66,372 clean tags per sample (**Supplementary Table S2**), and base quality values with a Q30 ratio greater than 83.16% (**Supplementary Table S2**). Four NC samples were sequenced to generate 105 OTUs (**Supplementary Table S3**), the 16S rRNA samples from NC generated 16,939 clean tags with an average of 4,234 clean tags per sample. The NC sequences were eliminated for subsequent analysis with the R package "decontam" (see text footnote 1). The result further demonstrates that it is critical to pay assiduous attention to controls, when sequencing samples were in a low microbial density. A rarefaction analysis including all samples revealed a curve approaching saturation (**Supplementary Figure S2**), and the Chao1 index indicated that approximately 98% of microbial genes were captured in the samples. The Chao1 index of the OTU level indicated that there was a significant difference in the α diversity index between the various age groups, except for Day 3 and Day 7 (**Table 1**). These data suggest that the communities exhibited a higher α diversity with increasing age.

Analysis of Bacterial Composition in Reticulum

From days 3 to 56, the proportions of Proteobacteria gradually decreased with age (P = 0.039) at the phylum level. The majority of annotated reads (45.46%) belonged to Bacteroidetes, followed by Proteobacteria (27.29%) at Day 3. At Day 7, Bacteroidetes (57.97%) comprised the dominant phyla, followed by Firmicutes (24.47%). However, at D56, Firmicutes (46.74%) and Bacteroidetes (42.29%) accounted for the majority of annotated reads (**Figure 1A** and **Supplementary Table S4**). At the genus level, *Mannheimia*, *Bacteroides*, *Fusobacterium,* and *Porphyromonas* were detected as the dominant genera at Day 3, but the ratios of these genera gradually decreased with increasing age (**Table 2**). On the contrary, *Prevotella_1*, *Rikenellaceae_RC9*, *Ruminococcus_2*, *Bacteroidales_S24-7* were relatively low in samples collected from the Day 3 group but became the most abundant genera in samples collected after Day 28 (**Table 2** and **Supplementary Table S5**). In addition, the annotated reads (about 1%) belonged to *Ruminococcaceae_UCG-005* at days 21 and 28, but were essentially absent at other ages (**Table 2**). Therefore, the microbiota community in reticulum data demonstrates a strong temporal specificity.

Analysis of Bacterial Composition in Omasum

It was observed that in omasum, the prokaryotic communities were dominated by the phyla Bacteroidetes (33.20%), Proteobacteria (15.62%) at the phylum level on Day 3. Interestingly, the Bacteroidetes (54.51%) had increased by Day 7, becoming the most abundant phylum in the samples collected after Day 7 (**Figure 1A** and **Supplementary Table S4**). At days 3 through 56, the populations of Proteobacteria exhibited a gradual but decrease with age. By contrast, the higher annotated reads (21.15% and 1.72%) belonged to Fusobacteria at days 3 and 7, respectively (**Figure 1A**). At the genus level, the majority of the reads (21.15% and 17.75%) were classified as *Fusobacterium* and *Bacteroide* at Day 3, *Prevotella_1 Rikenellaceae_RC9* and *Bacteroidales_S24-7* were observed to exhibit similar changes in the reticulum and omasum (**Supplementary Table S5**). The proportions of *Christensenellaceae_R-7* at Day 28 were higher than at all other ages ($P < 0.05$) (**Table 2**). The proportions of *Ruminococcaceae_NK4A214_group* at Day 14–28 were higher than that at all other stages (P = 0.0123) (**Table 2**). Taken together, these data demonstrate the heterogeneous nature of the omasum microbiota at different developmental stages in terms of genus-level composition.

Analysis of Bacterial Population Composition in Abomasum

At the phylum level, most of the annotated reads (62.10%) belonged to Firmicutes, followed by Bacteroidetes (24.15%) at Day 3. As the ages of the animals increased, the predominant phylum became Bacteroidetes (**Figure 1A**). Between days 3 and 56, the observed low abundance of the phylum Fusobacteria gradually decreased with age (**Supplementary Figure S3**). In contrast, Spirochaetae gradually increased with age. The proportion of Verrucomicrobia at Day 28 was greater than at all other ages (**Supplementary Figure S3**). At the genus level, *Bacteroides*, *Porphyromonas*, *Prevotella_1*, *Rikenellaceae_RC9* were observed to follow similar patterns of population dynamics in the reticulum (**Supplementary Figure S4**). Extraordinary, the proportion of *Conchiformibius* was observed to account for 3.57% at Day 3, whereas after Day 3, it only accounted for less than 0.34% (**Table 2**). The proportion of *Lactobacillus* (52.87%) at Day 3 was greater than at all other ages. The proportion of *Ruminococcaceae_NK4A214* at Day 14 was higher than at all other ages ($P < 0.001$) (**Table 2**). In ruminants, the abomasum is the last compartment of the multi-chambered stomach and serves as the gastric stomach where acid is secreted, and digestion as it occurs in non-ruminant mammalian species, begins. These data indicate that the primary role of *Lactobacillus* is to reduce the pH of the abomasum, allowing abomasum to function as a sort of barrier for bacterial transmission to the lower gastrointestinal tract.

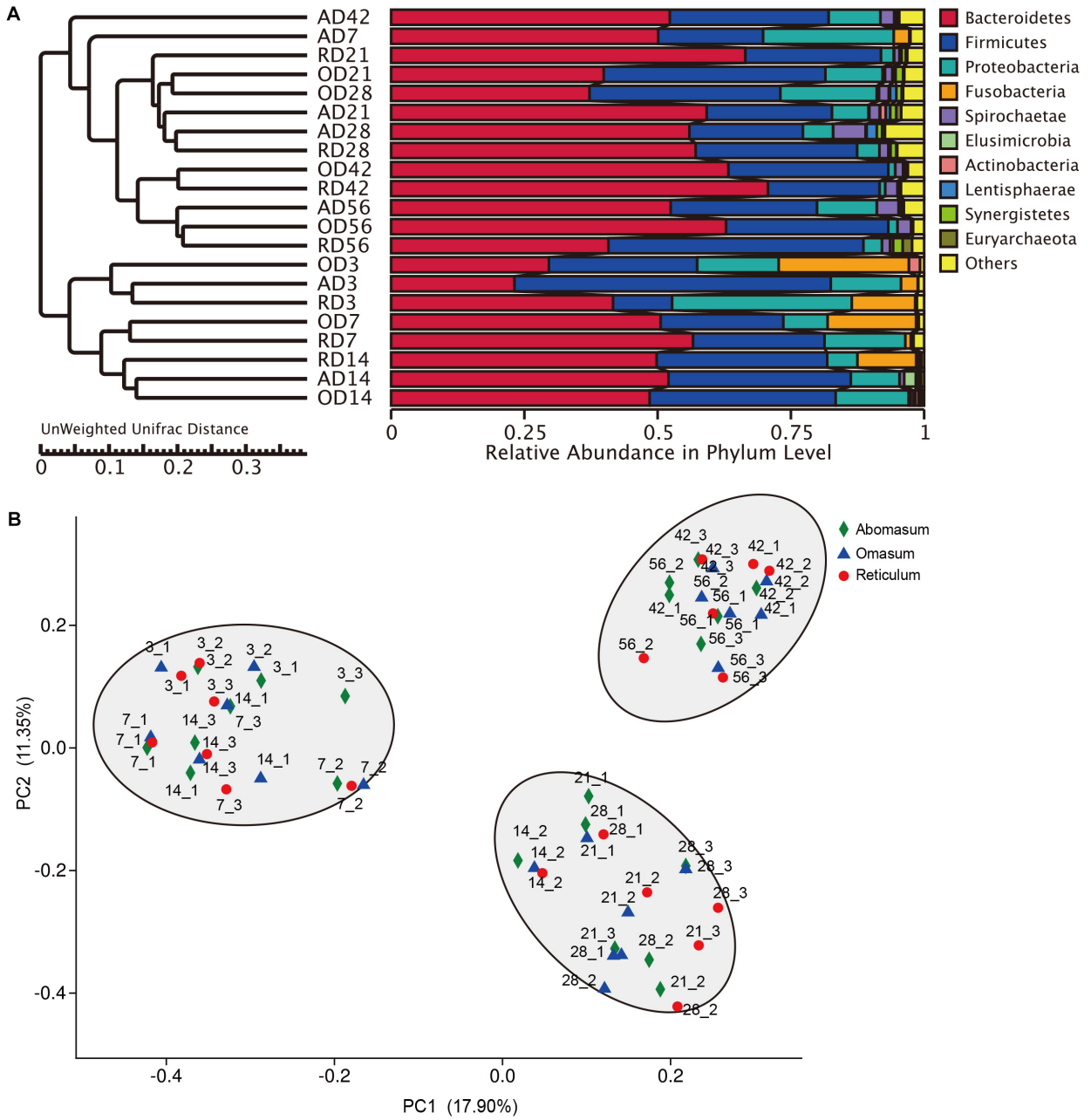


FIGURE 1 | Distribution of bacterial composition in goat compound stomachs. **(A)** UPGMA (Unweighted Pair-group Method with Arithmetic Mean) analysis of the similarities between different samples, and cluster analysis was used to construct the clustering tree of samples. the UPGMA cluster tree based on the Unweighted Unifrac distance. **(B)** Principal coordinate analysis (PCoA) profile of microbial diversity across all samples using an unweighted UniFrac metric, the percentage of variation explained by PC1 and PC2 are indicated in the axis.

OTU Diversity and Similarity Analysis

Community OTU comparisons by principal co-ordinates analysis (PCoA) of each group using the Bray–Curtis similarity metric revealed that the bacterial populations in each sample were best clustered together according to the age of the hosts, suggesting each group hosts its own distinct bacterial community. The average within-group similarity analysis showed a significant difference between the groups, which increased in an age-dependent manner (**Figure 1B**). Our results also revealed a sub-clustering within the Day 3–14 group, Day 21–28 group, and Day 42–56 group, the community structure was similar among the different of stomach compartments at the same age group (**Figure 1B**). The proportion of Fusobacteria in abomasum was lower than that in reticulum and omasum (**Supplementary Figure S5**). The function of Fusobacteria is mainly cellulose decomposition. At the family level in

TABLE 2 | Phylum and genus level composition in goat compound stomachs.

Tissue	Species	D3 (%)	D7 (%)	D14 (%)	D21 (%)	D28 (%)	D42 (%)	D56 (%)	*P*-value
Reticulum	p_Proteobacteria	27.290^{a}	15.250^{ab}	5.523^{b}	2.915^{b}	4.162^{b}	1.034^{b}	3.688^{b}	0.039
	g_Porphyromonas	14.840^{a}	8.238^{a}	13.210^{a}	6.339^{a}	0.226^{b}	0.016^{b}	0.010^{b}	0.003
	g_Rikenellaceae_RC9	0.005^{a}	0.364^{a}	0.446^{a}	14.780^{b}	7.339^{ab}	3.240^{a}	6.662^{ab}	0.003
	g_Ruminococcus_1	0.005^{a}	0.844^{a}	0.523^{a}	0.373^{a}	1.255^{b}	1.643^{b}	2.570^{b}	0.003
	g_Desulfovibrio	0.015^{a}	0.017^{a}	0.087^{a}	0.464^{a}	0.218*a*	0.252^{a}	2.089^{b}	<0.001
	g_Ruminococcaceae_UCG-005	0.002^{a}	0.167^{ab}	0.028^{a}	1.444^{b}	1.175^{bc}	0.108^{ac}	0.464^{abc}	<0.001
Omasum	*g_Prevotella_1*	0.012^{a}	0.062^{a}	1.125^{a}	1.794^{a}	5.493^{a}	27.430^{b}	19.580^{c}	0.001
	g_Christensenellaceae_R-7	0.077^{a}	1.083^{a}	4.730^{ab}	9.158^{b}	5.712^{ab}	1.323^{a}	0.980^{a}	0.009
	g_Ruminococcaceae_NK4A214	0.051^{a}	2.487^{ad}	8.075^{b}	7.405^{c}	5.804^{d}	0.994^{a}	0.937^{a}	0.012
	g_Prevotellaceae_UCG-001	0.003	2.529	2.095	2.443	0.457	4.440	8.60	0.013
Abomasum	*g_Ruminococcaceae_NK4A214*	0.162^{a}	2.437^{a}	10.590^{b}	4.933^{b}	4.278^{b}	1.667^{a}	1.890^{a}	<0.001
	g_Rikenellaceae_RC9_gut	0.003^{a}	0.218^{a}	0.249^{a}	6.488^{b}	5.511^{b}	2.110^{a}	3.010^{a}	0.001
	g_Prevotella_1	0.008^{a}	0.044^{a}	0.625^{a}	7.856^{a}	9.909^{ac}	25.940^{b}	24.580^{c}	<0.001
	g_SP3-e08	0.001^{ac}	0.003^{ac}	0.004^{ac}	7.566^{b}	2.576^{c}	0.152^{ac}	0.292^{ac}	0.005
	g_Conchiformibius	3.570^{a}	0.340^{b}	0.210^{b}	0.060^{b}	0.001^{b}	0.001^{b}	0^{b}	<0.001

The bacterial relative abundance values were analyzed using Tukey' test. a, b, c Means with different superscripts differ significantly each row.

TABLE 3 | Distribution differences at the family and genus level in goat compound stomachs.

Species	Reticulum (%)	Omasum (%)	Abomasum (%)	*P*-value
f_Christensenellaceae	2.118^{ab}	3.490^{a}	0.952^{b}	0.004
f_Lactobacillaceae	0.760^{a}	1.720^{a}	11.350^{b}	0.040
f_Acidaminococcaceae	1.486^{ab}	2.143^{a}	0.827^{b}	0.010
g_Clostridiales_vadinBB60	0.265^{b}	0.068^{b}	0.962^{a}	0.003
g_Christensenellaceae_R-7	2.096^{ab}	3.456^{b}	0.947^{a}	0.004
g_Lactobacillus	0.760^{b}	1.721^{b}	11.350^{a}	0.040
g_Erysipelotrichaceae	0.285^{b}	0.228^{ab}	0.041^{a}	0.002
g_Jeotgalicoccus	0.001^{b}	0.002^{b}	0.010^{a}	0.011
g_Anaerotruncus	0.054^{b}	0.050^{b}	0.009^{a}	0.001
g_Succiniclasticum	1.249^{ab}	1.829^{b}	0.705^{a}	0.038
g_Streptococcus	0.371^{ab}	0.682^{b}	0.096^{a}	0.005
g_Lachnospiraceae_UCG-010	0.120^{ab}	0.226^{b}	0.040^{a}	0.003

The bacterial relative abundance values were analyzed using Tukey' test. a, b, c Means with different superscripts differ significantly each row.

the abomasum, the proportion of Lactobacillaceae (11.35%) Clostridiales_vadinBB60_group (0.96%), Staphylococcaceae (0.01%), and Planococcaceae (0.03%) were significantly higher than in both the reticulum and omasum samples (**Table 3** and **Supplementary Figure S6**). In samples collected from the omasum, the proportion of Christensenellaceae (3.49%), Acidaminococcaceae (2.14%), Coriobacteriaceae (0.09%), and Peptococcaceae (0.55%) were significantly higher than in the samples collected from the reticulum and abomasum (**Table 3** and **Supplementary Figure S6**). At the genus level in the abomasum, the proportions of *Lactobacillus* (11.35%), *Clostridiales_vadinBB60* (0.96%), and *Jeotgalicoccus* (0.01%) were significantly higher than in reticulum and omasum samples (**Table 3**). In the omasum, the proportions of *Peptococcaceae* (0.51%), *Christensenellaceae_R-7_group* (3.45%), *Lachnospiraceae_UCG-008* (0.07%), *Lachnospiraceae_UCG-010* (0.22%), and *Anaerotruncus* (0.05%) were significantly higher than in samples collected from either the reticulum or abomasum (**Table 3**).

Within-Network Interactions Mirrored the Bacterial Microbiota Relationships

Network analysis was used to analyze the correlation between species abundance among different samples in this study. This would provide further insights into the mechanism of the formation of phenotypic differences between the samples. The reticulum network contains a node representing *Prevotella_1*. Its neighbors form seven mutually exclusive clusters: two positively correlated with *Prevotella_1*, consisting mainly of members of the *Prevotellaceae_UGG-03* and *Rikenellaceae_RC9*. The other was negatively correlated with *Prevotella_1*, consisting mainly of *Bacteroides*, *Porphyromonas*, *Fusobacterium*, *Conchiformibius*, *Mannheimia* (**Figure 2A**). The omasum network contains a node representing *Porphyromonas*. Its neighbors form six mutually exclusive clusters: three were positively correlated with *Porphyromonas*, consisting mainly of members of the *Escherichia-Shigella*, *Fusobacterium* and *Bacteroides*, and the other negatively correlated with *Porphyromonas*, consisting mainly of members of the *Prevotella_1*, *Rikenellaceae_RC9*,

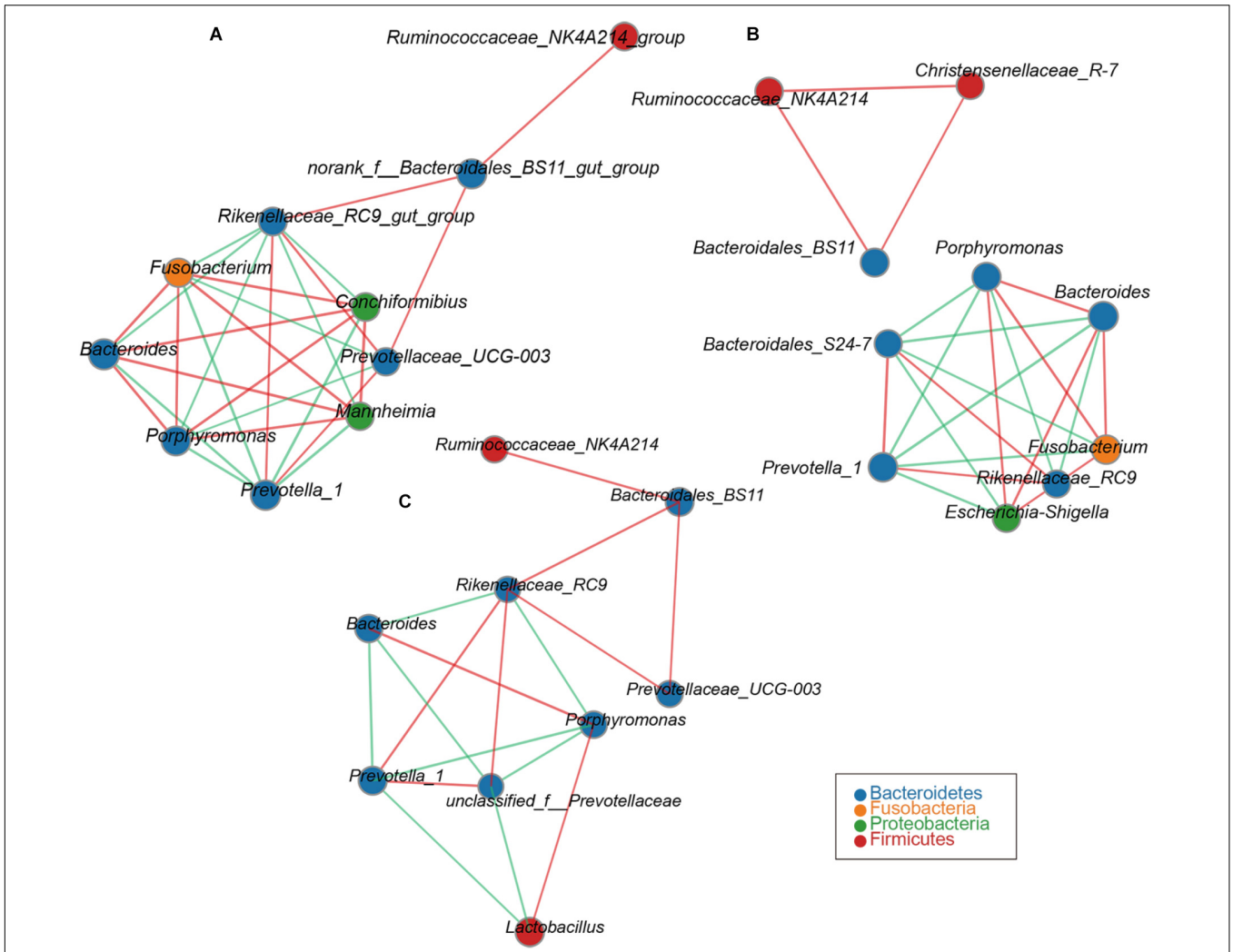


FIGURE 2 | Network analysis applied to the goat compound stomach. Reticulum **(A)**, omasum **(B)**, abomasum **(C)** microbiota at different ages. Genus correlation network maps mainly reflect the genus-relatedness of each taxonomic level under an environmental condition. By calculating the correlation coefficient between genera to reflect the correlation between genera. The size of the node is proportional to the abundance of the genera. Node color corresponds to phylum taxonomic classification. Edge color represents positive (green) and negative (red) correlations, and the edge thickness is equivalent to the correlation values.

and *Bacteroidales_S24-7* (**Figure 2B**). The genus of *Bacteroidales_BS11* was observed to be positively correlated with *Ruminococcaceae_NK4A214* and *Christensenellaceae_R-7* (**Figure 2B**). The network observed within the abomasum contains a node representing *Rikenellaceae_RC9*. Its neighbors form 6 mutually exclusive clusters. Four were positively correlated with *Rikenellaceae_RC9*, consisting members of *Bacteroidales_BS11*, *Prevotellaceae_UGG-03*, *Prevotella_1*, and *Prevotellaceae*. The remaining cluster was negatively correlated with *Rikenellaceae_RC9*, consisting mainly of members of the, *Bacteroides* and *Porphyromonas* (**Figure 2C**).

Predicted Molecular Functions of Bacterial Microbiota

To better understand the molecular functions of the bacterial microbiota across goat compound stomachs, PICRUSt was used to predict likely functions. A total of 278 gene families were identified in the three groups. Of the 278 gene families, most of the genes belonged to Transport (4.97% in reticulum, 5.31% in omasum, 4.96% in abomasum), DNA_repair_and_recombination_proteins (3.07% in reticulum, 3.04% in omasum, 3.15% in abomasum), Ribosome (2.72% in reticulum, 2.64% in omasum, 2.82% in abomasum), Purine metabolism (2.38% in reticulum, 2.36% in omasum, 2.52% in abomasum), and ABC transporters (2.54% in reticulum, 2.75% in omasum, 2.48% in abomasum). A principal component analysis (PCA) on the relative abundance values of the KEGG pathways represented from the different age group microbiota showed a clear distinction (**Figure 3**). Genes related to Glycolysis/Gluconeogenesis, Pyruvate metabolism, Aminoacyl_tRNA_biosynthesis, DNA repair and recombination proteins, Pyrimidine metabolism, and

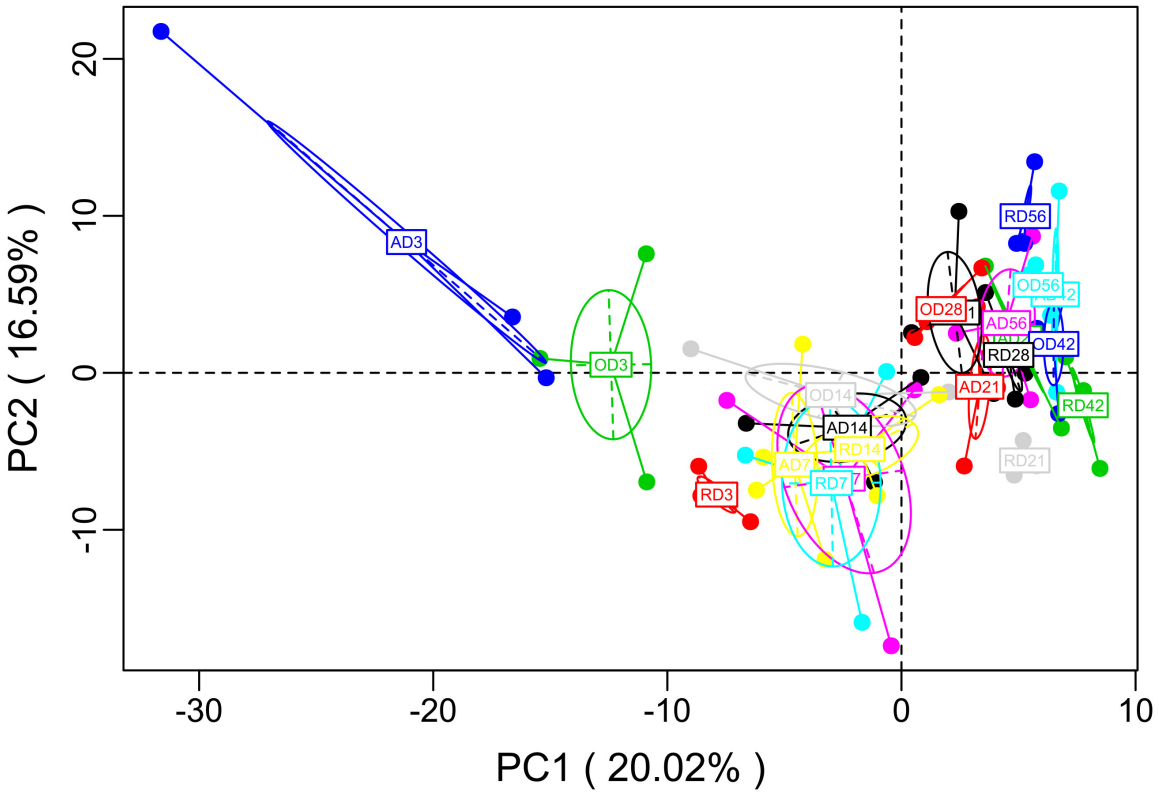


FIGURE 3 | PCA of microbial functional diversity across compound stomach samples. After the dimensionality analysis of the sample, there are relative coordinate points on the principal component, the distance of each sample point represents the distance of the sample, and the sample in the same area in the plane shows the similarity (A: abomasum; O: omasum; R: reticulum).

Purine metabolism were compared among the different age groups, and were observed to be enriched at Day 3 in abomasum (**Supplementary Figure S7**). The abundance of genes related to ABC transporters, Transporters, Transcription factors, and Other ion coupled transporters were enriched at Day 3 in the omasum (**Supplementary Figure S7**). These results suggest that the microbiota at Day 3 is primarily responsible for glucose transport, converting starch and sucrose into glucose and fructose through the enzymatic activities. In addition, the microbiota genes expressed at Day 3 serve to accelerate the conversion between Glycerate-2P and Phosphoenol-pyruvate.

DISCUSSION

In the present study, we sought to characterize the systematic microbiota dynamics in the compound stomachs of ruminants during the transition from birth to weaning, as well as to determine the compositional changes in the colonizing bacterial populations during early development. Our findings suggest that each sampled age group has its own distinct microbiota, which is reflected by the clustering of the samples by age group (**Figure 1**). In general, the microbiota in compound stomachs undergoes developmental changes that are independent of diet in pre-weaned goats. By combing our recent rumen bacterial population composition data (Zhang et al., 2018, unpublished), it was determined that the rumen microbial community exhibited clear spatial differences compared with the other three compartments (**Figure 4**). In contrast to the rumen, the other three stomach compartments exhibited larger temporal differences in the microbiota communities and smaller spatial differences (**Figure 4**).

Previous studies reported that the reagent and laboratory contamination may critically affect 16S rRNA gene sequencing and metagenomic results (Salter et al., 2014). To guarantee the stomach fluid DNA samples were not contaminated by reagents used during the DNA extraction process, NC samples (mainly DNA-free water and buffer used for DNA extraction) were subjected to PCR amplification and sequencing (**Supplementary Figure S1**). Consistent with previous reports (Kennedy et al., 2014; Lauder et al., 2016; Lim et al., 2018), we confirmed the microbial diversity in the NC samples (**Supplementary Table S3**). However, we demonstrate that the microbial diversity in NC only had an effect to the stomach samples with a low microbe density (**Supplementary Table S3**), indicating the microbes in control samples should be considered in metagenomic studies. Based on the observed differences between the newborn animals (Day 3–7) and their low similarity when compared with other age groups (**Figure 2**), we propose that only a few genera are

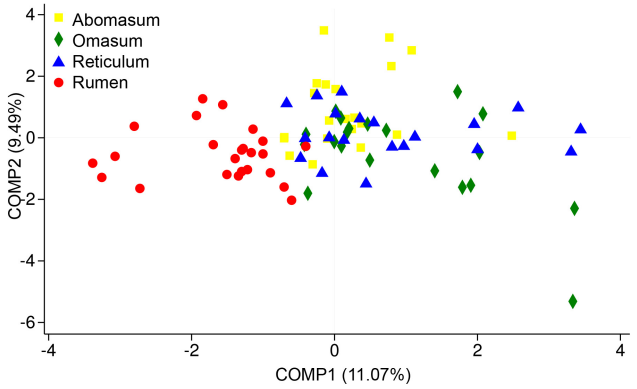


FIGURE 4 | Partial least squares discriminant analysis (PLS-DA) scores plots of the microbiota in four stomachs based on sequencing at the phylum level. PLS-DA is a partial least squares regression of a set of Y of binary variables describing the categories of a categorical variable on a set X of predictor variables. It is a compromise between the usual discriminant analysis and a discriminant analysis on the significant principal components of the predictor variables.

shared between the microbiota communities during the primary stages of colonization and those found in mature animals. This could in part be due to the fact that a pre-functioning rumen is unavailable to digest plant mass during first days of life (Jami et al., 2013). The diversity and richness of OTU numbers increased with the age. This could be explained by the animal's environment and colostrum intake being contributing factors to the varied diversity of the gastrointestinal microbiota. There is increasing evidence that milk is essential for the initial development of newborns, as it represents a great source of commensal bacteria (Khodayar-Pardo et al., 2014). It is assumed that mature milk is the main postnatal source of bacteria for the infant's intestines, and thus serves an important role in microbiota colonization after birth (Martín et al., 2003). The data presented here show that of the microbiota flora in the four stomachs is age specific from the 1st week after birth, with Proteobacteria and Fusobacteria detected as the dominant phyla (**Figure 1**).

Previous studies have shown the microbiota in the neonatal gut is of particular interest, because it reflects not only the fragile structure of the bacterial communities, but also the true origin of mammalian gut microbiota. Bacterial communities in the neonatal gut are unstable due to the nature of the system's rapid temporal variations. Due to the unique abundance of oxygen in the neonatal gut, the microbiota during the first week of life is frequently dominated by facultative anaerobes, mainly Proteobacteria species (Guaraldi and Salvatori, 2012). Other studies have reported that Proteobacteria in the neonatal gut may be derived from the maternal placenta through fetal swallowing of amniotic fluid *in utero* (Koren et al., 2012; Shin et al., 2015). In contrast, beyond the 28th day of life, the gut microbiota is composed almost exclusively of the genus *Prevotella* (**Table 1**). The most abundant bacteria in goats belong to two phyla, Firmicutes and Bacteroidetes. Of the members in Bacteroidetes, two genera dominate-*Bacteroides* and *Prevotella*. *Prevotella* are more common in ruminants who consume a fiber-rich diet (Kovatcheva-Datchary et al., 2015; Ley, 2016). A previous report demonstrated a positive effect on host metabolism by the *Prevotella*-dominated gut microbiota (Kovatcheva-Datchary et al., 2015). In fact, *Prevotella* is capable of metabolizing dietary fiber from plant cell walls, and thus produces significant amounts of short chain fatty acids (SCFAs) that are later absorbed by the host (Ramayo-Caldas et al., 2016). Previous study found that *Prevotella* and *Clostridium* have significantly different cellulose degrading modes, *Prevotella* would preferentially degrade hemicellulose, and when xylan compounds then become accessible, colonization of the fiber surface is gradually taken over by *Clostridium* (Rubino et al., 2017). Therefore, the rapid increase of *Prevotella* in the gastrointestinal tract has a positive effect for the animal's physiological development. Similarities in the composition of the microbiota were observed between the segments of the four stomach compartments. With increasing age, the same patterns of flora colonization were observed in the reticulum, omasum, and abomasum. This is in agreement with previous reports in which the microbiota within the four stomachs as the digesta passes from one segment to another (Mao et al., 2015; Wang et al., 2017). Lactobacillaceae revealed a higher proportion (11.35%) in abomasum (**Supplementary Figure S6**), whereas Lactobacillaceae accounted for a higher proportion (9.80%) in rumen at Day 56. It has been well established that kids mainly digest in the abomasum, and the rumen begins to share the digestive function until Day 28. Therefore, Lactobacillales may be more abundant in the main digestive organs of the body. An abundance of Lactobacillales has been implicated in arthritis, rheumatic disease, and diabetes (Yeoh et al., 2013; McLean et al., 2015). Furthermore, the relative abundance of Lactobacillales was predictive of higher host T-helper cell counts, suggesting an important link between Lactobacillales and host adaptive immunity (Snijders et al., 2016).

In summary, our results suggest that the microbial community in compound stomachs of ruminants at early ages shifts toward the mature ruminant state, the rumen microbiota community showed strong specificity compared to the other three stomach compartments. Together with our previously published data on rumen colonization (Zhang et al., 2018, unpublished), we show that the microbiome landscape represents three major mature stages in ruminant stomachs compartments early in life. We termed this period the primary stage (the first two weeks after birth) for the gain of access to foreign microorganisms, the secondary stage (Days 14–28) is the period of microbial transition, and the third stage (after Day 28) is the exogenous and endogenous microbial colonization stage. The diversity and within-group similarity increased with age, suggesting a more diverse but homogeneous and specific mature community, relative to the more heterogeneous and less diverse primary community. The findings presented here provide a clearer picture of the biochemical and microbial functions within the gastrointestinal tract in ruminants.

Furthermore, the data provide additional insights into diet formulation to promote ruminant animal health during the early stages of development.

AUTHOR CONTRIBUTIONS

KZ, YL, MG, YY, XW, and YC designed and conceived the experiments. KZ, YL, GL, and BL performed the experiments. KZ, YL, CL, and MG carried out microbial data processing, analysis, and interpretation. KZ and XW wrote the manuscript.

FUNDING

The present study was supported by the Key Research Program of Shaanxi Province (2017NY-072), the National Natural Science Foundation of China (31572369 and 31772571), as well as by China Agriculture Research System (CARS-39). XW was supported by the Innovative Talents Promotion Plan in Shaanxi Province, and is a Tang Scholar at Northwest A&F University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01846/full#supplementary-material

FIGURE S1 | Quality control of the microbial DNA extracted from stomach fluid samples. **(A)** Microbial DNA extracted with the EZNA® Stool DNA kit, NC: the mixture of reagents and water used. **(B)** Microbial DNA were extracted with the Fast-DNA® Spin kit for soil. NC: the mixture of reagents and water used. Amplified PCR products using the DNA extracted from rumen **(C)**, and abomasum and reticulum **(D)** fluid samples.

FIGURE S2 | Summary of rarefaction results based on operational taxonomic units (OTUs) (3% divergence) for each sample.

FIGURE S3 | Phylum level analysis of goat abomasum bacteria in different age groups. The ordinate indicates the species name under different classification levels, and the abscissa indicates the percentage of the abundance of a species of the sample. Different colors represent different groups.

FIGURE S4 | Genus level analysis of goat abomasum bacteria in different age groups. The ordinate indicates the species name under different classification levels, and the abscissa indicates the percentage of the abundance of a species of the sample. Different colors represent different groups (*$0.01 < p \leq 0.05$, **$0.001 < p < 0.01$, ***$p \leq 0.001$).

FIGURE S5 | Phylum level analysis of goat stomach compartment bacteria. The ordinate indicates the species name under different classification levels, and the abscissa indicates the percentage of the abundance of a species of the sample. Different colors represent different groups.

FIGURE S6 | Family level analysis of goat stomach compartment bacteria. The ordinate indicates the species name under different classification levels, and the abscissa indicates the percentage of the abundance of a species of the sample. Different colors represent different groups (*$0.01 < p \leq 0.05$, **$0.001 < p < 0.01$, ***$p \leq 0.001$).

FIGURE S7 | Metagenomic functional predictions for samples. Variations in KEGG metabolic pathways in functional bacterial communities throughout goat rumens.

TABLE S1 | Ingredients and nutrients of the experimental diets.

TABLE S2 | The sequence information of 63 different group sample.

TABLE S3 | The OTU classification statistics of the negative control sample.

TABLE S4 | The dynamics of bacterial community at the phyla level in different treatment groups.

TABLE S5 | The changes of bacterial community at the genus level in different treatment groups.

REFERENCES

Amato, K. R., Yeoman, C. J., Kent, A., Righini, N., Carbonero, F., Estrada, A., et al. (2013). Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. *ISME J.* 7, 1344–1353. doi: 10.1038/ismej.2013.16

Animut, G., and Goetsch, A. (2008). Co-grazing of sheep and goats: benefits and constraints. *Small Rumin. Res.* 77, 127–145. doi: 10.1016/j.smallrumres.2008.03.012

Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P., and Kauserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol. Ecol. Resour.* 13, 218–224. doi: 10.1111/1755-0998.12065

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2008). The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37(Suppl._1), D141–D145.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2013). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Davis, N. M., Proctor, D., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2017). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *bioRxiv* [Preprint]. doi: 10.1101/221499

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., et al. (2016). Population-level analysis of gut microbiome variation. *Science* 352, 560–564. doi: 10.1126/science.aad3503

Fernández, L., Langa, S., Martín, V., Maldonado, A., Jiménez, E., Martín, R., et al. (2013). The human milk microbiota: origin and potential roles in health and disease. *Pharmacol. Res.* 69, 1–10. doi: 10.1016/j.phrs.2012.09.001

Guaraldi, F., and Salvatori, G. (2012). Effect of breast and formula feeding on gut microbiota shaping in newborns. *Front. Cell. Infect. Microbiol.* 2:94. doi: 10.3389/fcimb.2012.00094

Han, X., Yang, Y., Yan, H., Wang, X., Qu, L., and Chen, Y. (2015). Rumen bacterial diversity of 80 to 110-day-old goats using 16S rRNA sequencing. *PLoS One* 10:e0117811. doi: 10.1371/journal.pone.0117811

Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J.* 7, 1069–1079. doi: 10.1038/ismej.2013.2

Jonsson, H. (2011). *Foraging Behaviour of Cattle, Sheep and Goats on Semi-Arid Pastures in Kenya*. Available at: http://ex-epsilon.slu.se

Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., et al. (2014). The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* 9:e88982. doi: 10.1371/journal.pone.0088982

Khan, M., Weary, D., and Von Keyserlingk, M. (2011). Invited review: effects of milk ration on solid feed intake, weaning, and performance in dairy heifers. *J. Dairy Sci.* 94, 1071–1081. doi: 10.3168/jds.2010-3733

Khodayar-Pardo, P., Mira-Pascual, L., Collado, M., and Martinez-Costa, C. (2014). Impact of lactation stage, gestational age and mode of delivery on breast milk microbiota. *J. Perinatol.* 34, 599–605. doi: 10.1038/jp.2014.47

Koren, O., Goodrich, J. K., Cullender, T. C., Spor, A., Laitinen, K., Bäckhed, H. K., et al. (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* 150, 470–480. doi: 10.1016/j.cell.2012.07.008

Kovatcheva-Datchary, P., Nilsson, A., Akrami, R., Lee, Y. S., De Vadder, F., Arora, T., et al. (2015). Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of Prevotella. *Cell Metab.* 22, 971–982. doi: 10.1016/j.cmet.2015.10.001

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Lauder, A. P., Roche, A. M., Sherrillmix, S., Bailey, A., Laughlin, A. L., Bittinger, K., et al. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 4:29. doi: 10.1186/s40168-016-0172-3

Ley, R. E. (2016). Gut microbiota in 2015: prevotella in the gut: choose carefully. *Nat. Rev. Gastroenterol. Hepatol.* 13, 69–70. doi: 10.1038/nrgastro.2016.4

Lim, E. S., Rodriguez, C., and Holtz, L. R. (2018). Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome* 6:87. doi: 10.1186/s40168-018-0475-7

Mao, S., Zhang, M., Liu, J., and Zhu, W. (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Sci. Rep.* 5:16116. doi: 10.1038/srep16116

Martín, R., Langa, S., Reviriego, C., Jimínez, E., Marín, M. L., Xaus, J., et al. (2003). Human milk is a source of lactic acid bacteria for the infant gut. *J. Pediatr.* 143, 754–758. doi: 10.1016/j.jpeds.2003.09.028

McLean, M. H., Dieguez, D., Miller, L. M., and Young, H. A. (2015). Does the microbiota play a role in the pathogenesis of autoimmune diseases? *Gut* 64, 332–341. doi: 10.1136/gutjnl-2014-308514

Morgavi, D. P., Rathahao-Paris, E., Popova, M., Boccard, J., Nielsen, K. F., and Boudra, H. (2015). Rumen microbial communities influence metabolic phenotypes in lambs. *Front. Microbiol.* 6:1060. doi: 10.3389/fmicb.2015.01060

Morowitz, M. J., Poroyko, V., Caplan, M., Alverdy, J., and Liu, D. C. (2010). Redefining the role of intestinal microbes in the pathogenesis of necrotizing enterocolitis. *Pediatrics* 125, 777–785. doi: 10.1542/peds.2009-3149

Peng, S., Yin, J., Liu, X., Jia, B., Chang, Z., Lu, H., et al. (2015). First insights into the microbial diversity in the omasum and reticulum of bovine using Illumina sequencing. *J. Appl. Genet.* 56, 393–401. doi: 10.1007/s13353-014-0258-1

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Ramayo-Caldas, Y., Mach, N., Lepage, P., Levenez, F., Denis, C., Lemonnier, G., et al. (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J.* 10, 2973–2977. doi: 10.1038/ismej.2016.77

Rey, M., Enjalbert, F., and Monteils, V. (2012). Establishment of ruminal enzyme activities and fermentation capacity in dairy calves from birth through weaning. *J. Dairy Sci.* 95, 1500–1512. doi: 10.3168/jds.2011-4902

Rubino, F., Carberry, C., Waters, S. M., Kenny, D., Mccabe, M. S., and Creevey, C. J. (2017). Divergent functional isoforms drive niche specialisation for nutrient acquisition and use in rumen microbiome. *ISME J.* 11, 932–944. doi: 10.1038/ismej.2016.172

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Schloss, P. D., Gevers, D., and Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One* 6:e27310. doi: 10.1371/journal.pone.0027310

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60

Shin, N.-R., Whon, T. W., and Bae, J.-W. (2015). *Proteobacteria*: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* 33, 496–503. doi: 10.1016/j.tibtech.2015.06.011

Snijders, A. M., Langley, S. A., Kim, Y.-M., Brislawn, C. J., Noecker, C., Zink, E. M., et al. (2016). Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nat. Microbiol.* 2:16221. doi: 10.1038/nmicrobiol.2016.221

Toscano, M., De Grandi, R., Peroni, D. G., Grossi, E., Facchin, V., Comberiati, P., et al. (2017). Impact of delivery mode on the colostrum microbiota composition. *BMC Microbiol.* 17:205. doi: 10.1186/s12866-017-1109-0

Trenkle, A. (2001). *Experience with an Electronic Identification System for Cattle*. Ames, IA: Iowa State University.

Wang, J., Fan, H., Han, Y., Zhao, J., and Zhou, Z. (2017). Characterization of the microbial communities along the gastrointestinal tract of sheep by 454 pyrosequencing analysis. *Asian Australas. J. Anim. Sci.* 30, 100–110. doi: 10.5713/ajas.16.0166

Wang, L., Xu, Q., Kong, F., Yang, Y., Wu, D., Mishra, S., et al. (2016). Exploring the goat rumen microbiome from seven days to two years. *PLoS One* 11:e0154354. doi: 10.1371/journal.pone.0154354

Wang, Y., Sheng, H.-F., He, Y., Wu, J.-Y., Jiang, Y.-X., Tam, N. F.-Y., et al. (2012). Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol.* 78, 8264–8271. doi: 10.1128/AEM.01821-12

White, J. R., Nagarajan, N., and Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* 5:e1000352. doi: 10.1371/journal.pcbi.1000352

Yeoh, N., Burton, J. P., Suppiah, P., Reid, G., and Stebbings, S. (2013). The role of the microbiome in rheumatic diseases. *Curr. Rheumatol. Rep.* 15:314. doi: 10.1007/s11926-012-0314-y

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Lei, Zhang, Guo, Li, Li, Li, Yang, Chen and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 June 2018
doi: 10.3389/fmicb.2018.01224

Comparative Genomics Reveals Evidence of Genome Reduction and High Extracellular Protein Degradation Potential in *Kangiella*

Jiahua Wang[1,2†], Ye Lu[1,2†], Muhammad Z. Nawaz[1,2] and Jun Xu[1,2]*

[1] *Institute of Oceanography, Shanghai Jiao Tong University, Shanghai, China,* [2] *State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China*

The genus *Kangiella* has recently been proposed within the family Kangiellaceae, belonging to order Oceanospirillales. Here, we report the complete genome sequence of a novel strain, *Kangiella profundi* FT102, which is the only *Kangiella* species isolated from a deep sea sediment sample. Furthermore, gaps in the publicly available genome scaffold of *K. aquimarina* DSM 16071 (NCBI Reference Sequence: NZ_ARFE00000000.1) were also filled using polymerase chain reaction (PCR) and Sanger sequencing. A comparative genomic analysis of five *Kangiella* and 18 non-*Kangiella* strains revealed insights into their metabolic potential. It was shown that low genomic redundancy and *Kangiella*-lineage-specific gene loss are the key reasons behind the genome reduction in *Kangiella* compared to that in any other free-living Oceanospirillales strain. The occurrence of relatively diverse and more frequent extracellular protease-coding genes along with the incomplete carbohydrate metabolic pathways in the genome suggests that *Kangiella* has high extracellular protein degradation potential. Growth of *Kangiella* strains has been observed using amino acids as the only carbon and nitrogen source and tends to increase with additional tryptone. Here, we propose that extracellular protein degradation and amino acid utilization are significant and prominent features of *Kangiella*. Our study provides more insight into the genomic traits and proteolytic metabolic capabilities of *Kangiella*.

Keywords: marine bacteria, Oceanospirillales, *Kangiella*, genome reduction, protein degradation

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Alfonso Benítez-Páez, Instituto de Agroquímica y Tecnología de Alimentos (IATA), Spain
David Correa Galeote, Universidad Nacional Autónoma de México, Mexico

***Correspondence:**
Jun Xu
xujunn@sjtu.edu.cn

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *21 September 2017*
Accepted: *22 May 2018*
Published: *07 June 2018*

Citation:
Wang J, Lu Y, Nawaz MZ and Xu J (2018) Comparative Genomics Reveals Evidence of Genome Reduction and High Extracellular Protein Degradation Potential in Kangiella. Front. Microbiol. 9:1224. doi: 10.3389/fmicb.2018.01224

INTRODUCTION

The ocean covers 71% of the earth's surface and is regarded as the largest habitat for life on the planet Earth. Marine microorganisms are known to play an essential role in energy conservation and biogeochemical cycling in the oceans. Heterotrophic prokaryotes are considered key players in the decomposition of the dissolved organic matter (DOM) and particulate organic matter (POM) present therein (DeLong and Karl, 2005; Azam and Malfatti, 2007).

Oceanospirillales is an order of proteobacteria with seven families of heterotrophic marine bacteria that are usually associated with oil spills and are known to be involved in xylan and hydrocarbon utilization (Choi et al., 2012; Cao et al., 2014). Recently, a new family named Kangiellaceae has been proposed within the order Oceanospirillales based on phylogenetic,

chemotaxonomic and physiological characteristics; this family comprises three genera: *Kangiella, Aliikangiella,* and *Pleionea* (Wang et al., 2015). Although the *Kangiella* genus was reclassified within Kangiellaceae instead of Alcanivoracaceae based on 16S rRNA gene phylogeny (Wang et al., 2015), taxonomic signatures at the genomic level are still needed.

Kangiella are gram-negative long rods that are non-motile and non-spore-forming bacteria. Nine strains belonging to the genus *Kangiella* have been isolated from various marine environments, including marine sand (Kim et al., 2015), tidal flat sediments and coastal regions (Yoon et al., 2004, 2012; Romanenko et al., 2010; Ahn et al., 2011; Jean et al., 2012; Lee et al., 2013). For example, both *K. koreensis* DSM 16069 and *Kangiella aquimarina* DSM 16071 were isolated from tidal flat sediment at Daepo Beach, Yellow Sea, Korea, and were shown to grow optimally at 30–37°C and pH 7.0–8.0 (Yoon et al., 2004). *K. sediminilitoris* KCTC 23892 was isolated from tidal flat sediment from the South Sea in South Korea, which was shown to grow optimally at 30–37°C and pH 7.0–7.5 (Lee et al., 2013). *Kangiella geojedonensis* KCTC 23420 was isolated from seawater off the southern coast of Korea, and was shown to grow optimally at 10–40°C and pH 7.0–7.5 (Yoon et al., 2004). Recently, we isolated a novel strain named *Kangiella profundi* from a deep sea sediment sample that was collected from the southwest Indian Ocean at a depth of 2784 m (Xu et al., 2015). Due to limited research, our understanding about the metabolic potential and ecological functions of bacteria belonging to the family Kangiellaceae is still obscure.

To date, no genome has been completely sequenced from either of the two genera *Aliikangiella* and *Pleionea*. Moreover, only three completely sequenced genomes of the genus *Kangiella*, i.e., *K. koreensis* DSM 16069 (GenBank accession: GCA_000024085.1), *K. sediminilitoris* KCTC 23892 (GenBank accession: GCA_001708405.1), and *K. geojedonensis* KCTC 23420 (GenBank accession: GCA_000981765.1), are available. In this study, we report the complete genome sequence of *K. profundi* FT102 (GenBank accession: CP025120). Comparative genomics-based approaches were utilized to explore the metabolic potential of *Kangiella*. We show that featured genomic reduction in *Kangiella* is due to low genomic redundancy as well as *Kangiella* lineage-specific gene loss. The metabolic base of *Kangiella* as a powerful extracellular protein degrader was investigated. Our study provides the first insights into the genomic and metabolic capabilities of *Kangiella*.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Kangiella profundi FT102^T (= CGMCC 1.12959^T = KCTC 42297^T = JCM 30232^T) is from our laboratory stock. The strains *K. koreensis* DSM 16069 (JCM 12317), *K. aquimarina* DSM 16071 (JCM 12318) and *K. geojedonensis* KCTC 23420 were obtained from Japan Collection of Microorganisms (JCM) and Korean Collection for Type Cultures (KCTC). These *Kangiella* strains were grown in marine broth 2216 (Xu et al., 2015) and marine peptone medium (1 L of broth containing 1 g of yeast extract, 10 g of tryptone, and 1000 mL of artificial sea water; pH adjusted to 7.3) at 37°C and 200 rpm under aerobic conditions. Bacterial growth was measured by an automatic turbiditimetry using a Bioscreen C analyzing system (Labsystems).

Cultivation in Defined Medium

The defined cultivation method for *Kangiella* has been described elsewhere (Widdel et al., 2006). All 20 common amino acids, as a sole source of carbon, nitrogen and energy, were filter-sterilized and added to the media at a final concentration of 0.2 g/L of each amino acid. The medium contained (g/L): NaCl (26.0), $MgCl_2{\cdot}6H_2O$ (5.0), $CaCl_2{\cdot}2H_2O$ (1.4), NH_4Cl (0.3), KH_2PO_4 (0.1), KCl (0.50), $NaNO_3$ (1.0), with (mg/L), EDTA, disodium salt (5.20), $FeSO_4{\cdot}7H_2O$ (2.10), and (μg/L): H_3BO_3 (10.0), $MnCl_2{\cdot}4H_2O$ (5.0), $CoCl_2{\cdot}6H_2O$ (190.0), $ZnSO_4{\cdot}7H_2O$ (144.0), $CuCl_2{\cdot}2H_2O$ (10.0), $Na_2MoO_4{\cdot}2H_2O$ (36.0), $NiCl_2{\cdot}6H_2O$ (24.0). Vitamins were filter-sterilized and added to final concentrations (μg/L) of 4-Aminobenzoic acid (4.0), D(+)-biotin (1.0), Nicotinic acid (10.0), D(+)-Pantothenic acid, calcium salt (5.0), Pyridoxine-HCl (15.0), Cyanocobalamin (5.0), and Thiamine-HCl (10.0).

Genomic DNA Extraction

Genomic DNA was extracted as described by Chen et al. (2005). The quality and quantity of the extracted genomic DNA were checked by agarose gel electrophoresis and a NanoDrop spectrophotometer (Thermo Scientific).

Genome Sequencing and Assembly

Genomic DNA of *K. profundi* FT102 was sequenced using a next-generation sequencing platform. Illumina HiSeq 2000 (2*120bp, 0.5-kb insert size) sequencing was performed by BGI (Shenzhen, China), and *de novo* assembly was performed by using the Velvet assembler v.1.2.10 (Zerbino and Birney, 2008) with the following parameters: min kmer 31; max kmer 99; kmer step 6; and insert length 500. Eight contigs were assembled, including one containing an rRNA operon with doubled coverage and two short ones (< 550 bp). Then, these contigs were ordered according to the completed genome of *K. koreensis* DSM 16069. All gaps were filled using polymerase chain reaction (PCR) and Sanger sequencing (**Supplementary Material S1**). The coding sequence of *dnaA* was set as the genomic start in the positive strand.

Genomic Data Collection

We collected five *Kangiella* (including *K. profundi* FT102) and 18 genomes of other Oceanospirillales (**Table 1**) from the NCBI database[1]. Both nucleotide and protein sequences were used in this study. The scaffold genome of *K. aquimarina* DSM 16071 was completed by PCR and sequencing in the present study.

Genome Annotation and Comparative Analysis

Replication origins (*oriC*s) in bacterial genomes were predicted using Ori-Finder[2], which predicts based on analysis of the base

[1]http://ftp.ncbi.nlm.nih.gov/genomes
[2]http://tubic.tju.edu.cn/Ori-Finder/

TABLE 1 | Genomic features of the *Kangiella* strains used in present study.

Species name	*Kangiella aquimarina* DSM 16071	*Kangiella geojedonensis* KCTC 23420	*Kangiella koreensis* DSM 16069	*Kangiella profundi* FT102	*Kangiella sediminilitoris* KCTC 23892
Genome Size	2,686,123	2,495,242	2,852,073	2,653,010	2,496,140
%GC	43.72%	43.78%	43.69%	43.81%	43.40%
tRNA operon	41	41	41	41	41
rRNA number	2	2	2	2	2
ORF number	2446	2208	2632	2484	2308
ORF total length	2,438,865	2,244,783	2,594,526	2,382,009	2,250,390
%ORF length	90.79%	89.96%	90.97%	89.79%	90.15%
SignalP	317	298	339	305	283
%signalP	12.96%	13.50%	12.88%	12.28%	12.26%
TM	606	533	701	610	581
%TM	24.78%	24.14%	26.63%	24.56%	25.17%
No. of Paralogs	130	121	150	142	133
ORF in paralogs	344	296	396	370	332

SingnalP, means signal peptide-fused protein; %SignalP, means percentage of signal peptide-fused proteins; TM, means transmembrane protein; %TM, means percentage of transmembrane proteins.

composition asymmetry using the Z-curve method, distribution of DnaA boxes, and the occurrence of genes frequently close to *oriC*s (Gao and Zhang, 2008). Putative protein-coding sequences were predicted using Glimmer 3 (Delcher et al., 1999), which was trained with the CDS (coding sequences) from four completely sequenced *Kangiella* genomes. Manual curation of all coding sequences was performed by examining the database hits of BLAST 2.2.25+against the NR database and the RAST server (Aziz et al., 2008). tRNA and rRNA genes were predicted using tRNAscan-SE 1.3.1 (Lowe and Eddy, 1997) and RNAmmer 1.2 (Lagesen et al., 2007), respectively. Signal peptides of the ORFs were predicted with SignalP 4.1 (Petersen et al., 2011; **Supplementary Material S3**). Peptidase genes were predicted with MEROPS batch BLAST (Rawlings et al., 2016; **Supplementary Material S3**).

Protein families of 23 Oceanospirillales strains including 5 *Kangiella* were clustered using a local OrthoMCL 2.0.9 (Li et al., 2003) with the following cut-off values: identity, 30%; coverage, 50%; *E*-value, 1e-5; score, 40; and MCL Markov clustering inflation index, 1.5. To identify the *Kangiella* signature proteins, the core genome components were blasted using BLASTP against the non-redundant protein sequence (nr) database, excluding the *Kangiella* genus. A protein was considered a *Kangiella* signature protein if there were no BLAST hits with acceptable *E*-values ($<10^{-1}$), similarity (>20%), or coverage (>50%). GO (gene ontology) term enrichment was performed using Blast2GO PRO 3.0. COG (cluster of orthologous genes) classification was performed using the WebMGA online server[3].

Metabolic Network of *Kangiella*

The core metabolic network of *Kangiella* was constructed based on the core genome information of the five *Kangiella* strains using KEGG BlastKOALA[4].

[3]http://weizhong-lab.ucsd.edu/metagenomic-analysis/server/cog/

[4]http://www.kegg.jp/blastkoala/

Genome Rearrangement Analysis

Locally collinear blocks (LCBs) of the five *Kangiella* strains were investigated using Mauve aligner 2.4.0 (Darling et al., 2004). Each LCB represents a region of homologous sequence without rearrangement among genomes.

GC-Skew Analysis

GenSkew online application[5] was used to compute and plot GC-skew data.

RESULTS

Genomic Features of *Kangiella profundi* FT102

The complete genome sequence of *K. profundi* FT102 is composed of a circular chromosome of 2,653,010 bp with 43.81% GC content (**Figure 1**). The coding region covers 89.79% of the genome, encoding 2,484 proteins, of which 2,042 proteins were annotated with COG classification (**Supplementary Table S1**). Translation, ribosomal structure, general function prediction only (R), biogenesis (J), and amino acid transport, and metabolism (E) were the most abundant COG categories (7.81, 6.60, and 6.42%, respectively). The genome also encodes 41 tRNAs, as well as 2 rRNA operons. Signal peptides and transmembrane helices account for 12.28% (305) and 24.56% (610) of the protein-coding genes, respectively (**Table 1**).

Moreover, eight genomic islands (GIs) were predicted in the *K. profundi* FT102 strain (**Figure 1**). One of the GIs encodes 15 proteins (orf01390 to orf01530), including a putative type VI secretion system and a related lysozyme, which might provide antibacterial ability to *K. profundi* FT102. Another GI with the least GC content compared to the rest of the genome was

[5]http://genskew.csb.univie.ac.at/

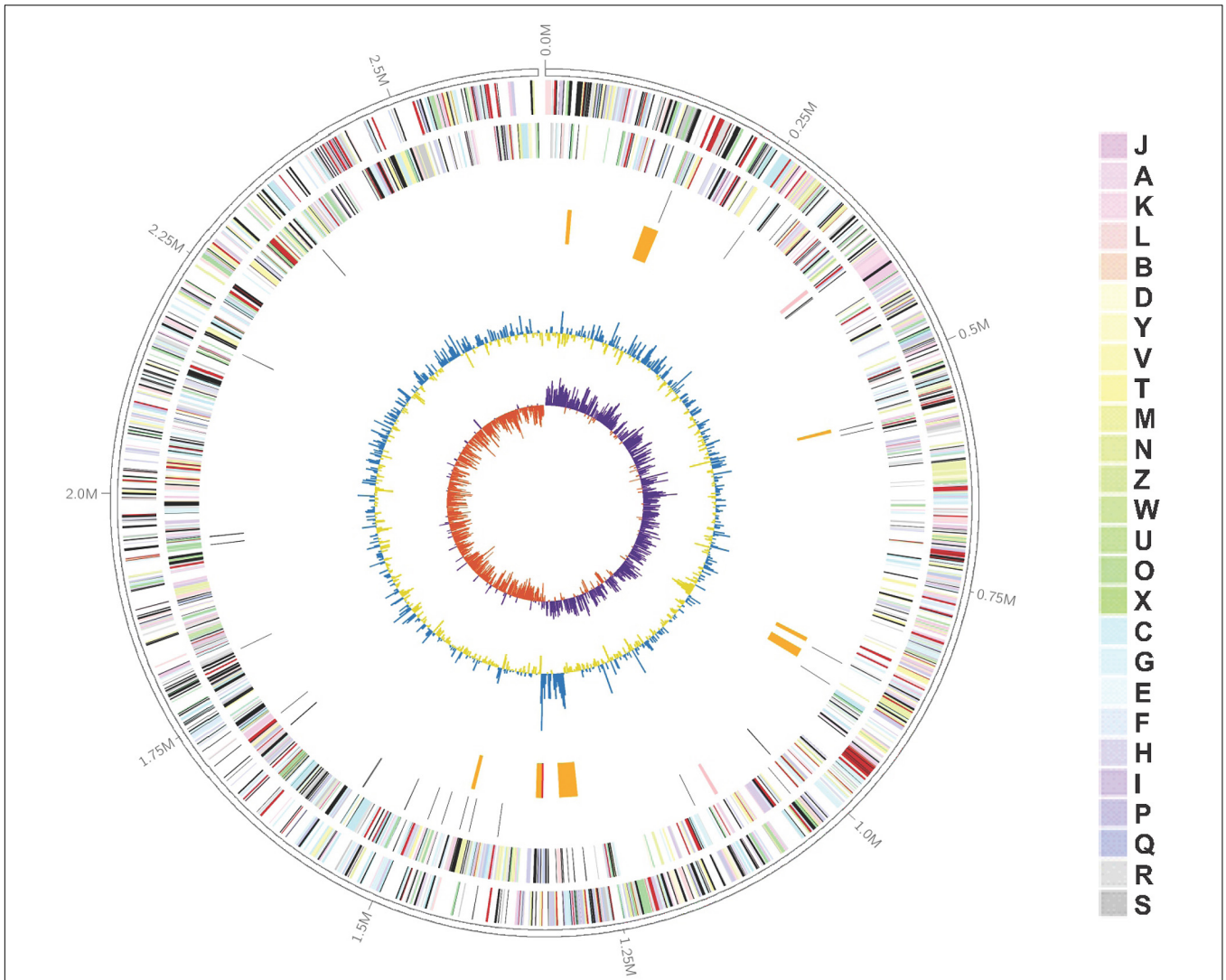


FIGURE 1 | Graphical representation of the *Kangiella profundi* FT102 genome. Genes on the forward (shown in **Outer** circle) and reverse (shown in **Inner** circle) strands are colored according to their cluster of orthologous genes (COG) categories, RNA genes are highlighted with different colors (tRNAs orange, rRNAs red, other RNAs green), gene islands and GC content are shown in black, and GC skew is shown in light green/orange.

predicted to harbor gene clusters for the biosynthesis of capsular polysaccharides. In contrast, the region with the highest GC content, located next to the GC-skew, was annotated as a putative heavy-metal resistance island against Zn/Co/Cd.

Signature Proteins of the Genus *Kangiella*

The *Kangiella* genus is of phylogenetic interest because of its very isolated location in Oceanospirillales. *K. koreensis* is the first type species of this genus to be described and was selected for whole-genome sequencing as part of the Genomic Encyclopedia of Bacteria and Archaea project (Han et al., 2009). With five complete genomes at hand, we identified the signature proteins of *Kangiella* by BLAST analysis on the core genome with the following cutoff: at least 60% identity (*e*-value: 1e-30) with each other in the same orthologous families and less than 30% identity (*e*-value: 1e-1) with any non-*Kangiella* hits from the NR database. Under the cutoff, four orthologous families (KangOF24, 566, 656, and 804) appeared as *Kangiella* specific and were annotated as hypothetical proteins. Notably, two copies of KangOF24 were observed in *K. sediminilitoris* KCTC 23892 (AOE50183.1 and AOE50181.1). These signature proteins could be used as genus-specific molecular markers to identify *Kangiella* independently of detection of 16S rRNA gene as a phylogenic marker.

Analysis of Orthologous Gene Families

The genomic features of the five *Kangiella* species used in the present study are shown in **Table 1**. The Markov clustering algorithm OrthoMCL was used to identify orthologous gene families. Further analysis of these families in the five *Kangiella* genomes revealed 3,544 orthologs (**Figure 2**

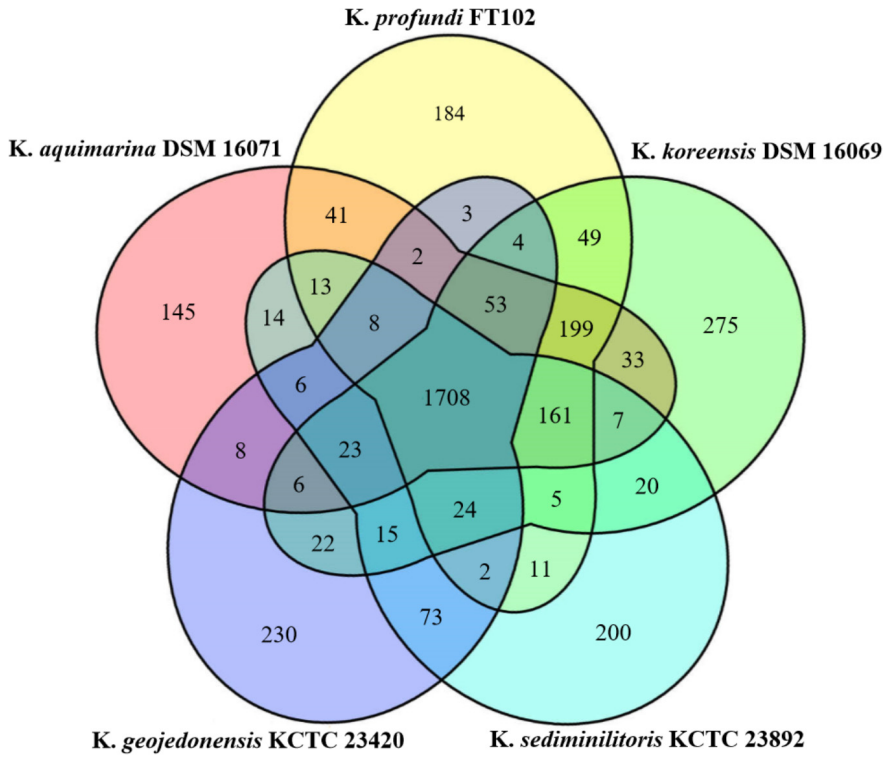


FIGURE 2 | Venn diagram of the orthologous families of the five *Kangiella* species. The size of the core genome is shown in the center, while the numbers of strain-specific genes in *Kangiella* strains are shown in the petals.

and **Supplementary Table S2**), which includes (1) 1,708 core genome components, of which only 21 orthologs contain duplication in at least one genome (**Supplementary Table S3**); (2) 802 components in the dispensable genome, which have representatives in at least two but not all of the five genomes; and (3) 1,034 components that were uniquely present only in one genome, which include only 10 lineage-unique (LSE) gene families and 1,024 singletons. According to the COG classification (**Supplementary Figure S1** and **Supplementary Table S4**), it is evident that core genes mainly belong to categories J (translation, ribosome structure, and biogenesis), C (energy production and conversion), D (cell cycle control, cell division, and chromosome partitioning), F (nucleotide transport and metabolism), G (carbohydrate transport and metabolism), H (coenzyme transport and metabolism), I (lipid transport and metabolism), and L (replication, recombination, and repair); the unique genes are involved in V (defense mechanisms); and dispensable genes are found more often associated with P (inorganic ion transport and metabolism), and T (signal transduction mechanisms).

Strain-Specific Losses and Gains in the Metabolic Pathways in *Kangiella*

To determine the differences in the metabolic potential among *Kangiella* strains, we compared the GO terms from each strain, and strain-specific gene gains and losses were identified. To avoid the biases of annotation, "gained" GO terms were derived from strain-specific orthologous families that were not possessed by any of the other four strains. On the other hand, "lost" GO terms were retrieved only from those dispensable orthologous families that were owned by each of the four strains and missing in only one. Genes related to these GO terms were manually checked using both KEGG pathways and BLASTP analysis.

The results suggest that *K. profundi* FT102 gained a type III restriction endonuclease (orf15860) and a related DNA methyltransferase (orf15870), as well as a type VI secretion system. The gene clusters related to sulfide reduction and arginine biosynthesis were missing in the FT102 strain; this absence was also validated through PCR and sequencing (**Supplementary Figure S2** and **Supplementary Material S2**).

Kangiella aquimarina DSM 16071 was found to be the only strain with a type I restriction-modification system, including site-specific deoxyribonuclease activity (WP_033414025.1) and related methyltransferase (WP_033414024.1). Moreover, 2-isopropylmalate synthases were observed in other analyzed strains but were absent in *K. aquimarina* DSM 16071. 2-isopropylmalate is an important substrate for the biosynthesis of leucine from pyruvate and could be catalyzed to leucine along with acetyl-CoA, 3-methyl-2-oxobutanoate, and water. However, *K. aquimarina* DSM 16071 contains isopropylmalate isomerase (WP_018624573.1), which could transform (2R,3S)-3-isopropylmalate to (2S)-2-isopropylmalate. Therefore, it was suggested that *K. aquimarina* DSM 16071 could synthesize leucine, similar to other species, using an alternate pathway.

Kangiella koreensis DSM 16069 represents the largest genome size compared to the other four strains used in this study, and no strain-specific GO term loss was found in *K. koreensis* DSM 16069. Furthermore, *K. koreensis* DSM 16069 was found to exhibit putative nitrate reduction capability. A gene cluster with at least 19 proteins (ACV26744.1-ACV26762.1) and 12 GO terms related to Mo-molybdopterin cofactor biosynthesis and dissimilatory nitrate reduction were found limited to *K. koreensis* DSM 16069.

Although, the genome size of *K. geojedonensis* KCTC 23420 is the smallest among all the strains used herein, it was the only one predicted to have a putative capacity for synthesizing all 20 amino acids. A gene cluster (AKE51219.1–AKE51221.1) related to branched-chain amino acid biosynthetic and metabolic processes was found in *K. geojedonensis* KCTC 23420, while it was missing in the rest of the four *Kangiella* strains. Furthermore, the presence of homoserine kinase (AKE52760.1) possibly represents a strain-specific biosynthetic pathway from L-aspartate to threonine via O-phospho-L-serine in *K. geojedonensis* KCTC 23420. Additionally, 10 strain-specific GO terms in 6 orthologous families were found in the second island of *K. geojedonensis* KCTC 23420 (consisting of 21 protein-coding genes) and are involved in capsule biosynthesis. However, they lack genes involved in bacterial respiration (such as cytochrome c oxidase subunits and HemN), oxidative resistance (such as superoxide dismutase), heavy metal transport, and nitrite utilization.

Kangiella sediminilitoris KCTC 23892 was the only strain with an ammonium transmembrane transporter (AOE50237.1), while it was also the only one among the five *Kangiella* strains that lacked aspartate-ammonia ligase, whose function could be supplemented by asparagine synthase (AOE49243.1).

Reconstruction of the Basic Metabolic Pathways of Carbohydrate and Amino Acid

Based on the comparison KEGG pathway analysis described above, the main metabolic network of the *Kangiella* strains were reconstructed. The glycolysis, tricarboxylic acid (TCA) cycle, and amino acid biosynthesis and transport pathways are shown in **Figure 3**. Although *Aliikangiella marina* GYP-15^T was reported as a carbohydrate utilizer (Wang et al., 2015), the gene encoding carbohydrate transporter or glucose kinase was not predicted in the *Kangiella* strains, which might be the reason that the heterotrophic growth of *Kangiella* strains could not be supported when using glucose as the sole carbon source (data not shown). Therefore, it is possible that carbohydrates are derived from central carbon metabolism (such as pyruvate) and transformed via reverse glycolysis and the pentose phosphate pathway. Fatty acids could be one of the carbon sources, as the genes coding for the key enzymes in the fatty acid beta-oxidation pathway could be found in the genomes of *Kangiella*. Amino acids could also feed central carbon metabolism via deamination and/or transamination, and most of them could be synthesized by each strain.

Diversity in metabolic capability in the biosynthesis of certain amino acids, including valine, isoleucine, arginine, and phenylalanine, was observed among the analyzed strains (**Figure 3**). *K. geojedonensis* KCTC 23420 appeared to be able to synthesize valine and isoleucine, while *K. profundi* FT102 could not synthesize arginine. Additionally, biosynthesis of phenylalanine was found to be missing in both *K. profundi* FT102 and *K. sediminilitoris* KCTC 23892 because of the loss of a gene cluster containing phospho-2-dehydro-3-deoxyheptonate aldolase and prephenate dehydratase (**Supplementary Figure S2**). Particularly, the predicted disability of the phenylalanine and arginine biosynthesis in *K. profundi* FT102, were experimentally validated by growth curve analysis under various defined media with amino acids as the sole carbon and nitrogen source (**Figure 4**). Interestingly, we detected various kinds of secretory peptidases, which are exported by TIISS/Tat/Sac system to degrade extracellular proteins into amino acids and peptides in these strains. Furthermore, seven types of amino acid/peptide transporters were also identified, which could support the intake of these amino acids/peptides from the extracellular environment.

Genome Reduction and Metabolic Simplicity of *Kangiella* Strains

The genomic size of the *Kangiella* strains is smaller than that of any other free-living Oceanospirillales strains. To gain insights into the genome reduction in *Kangiella*, we performed comparative genome analysis with 18 non-*Kangiella* completely sequenced genomes of Oceanospirillales strains (**Supplementary Table S5**). We particularly focused on the investigating genomic redundancy, functional gene categories and genome function reduction. OrthoMCL clustering results showed that each *Kangiella* strain possesses at most 150 paralogous gene families within 396 genes, whereas non-*Kangiella* strains exhibit at least 351 paralogous families within 1,118 genes, demonstrating that low genomic redundancy is the primary reason behind the genome reduction. There were fewer COG categories in *Kangiella* strains than in their non-*Kangiella* counterparts (**Supplementary Table S1**), especially for carbohydrate transport and metabolism (COG category G), genes in the mobilome (COG category X) and cell motility (COG category N). Each *Kangiella* strain exhibits no more than 41 COG category G orthologous proteins, while non-*Kangiella* strains have at least 60 COG category G, demonstrating that *Kangiella* has a decreased preference for carbohydrate transport and metabolism. On the average, *Kangiella* strains have only four COG category X orthologous proteins, while non-*Kangiella* strains have 48, which reflects less horizontal gene transfer during the evolution of *Kangiella*. There also seemed to be fewer COG category N orthologous proteins than the average in non-*Kangiella* strains, which might be mainly because *Kangiella* strains lack flagellum (**Figure 5**).

To gain insights into the reduction in *Kangiella*-specific function, we clustered all five *Kangiella* protein sequences together with their 18 counterparts using OrthoMCL. Among 23,058 Oceanospirillales orthologous families (OceaOFs) (**Supplementary Table S6**), we found 96 families that were only absent in *Kangiella* and called them "*Kangiella*-specific-absences." Genes both from these families and all five *Kangiella*

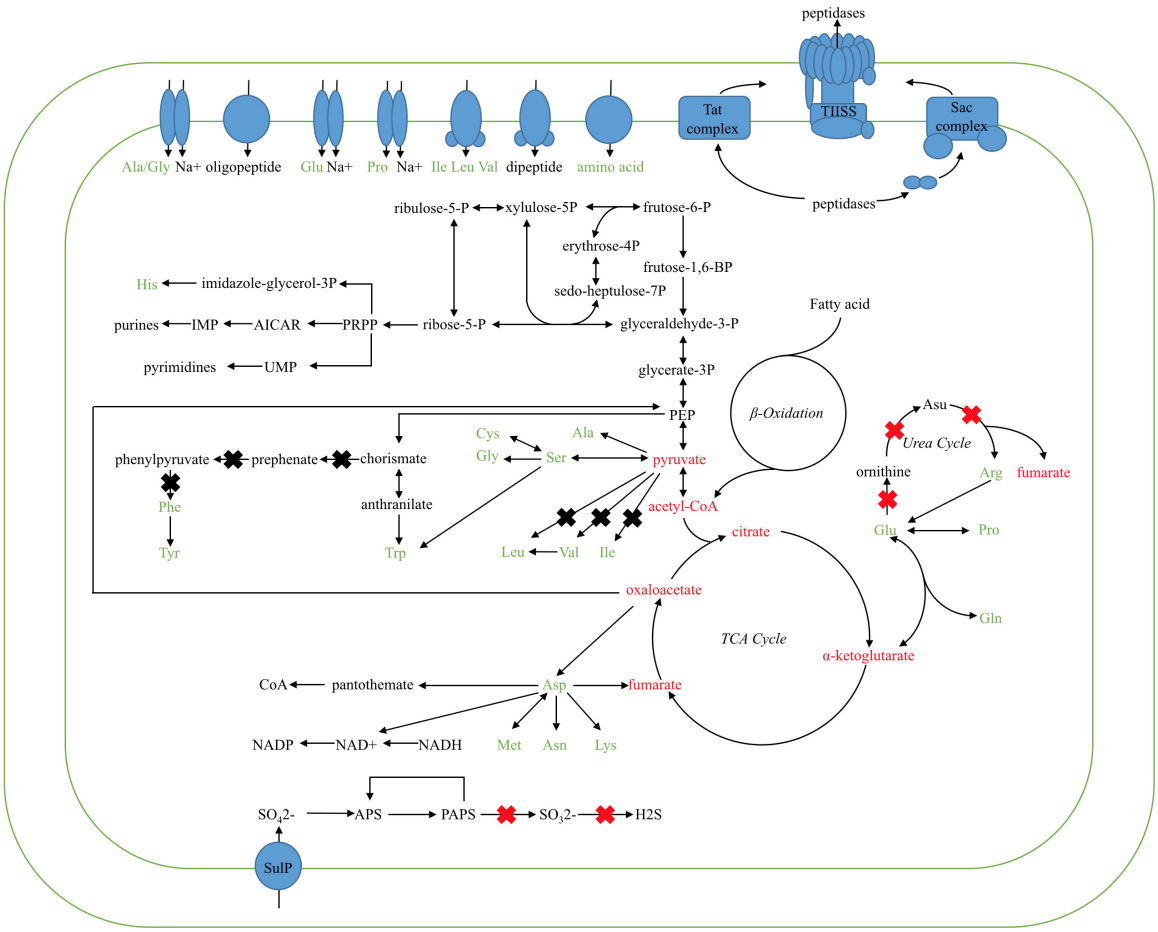


FIGURE 3 | Proposed metabolic pathways of the TCA cycle, inferred from the *Kangiella* core genome. Ovals in blue represent the amino acid/oligopeptide transporters and protein-secreting systems. Black arrows show the flow of the metabolic pathway. X marks in the red color show the genes absent only in *K. profundi* FT102 (confirmed by PCR and growth experiments), whereas X marks in black represent the genes absent in *K. profundi* FT102 and at least one other species (but not all the species used here).

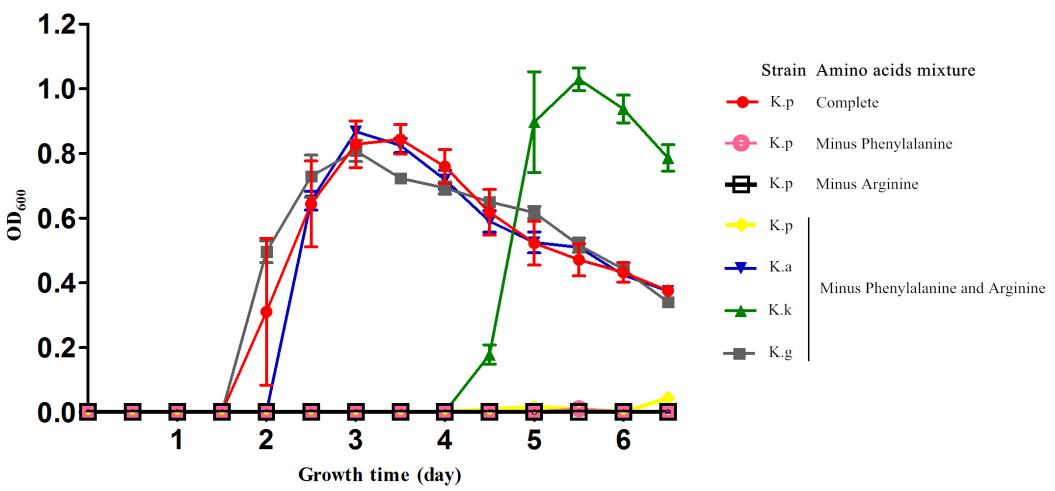


FIGURE 4 | Growth curve of the four *Kangiella* strains cultured in the defined amino acids medium. K.p, *Kangiella aquimarina* DSM 16071; K.k, *Kangiella koreensis* DSM 16069; K.g, *Kangiella geojedonensis* KCTC 23420; and K.p, *Kangiella profundi* FT102. Complete: 20 amino acids; Minus Phenylalanine: phenylalanine-absent 19 amino acids; Minus Arginine: arginine-absent 19 amino acids; Minus Phenylalanine and Arginine: phenylalanine- and arginine-absent 18 amino acids.

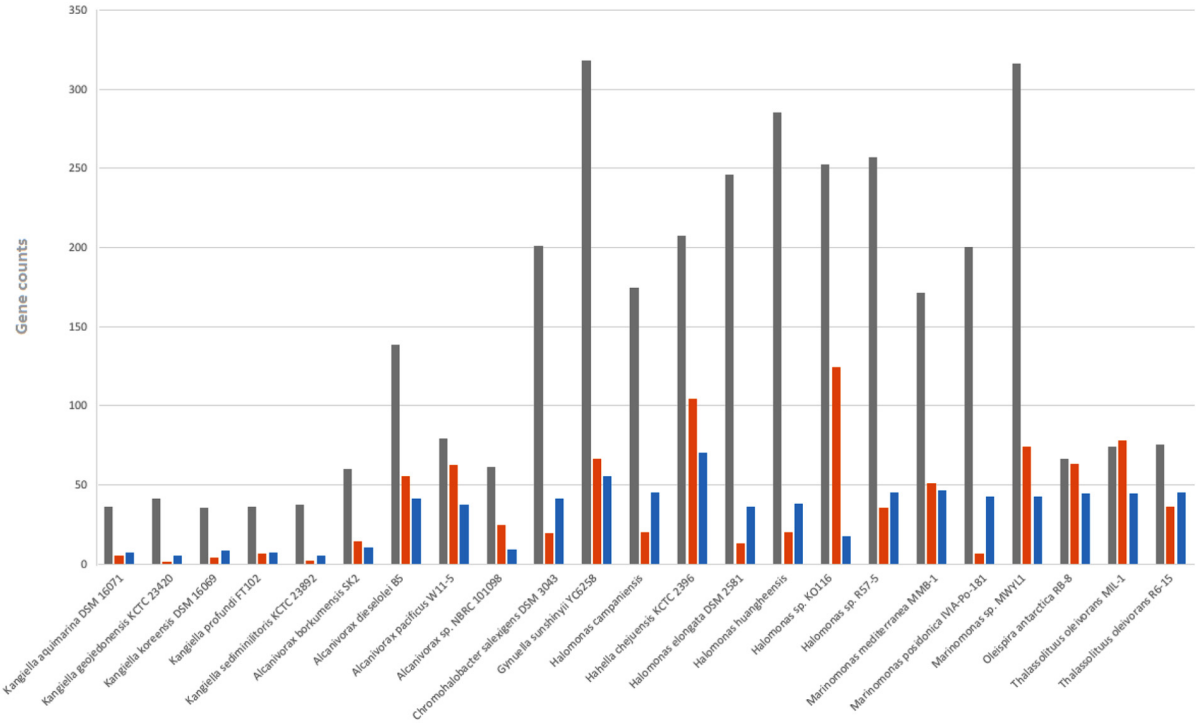


FIGURE 5 | Number of genes belonging to selected COG categories in all 23 strains in Oceanospirillales. G shown in gray is related to carbohydrate transport and metabolism; N shown in orange is related to cell motility; and X shown in blue is related to mobility.

genomes were annotated against the GO database. If a GO term from these "*Kangiella*-specific-absent" families was also found to be absent in all the *Kangiella* genomes, this pathway was inferred to be lost in *Kangiella*. Approximately 31 GO terms from 24 OceaOFs were found missing in *Kangiella* only (**Supplementary Table S7**). All the non-*Kangiella* Oceanospirillales were found to contain a phosphoenolpyruvate-dependent sugar phosphotransferase system; in contrast, this protein was missing in *Kangiella*. Glycerol is the primary carbon source available to halophilic heterotrophic communities. Glycerol kinase (EC 2.7.1.30; ATP-glycerol 3-phosphotransferase) is required for glycerol metabolism, and *Kangiella* strains were found to be deprived of it, suggesting that glycerol is not the carbon source in *Kangiella*. Malate synthase works together with isocitrate lyase in the glyoxylate cycle to bypass two oxidative steps of the Krebs cycle and permit carbon incorporation from acetate or fatty acids in many microorganisms. *Kangiella* is also lacking these enzymes, which demonstrates the inertness of gluconeogenesis in *Kangiella*. Phosphoenolpyruvate carboxylase was also missing in *Kangiella*. In the aspect of amino acid biosynthesis and metabolism, the glutamate synthase, L-alanine:2-oxoglutarate aminotransferase and threonine dehydratase are missing in all *Kangiella* genomes. *Kangiella* strains also lack GO:0008942 (F:nitrite reductase [NAD(P)H] activity) and GO:0042128 (P:nitrate assimilation), which suggests that *Kangiella* cannot perform assimilatory nitrate reduction. Many halophilic microorganisms accumulate ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidine carboxylic acid) to counteract heat, cold, desiccation, and high salinity. The operon with L-2,4-diaminobutyric acid acetyltransferase, 4-aminobutyrate aminotransferase, and ectoine synthase is missing in *Kangiella*.

High Extracellular-Protein/Peptide Degradation Potential of *Kangiella*

The low genomic redundancy in *Kangiella*, was observed in the present study (**Supplementary Table S3**). Surprisingly, only 21 (1.23%) orthologs with duplications in the core genome and five of them (KangOF4, 8, 9, 15, and 19) were annotated to be SignalP-fused peptidases (SignalP) that involved in proteolysis (GO:0006508). Our experimental results also demonstrated that *Kangiella* strains could grow solely on amino acids as the carbon and nitrogen source (**Figure 4**). In addition, the biomass of *Kangiella* was increased if tryptone in marine broth 2216 medium has been doubled to make a enriched marine peptone medium (**Figure 6**). Therefore, we proposed that extracellular protein degradation and amino acid utilization are essential and prominent features of *Kangiella*.

Although the genome sizes of *Kangiella* strains were smaller than those of the other genus in Oceanospirillales, the percentage of SignalP-fused peptidases in *Kangiella* strains (1.68–2.36%) was significantly higher than that in the other genomes belonging to the order Oceanospirillales (0.25–0.80%) (**Table 2** and **Supplementary Table S8**). Furthermore, *Kangiella* strains also exhibit diverse types of peptidases. Approximately 67–71

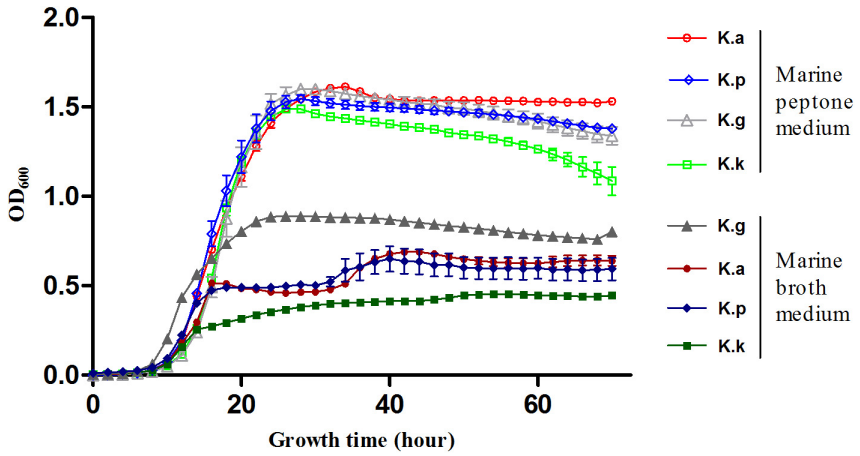


FIGURE 6 | Growth curve of four *Kangiella* strains cultured in marine peptone medium grouped in "Marine peptone medium" and marine broth 2216 grouped in "Marine broth medium." K.a, *K. aquimarina* DSM 16071; K.k, *K. koreensis* DSM 16069; K.g, *K. geojedonensis* KCTC 23420; and K.p, *K. profundi* FT102.

TABLE 2 | Comparison of peptidases in all the 23 genomes used in this study.

Sr. No.	Strain name	Total No. of Peptidase	Secretory peptidase	No. of Secretory Peptidase/MB
1	*Kangiella aquimarina* DSM 16071	142	56	20.85*
2	*Kangiella geojedonensis* KCTC 23420	140	50	20.04*
3	*Kangiella koreensis* DSM 16069	156	62	21.74*
4	*Kangiella profundi* FT102	142	54	20.40*
5	*Kangiella sediminilitoris* KCTC 23892	142	43	17.23*
6	*Alcanivorax borkumensis* SK2	108	20	6.41
7	*Alcanivorax dieselolei* B5	151	21	4.26
8	*Alcanivorax pacificus* W11-5	121	20	4.80
9	*Alcanivorax* sp. NBRC 101098	107	21	6.78
10	*Chromohalobacter salexigens* DSM 3043	123	15	4.06
11	*Gynuella sunshinyii* YC6258	156	29	4.48
12	*Halomonas campaniensis*	133	25	3.46
13	*Hahella chejuensis* KCTC 2396	219	54	13.25
14	*Halomonas elongata* DSM 2581	122	12	2.95
15	*Halomonas huangheensis*	147	23	4.83
16	*Halomonas* sp. KO116	132	21	4.06
17	*Halomonas* sp. R57-5	141	20	3.97
18	*Marinomonas mediterranea* MMB-1	120	15	3.20
19	*Marinomonas posidonica* IVIA-Po-181	111	15	3.85
20	*Marinomonas* sp. MWYL1	122	11	2.16
21	*Oleispira antarctica* RB-8	119	15	3.40
22	*Thalassolituus oleivorans* MIL-1	117	20	5.10
23	*Thalassolituus oleivorans* R6-15	115	22	5.84

"" means p-value < 0.01.*

peptidase subfamilies were predicted in *Kangiella*, of which 26 subfamilies were signal-fused peptidases, whereas only 55 subfamilies were identified in their non-*Kangiella* counterparts on average, and only 14 were signal-fused peptidases. Moreover, eight families of peptidases with SignalP (S8A, S9B, S9C, S41A, M19, M16B, and M38) were found to be remarkably abundant in *Kangiella*. We also identified three *Kangiella*-specific SignalP-fused protease families (S10, S46, and M28D).

S8A peptidases are the most abundant peptidase superfamily in *Kangiella* genomes. A gene cluster containing several tandem S8A superfamily peptidases (**Supplementary Figure S3**), including two orthologs (KangOF8 and 9) having multiple copies in the core genome was also identified in the *Kangiella* strains. KangOF9, which was annotated as a cold-active alkaline serine protease, was found with up to three copies per *Kangiella* genome. Interestingly, the paralogs of these proteases in KangOF9 showed

overall high sequence identity, except that an additional PKD domain was observed in each of the longer ones.

DISCUSSION

Genomic reduction is a significant characteristic of *Kangiella* among Oceanospirillales. It was reported that both loss of entire gene families and deletion of paralogs within multigene families could contribute to genome size reduction in marine bacteria (Lerat et al., 2005), and the smaller genome size of the marine cyanobacterial genus *Prochlorococcus* resulted from gene loss compared with its closely related genus *Synechococcus* (Luo et al., 2011). Our results showed a similar mechanism of genomic reduction, including de-redundancy and genus-specific loss of orthologs, in genus *Kangiella* compared to other non-Kangiellaceae Oceanospirillales.

It should be noted that in the background of global genomic reduction and de-redundancy of *Kangiella*, genomic plasticity and expansion also exist. Multigenome alignments (**Figure 7**) revealed that in the central position near the GC-skew reversion points of each *Kangiella* genome, there is an aberrant GC content genomic region (0.05–0.2 Mb) with poor colinearity. Such "hot block" for insertion and recombination was found to contain a lot of heavy metal and arsenic resistance genes in all *Kangiella* genomes, except *K. geojedonensis* KCTC 23420.

In the aspect of central metabolism, we predicted and experimentally confirmed that different *Kangiella* strains showed different defects in the capabilities of amino acid biosynthesis. For example, *K. profundi* FT102 lacks biosynthetic pathways of isoleucine, valine and phenylalanine and arginine, while *K. koreensis* DSM 16069 and *K. aquimarina* DSM 16071 only lack the first two. Interestingly, the phylogenetic tree based on 16S rRNA gene sequences showed that *K. aquimarina* DSM 16071 and *K. profundi* TF102 are more closely related than *K. koreensis* DSM 16069 (Choe et al., 2015). However, *K. aquimarina* DSM 16071 and *K. koreensis* DSM 16069 from the same sampling location showed the same capability of amino acids biosynthesis.

Furthermore, the subunits of sulfite reductase were strain-specifically missing in *K. profundi* FT102 (**Supplementary Figure S2**) as mentioned above, and we did not identify any functionally complementary gene in its genome.

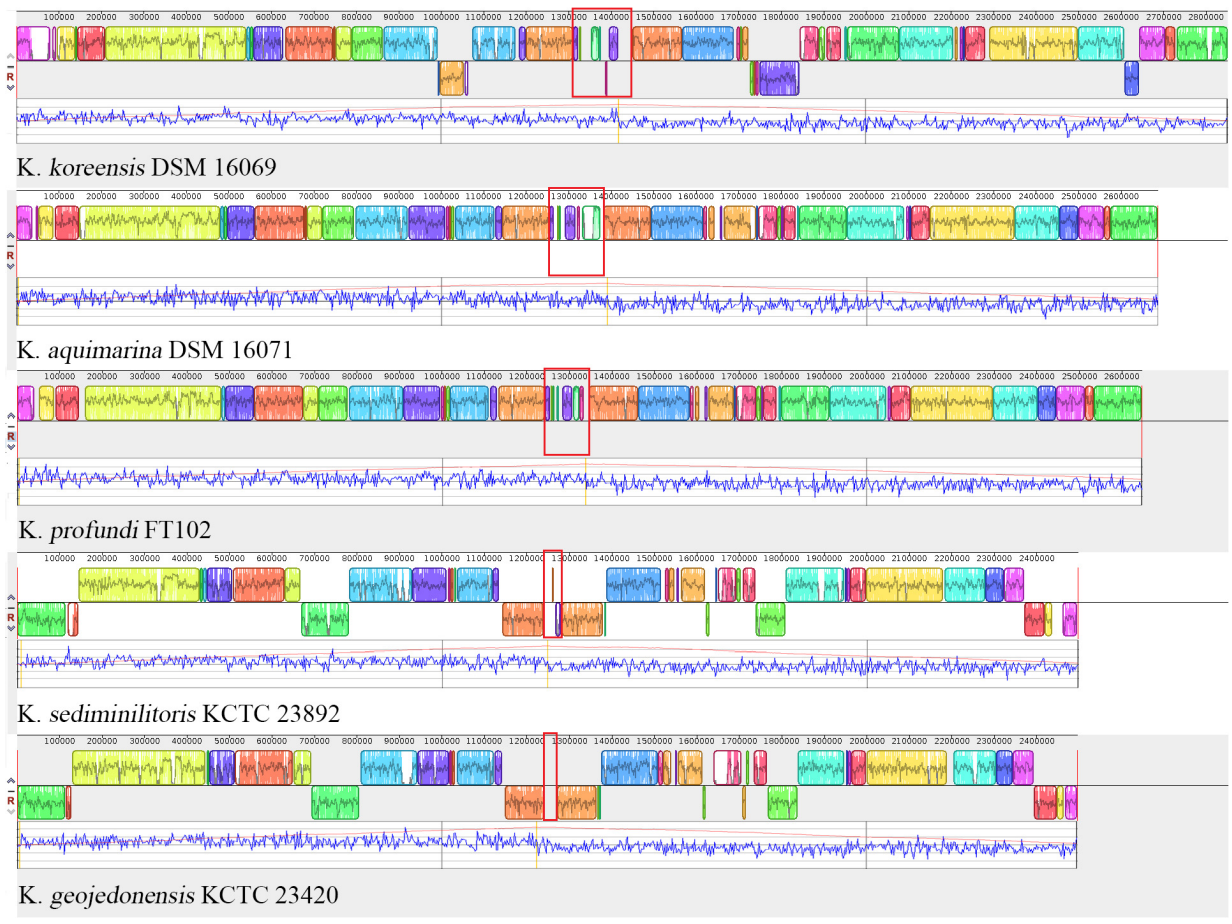


FIGURE 7 | Collinearity of the *Kangiella* genomes. Highlighted sites in red show the highly variable region of each genome. Blocks in the different colors represent different locally collinear blocks (LCBs). Local GC-skew (blue) and cumulative GC-skew (red) were plotted below LCBs, and the peak position of the cumulative GC-skew was highlighted in orange.

Actually, the sediment samples used to isolate *K. profundi* FT102 were collected from the southwest Indian InterRidge (Xu et al., 2015), which contain a large amount of sulfides and iron hydroxides (Tao et al., 2011, 2012). As long as the demand of sulfide has been satisfied from the environment, loss of those related genes seems to have no detrimental effect in *K. profundi* FT102.

In the aspect of proteolysis, our previous study confirmed the extracellular protease activities of four *Kangiella* strains (*K. profundi* FT102, *K. koreensis* DSM 16069, *K. aquimarina* DSM 16071, and *K. geojedonensis* KCTC 23420). In general, a positive correlation of the biomass and the extracellular protease activities was found when these strains were cultured in marine broth 2216 (Xu et al., 2018).

It is also worth mentioning that unlike *Aliikangiella marina* GYP-15^T, which represents another genus in Kangiellaceae, and was confirmed to utilize carbohydrates (Wang et al., 2015), *Kangiella* could not grow in defined media using glucose as the sole carbon source (data not shown). Taken together, proteolysis of extracellular proteins rather than carbohydrate utilization plays an important role in the life style of *Kangiella*.

Peptidase constitute a very large and complex group of enzymes that differ in properties such as substrate specificity, active site and catalytic mechanism, pH and temperature optima, and stability profiles (Jisha et al., 2013). In the *Kangiella* strains, there are at least 43 extracellular peptidases that could be classified into 26 subfamilies. Most of them are non-redundant, indicating wide substrates specificity of peptidases in *Kangiella*. Besides, among the merely 21 gene families with redundancy in core genome of *Kangiella*, five are secreted peptidases. In detail, four (KangOF4, 8, 9 and 15) of them belong to S8 peptidase family, and KangOF19 belongs to M1 peptidase family.

The S8A family peptidases are the most abundant in *Kangiella* genomes, including the tandem gene cluster constituted with the genes of KangOF190, 9 and 8 (**Supplementary Figure S3**). S8 peptidase is the second largest family of characterized serine peptidases, and they have catalytic triad consisted of aspartate, histidine and serine (Yamagata et al., 1994), which differs from that S1, S9 and S10 serine peptidases families. S8 peptidase family could be divided into two subfamilies, i.e., subtilisin (S8A) and kexin (S8B) according to MEROPS peptidase database. The S8 peptidases are mostly secreted and non-specific peptidases, and probably involved in uptaking of nutrition, which could support the obligate proteolytic lifestyle of *Kangiella.*

Moreover, S9 family serine peptidases are also abundant in *Kangiella* (**Supplementary Table S8**). However, no redundancy of S9 family was found in the core genome of *Kangiella*. Different from S8 peptidases, most members of the S9 peptidases family show strict substrate specificity (Rawlings et al., 1991). The tendency to be streamline in the genome of *Kangiella* was evident in keeping many heterogeneous genes of S9 family members instead of gene duplication.

More interestingly, the S10 peptidase family members were only identified and are conserved in *Kanagiella* genomes compared to other genus in Oceanospirillales. Different from most other serine peptidase families (S53 being the exception), the S10 peptidases are only active at acidic pH, e.g., carboxypeptidase Y with max activity at pH 5.5 (Jung et al., 1998). As the optimal pH for the growth of *Kangiella* is 7.0–8.0, the biological function of the putative S10 peptidases in *Kangiella* is still unclear. Quantitative analysis of the expression level of S10 peptidases in *Kangiella* challenged by different pH are still needed.

In the view of evolution, it is also interesting that the duplications of the peptidase families are still obvious in the *Kangiella* genomes under the background of genomic de-redundancy. We proposed that such gene duplications are partly driven by the obligate carbon source acquisition strategy of *Kangiella*. However, *K. sediminilitoris* KCTC 23892 was an exception, which showed no abovementioned duplication of KangOF8 or KangOF9 (**Supplementary Figure S3**), as well as carriers the fewest SignalP-fused peptidases (only 69–86% of other stains) (**Table 2**). The reason might be that the duplication is not only driven by carbon acquisition but also by nitrogen acquisition. Actually, a strain-specific gene coding ammonia transporter could be only identified in *K. sediminilitoris* as mentioned above. Absorbing inorganic nitrogen from the environment might alleviate its dependency on protein-derived ammonia, thus alleviating the selective pressure that drives the multiplication of extracellular peptidase genes.

CONCLUSION

In the present study, the complete genome of the novel strain *Kangiella profundi* FT102 was sequenced, and gaps in the scaffold genome of *K. aquimarina* DSM 16071 were filled using PCR. Five sequenced *Kangiella* genomes, including the abovementioned two, were utilized along with 18 non-*Kangiella* genomes in a comparative genomics approach to explore the metabolic capabilities of *Kangiella* and gain insight into genome reduction in *Kangiella*. We reported here that low genomic redundancy and *Kangiella*-lineage-specific gene loss are two key factors behind genome reduction in *Kangiella*. Furthermore, a highly enriched extracellular protein degradation system compared to that of any other non-*Kangiella* species was identified in *Kangiella*. Despite the low genome redundancy in the core genome (only 21 paralogs), five of these paralogs were signal-fused proteases. The absence of a complete pathway for carbohydrate metabolism led to the belief that proteases are not only abundant but also essential for *Kangiella* survival, demonstrating the extraordinary protein degradation capabilities of *Kangiella*.

AUTHOR CONTRIBUTIONS

JX and JW designed the experiments and analysis. JW and MN performed the computational analysis. YL conducted the experiments. JW, YL, and JX wrote the manuscript, in consultation with all other authors.

FUNDING

This study was supported by the National Basic Research Program of China ("973" Program 2014CB441503) and the National Natural Science Foundation of China (41676121 and 41376137).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01224/full#supplementary-material

FIGURE S1 | Percentage of the genes belonging to different cluster of orthologous genes (COG) categories in the orthologous gene families in pangenome of the five *Kangiella* strains. Genes belonging to the core genome, that are dispensable and that are specific to species are shown in blue, orange and gray, respectively. A: RNA processing and modification; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure, and biogenesis; K: Transcription; L: Replication, recombination, and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport, and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures; X: Mobilome: prophages and transposons; Z: Cytoskeleton.

FIGURE S2 | Verification by PCR amplification of four *Kangiella* species, including genes involved in sulfite reduction **(A,D)** corresponding to Kkor_0726 to Kkor_0729 in the *K. koreensis* DSM 16069, genes involved in phenylalanine biosynthesis **(B,E)** corresponding to Kkor_1584 and Kkor_1585 in *K. koreensis* DSM 16069, genes involved in arginine biosynthesis (**C,F**, shown in green color) corresponding to Kkor_0543 to Kkor_0548 in *K. koreensis* DSM 16069, and genes of cytochrome c oxidase subunits (**C,G**, shown in red color) corresponding to Kkor_0534 to Kkor_0542 in *K. koreensis* DSM 16069. Ka, *K. aquimarina* DSM 16071; Kk, *K. koreensis* DSM 16069; Kg, *K. geojedonensis* KCTC 23420; Kp, *K. profundi* FT102. The numbers in **(A–C)** represent genes listed in the following: 1: cysJ, FAD-binding domain protein; 2: sulfite reductase subunit beta; 3: cysH, phosphoadenosine phosphosulfate reductase; 4: cysG, uroporphyrin-III C-methyltransferase; 5: prephenate dehydratase; 6: phospho-2-dehydro-3-deoxyheptonate aldolase; 7: cytochrome-c oxidase, cbb3-type subunit I; 8: cytochrome-c oxidase, cbb3-type subunit II; 9: CcoQ/FixQ family Cbb3-type cytochrome c oxidase assembly chaperone; 10: cytochrome-c oxidase, cbb3-type subunit III; 11: cytochrome c oxidase accessory protein CcoG; 12: nitrogen fixation protein FixH; 13: cadmium-translocating P-type ATPase; 14: cbb3-type cytochrome oxidase assembly protein CcoS; 15: sulfite exporter TauE/SafE family protein; 16: N-acetylornithine carbamoyltransferase; 17: argininosuccinate synthase; 18: acetylornithine deacetylase; 19: acetylglutamate kinase; 20: N-acetyl-gamma-glutamyl-phosphate reductase; 21: argininosuccinate lyase; 22: oxygen-independent coproporphyrinogen III.

FIGURE S3 | Peptidase gene clusters (S8 FML) in the five *Kangiella* species. Arrows shown in black represent non-peptidase conserved genes. Arrows in the other colors represent the genes of different serine proteases belonging to the S8 family in the five *Kangiella* strains, including KangOF190 (orange), KangOF9 (green), KangOF8 (pink), and KangOF2260 (blue). Highlighted regions within the arrow (in red color) show the PKD domain.

TABLE S1 | Number of genes belonging to each cluster of orthologous genes (COG) category in all 23 strains used in the present study.

TABLE S2 | Orthologous families of the five *Kangiella* strains.

TABLE S3 | Orthologous families having paralogs in the core genome of the five *Kangiella* strains.

TABLE S4 | Cluster of orthologous genes (COG) distribution of the *Kangiella* pangenome.

TABLE S5 | Genomic features of the 23 Oceanospirillales strains used herein.

TABLE S6 | Orthologous families of the 23 Oceanospirillales strains.

TABLE S7 | *Kangiella*-specific missing GOs.

TABLE S8 | Number of SignalP-fused proteases identified in all 23 genomes.

MATERIAL S1 | ".ab1" files and PCR primers for genomic gap filling.

MATERIAL S2 | Raw gel electrophoresis maps and PCR primers corresponding to **Supplementary Figure S3**.

MATERIAL S3 | Peptidase genes predicted with MEROPS batch BLAST and SignalP prediction of 23 genomes corresponding to **Supplementary Table S8**.

REFERENCES

Ahn, J., Park, J.-W., Mcconnell, J. A., Ahn, Y.-B., and Häggblom, M. M. (2011). *Kangiella spongicola* sp. nov., a halophilic marine bacterium isolated from the sponge *Chondrilla nucula*. *Int. J. Syst. Evol. Microbiol.* 61, 961–964. doi: 10.1099/ijs.0.021733-0

Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* 5, 782–791. doi: 10.1038/nrmicro1747

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Cao, Y., Chastain, R. A., Eloe, E. A., Nogi, Y., Kato, C., and Bartlett, D. H. (2014). Novel psychropiezophilic *Oceanospirillales* species *Profundimonas piezophila* gen. nov., sp. nov., isolated from the deep-sea environment of the Puerto Rico trench. *Appl. Environ. Microbiol.* 80, 54–60. doi: 10.1128/AEM.02288-13

Chen, M., Xiao, X., Wang, P., Zeng, X., and Wang, F. (2005). *Arthrobacter ardleyensis* sp. nov., isolated from Antarctic lake sediment and deep-sea sediment. *Arch. Microbiol.* 183, 301–305. doi: 10.1007/s00203-005-0772-y

Choe, H., Kim, S., Oh, J., Nasir, A., Kim, B. K., and Kim, K. M. (2015). Complete genome of *Kangiella geojedonensis* KCTC 23420(T), putative evidence for recent genome reduction in marine environments. *Mar. Genomics* 24(Pt 3), 215–217. doi: 10.1016/j.margen.2015.05.015

Choi, Y. U., Kwon, Y. K., Ye, B. R., Hyun, J. H., Heo, S. J., Affan, A., et al. (2012). Draft genome sequence of *Marinobacterium stanieri* S30, a strain isolated from a coastal lagoon in Chuuk state in Micronesia. *J. Bacteriol.* 194:1260. doi: 10.1128/JB.06703-11

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636–4641. doi: 10.1093/nar/27.23.4636

DeLong, E. F., and Karl, D. M. (2005). Genomic perspectives in microbial oceanography. *Nature* 437, 336–342. doi: 10.1038/nature04157

Gao, F., and Zhang, C. T. (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 9:79. doi: 10.1186/1471-2105-9-79

Han, C., Sikorski, J., Lapidus, A., Nolan, M., Glavina Del Rio, T., Tice, H., et al. (2009). Complete genome sequence of *Kangiella koreensis* type strain (SW-125(T)). *Stand. Genomic Sci.* 1, 226–233. doi: 10.4056/sigs.36635

Jean, W. D., Huang, S.-P., Chen, J.-S., and Shieh, W. Y. (2012). *Kangiella taiwanensis* sp. nov. and *Kangiella marina* sp. nov., marine bacteria isolated from shallow coastal water. *Int. J. Syst. Evol. Microbiol.* 62, 2229–2234. doi: 10.1099/ijs.0.037010-0

Jisha, V. N., Smitha, R. B., Pradeep, S., Sreedevi, S., Unni, K. N., Sajith, S., et al. (2013). Versatility of microbial proteases. *Adv. Enzyme Res.* 1, 39–51. doi: 10.4236/aer.2013.13005

Jung, G., Ueno, H., and Hayashi, R. (1998). Proton-relay system of carboxypeptidase Y as a sole catalytic site: studies on mutagenic replacement of his 397. *J. Biochem.* 124, 446–450. doi: 10.1093/oxfordjournals.jbchem.a022133

Kim, J.-H., Ward, A. C., and Kim, W. (2015). *Kangiella chungangensis* sp. nov. isolated from a marine sand. *Antonie Van Leeuwenhoek* 107, 1291–1298. doi: 10.1007/s10482-015-0423-5

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Lee, S. Y., Park, S., Oh, T. K., and Yoon, J. H. (2013). *Kangiella sediminilitoris* sp. nov., isolated from a tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* 63, 1001–1006. doi: 10.1099/ijs.0.040691-0

Lerat, E., Daubin, V., Ochman, H., and Moran, N. A. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* 3:e130. doi: 10.1371/journal.pbio.0030130

Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955

Luo, H., Friedman, R., Tang, J., and Hughes, A. L. (2011). Genome reduction by deletion of paralogs in the marine cyanobacterium *Prochlorococcus*. *Mol. Biol. Evol.* 28, 2751–2760. doi: 10.1093/molbev/msr081

Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Rawlings, N. D., Barrett, A. J., and Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 44, D343–D350. doi: 10.1093/nar/gkv1118

Rawlings, N. D., Polgar, L., and Barrett, A. J. (1991). A new family of serine-type peptidases related to prolyl oligopeptidase. *Biochem. J.* 279(Pt 3), 907–908. doi: 10.1042/bj2790907

Romanenko, L. A., Tanaka, N., Frolova, G. M., and Mikhailov, V. V. (2010). *Kangiella japonica* sp. nov., isolated from a marine environment. *Int. J. Syst. Evol. Microbiol.* 60, 2583–2586. doi: 10.1099/ijs.0.017087-0

Tao, C., Li, H., Huang, W., Han, X., Wu, G., Su, X., et al. (2011). Mineralogical and geochemical features of sulfide chimneys from the 49°39′E hydrothermal field on the Southwest Indian Ridge and their geological inferences. *Chin. Sci. Bull.* 56, 2828–2838. doi: 10.1007/s11434-011-4619-4

Tao, C., Lin, J., Guo, S., Chen, Y. J., Wu, G., Han, X., et al. (2012). First active hydrothermal vents on an ultraslow-spreading center: Southwest Indian Ridge. *Geology* 40, 47–50. doi: 10.1130/G32389.1

Wang, G., Tang, M., Wu, H., Dai, S., Li, T., Chen, C., et al. (2015). *Aliikangiella marina* gen. nov., sp. nov., a marine bacterium from the culture broth of *Picochlorum* sp. 122, and proposal of *Kangiellaceae* fam. nov. in the order *Oceanospirillales*. *Int. J. Syst. Evol. Microbiol.* 65, 4488–4494. doi: 10.1099/ijsem.0.000601

Widdel, F., Boetius, A., and Rabus, R. (2006). "Anaerobic biodegradation of hydrocarbons including methane," in *The Prokaryotes: Ecophysiology and Biochemistry*, Vol. 2, eds M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (New York, NY: Springer), 1028–1049.

Xu, F. D., Li, X. G., Xiao, X., and Xu, J. (2015). *Kangiella profundi* sp nov., isolated from deep-sea sediment. *Int. J. Syst. Evol. Microbiol.* 65, 2315–2319. doi: 10.1099/ijs.0.000257

Xu, L., Wang, J., Lu, Y., and Xu, J. (2018). The growth, extracellular protease activity and the transcription of protease coding gene of four *Kangiella* strains affected by the major metal ions in the sea water. *Microbiol. China* 45, 1–13.

Yamagata, H., Masuzawa, T., Nagaoka, Y., Ohnishi, T., and Iwasaki, T. (1994). Cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor. *J. Biol. Chem.* 269, 32725–32731.

Yoon, J. H., Kang, S. J., Lee, S. Y., Lee, J. S., and Oh, T. K. (2012). *Kangiella geojedonensis* sp. nov., isolated from seawater. *Int. J. Syst. Evol. Microbiol.* 62, 511–514. doi: 10.1099/ijs.0.029314-0

Yoon, J. H., Oh, T. K., and Park, Y. H. (2004). *Kangiella koreensis* gen. nov., sp. nov. and *Kangiella aquimarina* sp. nov., isolated from a tidal flat of the Yellow Sea in Korea. *Int. J. Syst. Evol. Microbiol.* 54, 1829–1835. doi: 10.1099/ijs.0.63156-0

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wang, Lu, Nawaz and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 February 2018
doi: 10.3389/fmicb.2018.00251

Recovering Genomics Clusters of Secondary Metabolites from Lakes Using Genome-Resolved Metagenomics

Rafael R. C. Cuadrat[1,2,3]*, *Danny Ionescu*[2], *Alberto M. R. Dávila*[4] and *Hans-Peter Grossart*[2,5]

[1] Bioinformatics Core Facility, Max Plank Institute for Biology of Ageing, Köln, Germany, [2] Experimental Limnology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Stechlin, Germany, [3] Berlin Center for Genomics in Biodiversity Research, Berlin, Germany, [4] Computational and Systems Biology Laboratory, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil, [5] Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany

Metagenomic approaches became increasingly popular in the past decades due to decreasing costs of DNA sequencing and bioinformatics development. So far, however, the recovery of long genes coding for secondary metabolites still represents a big challenge. Often, the quality of metagenome assemblies is poor, especially in environments with a high microbial diversity where sequence coverage is low and complexity of natural communities high. Recently, new and improved algorithms for binning environmental reads and contigs have been developed to overcome such limitations. Some of these algorithms use a similarity detection approach to classify the obtained reads into taxonomical units and to assemble draft genomes. This approach, however, is quite limited since it can classify exclusively sequences similar to those available (and well classified) in the databases. In this work, we used draft genomes from Lake Stechlin, north-eastern Germany, recovered by MetaBat, an efficient binning tool that integrates empirical probabilistic distances of genome abundance, and tetranucleotide frequency for accurate metagenome binning. These genomes were screened for secondary metabolism genes, such as polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS), using the Anti-SMASH and NAPDOS workflows. With this approach we were able to identify 243 secondary metabolite clusters from 121 genomes recovered from our lake samples. A total of 18 NRPS, 19 PKS, and 3 hybrid PKS/NRPS clusters were found. In addition, it was possible to predict the partial structure of several secondary metabolite clusters allowing for taxonomical classifications and phylogenetic inferences. Our approach revealed a high potential to recover and study secondary metabolites genes from any aquatic ecosystem.

Keywords: metagenomics 2.0, PKS, NRPS, freshwater, environmental genomics

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Hans Uwe Dahms, Kaohsiung Medical University, Taiwan
Steven Singer, Lawrence Berkeley National Laboratory (LBNL), United States

***Correspondence:**
Rafael R. C. Cuadrat
rafaelcuadrat@gmail.com

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *24 October 2017*
Accepted: *31 January 2018*
Published: *20 February 2018*

Citation:
Cuadrat RRC, Ionescu D, Dávila AMR and Grossart H-P (2018) Recovering Genomics Clusters of Secondary Metabolites from Lakes Using Genome-Resolved Metagenomics. Front. Microbiol. 9:251. doi: 10.3389/fmicb.2018.00251

INTRODUCTION

Metagenomics, also known as environmental genomics, describes the study of a microbial community without the need of *a priori* cultivation in the laboratory. It has the potential to explore uncultivable microorganisms by accessing and sequencing their nucleic acid (Rodríguez-Valera, 2004). In recent years, due to decreasing costs of DNA sequencing–metagenomic databases

(Vincent et al., 2016) (e.g., MG-RAST) have rapidly grown and archive billions of short read sequences (Meyer et al., 2008). Many metagenomic tools and pipelines were proposed to better analyse these enormous datasets (Huson et al., 2007). Additionally, these tools allow to (i) infer ecological patterns, alfa-, and beta-diversity and richness (Caporaso et al., 2010); (ii) assemble environmental contigs from the reads (Li et al., 2015), and more recently, (iii) recover draft genomes from metagenomic bins (Strous et al., 2012; Wu et al., 2014; Kang et al., 2015). By recovering a high number of draft genomes from these so far uncultivable organisms, it is now possible to screen for new genes and clusters, unlocking a previously underestimated metabolic potential such as secondary metabolite gene clusters by using a metagenomic approach called Metagenomics 2.0 (McMahon, 2015).

Lake Stechlin is an oligo-mesotrophic ecosystem located in the Brandenburg-Mecklenburg lake district, northeast Germany, in a glacial region with numerous dead-ice lakes greatly varying in water quality according to the European Water Framework Directive (WFD) (Dadheech et al., 2014). The lake is monitored since 1959 for several physical, chemical and biological variables, ranging from phyto- and zooplankton species composition to cell number and biomass calculations (Jost Casper, 1985), but only since the late 20th, early 21st century mass developments of cyanobacteria were found (Salmaso and Padisák, 2007).

Many species of Cyanobacteria can form blooms, and it has been estimated that 25–75% of cyanobacterial blooms are toxic, specially in warm water (Bláhová et al., 2007, 2008) and the frequency of these blooms has also risen in Lake Stechlin in recent decades (Jost Casper, 1985) whereby its toxicity seems to increase in the lake (Dadheech et al., 2014).

Polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS) are two families of modular mega-synthases, both are very important for the biotechnological and pharmaceutical industry due to their broad spectrum of products, spanning from antibiotics and antitumor drugs to food pigments and also harmful toxins, like Anatoxin-a, and Microcystins (Dadheech et al., 2014).

Both families of mega-synthases act in an analogous way, producing polyketides (using acyil-coA monomers) and peptides (using aminoacyl monomers), and they are broadly distributed in many taxonomical groups, ranging from bacteria (alphaproteobacteria, cyanobacteria, actinobacteria) to fungi (Gokhale et al., 2007; Koglin and Walsh, 2009).

PKS enzymes can be classified in types (I, II, and III), whereby type I can be further classified into modular or iterative classes. The iterative PKS use the same domain many times, iteratively, to synthetize the polyketide. The modular PKS are large multi-domain enzymes in which each domain is used only once in the synthesis process (Cane et al., 1998; Lal et al., 2000). The production of the polyketide follows the co-linearity rule, each module being responsible for the addition of one monomer to the growing chain (Minowa et al., 2007).

Type I PKS are characterized by multiple domains in the same open reading frame (ORF) while in type II each domain is encoded in a separate ORF, acting interactively (Sun et al., 2012). Type III is also known as Chalcone synthase and has different evolutionary origin from type I and II (Austin and Noel, 2003). Type III PKSs are self-contained enzymes that form homodimers. Their single active site in each monomer catalyzes the priming, extension, and cyclization reactions iteratively to form polyketide products (Austin and Noel, 2003). Hybrid PKS/NRPS and NRPS/PKS are also modular enzymes, encoding lipopeptides (hybrid between polyketides and peptides) and occur in bacterial as well as fungal genomes (Fisch, 2013; Masschelein et al., 2013; Mizuno et al., 2013).

PKS and NRPS are very well explored in genomes from cultivable organisms, mainly *Actinomycetes* (Komaki et al., 2014) and *Cyanobacteria* (Micallef et al., 2015). Recently, by using a metagenomic approach, studies have demonstrated the presence of these metabolite-genes in aquatic environments, as for example, Brazilian coastal waters (in free living and particle-associated bacteria) and from the microbiomes of Australian marine sponges (Woodhouse et al., 2013). However, there are few metagenomic studies whose scope is to find these gene families in freshwater environments where most studies are based on isolation approaches (Silva-Stenico et al., 2011; Zothanpuia et al., 2016).

In addition, due to the rather large size of genes involved in these pathways, yet, it is not possible to recover the full genes by using traditional read-based metagenomics or the single sample assembly approach. Most of the studies aim to solely find specific domains, like Keto-synthase (KS) in PKS and Condensation domain (C) in NRPS, due to the high conservation of these domains (Selvin et al., 2016).

Thus, it is important to develop new methods to recover full sequences of these families, especially because of their modular nature, i.e., the final compound can only be inferred with information about all the modules and domains based on the co-linearity role of polyketides and non-ribosomal peptides (Minowa et al., 2007). By detecting the full sequences, the approach can be used, for example, for environmental long-term monitoring of the toxicity potential of harmful algal blooms (HABs) (Meriluoto et al., 2017).

We used a metagenomics 2.0 approach to overcome these limitations and improve the screening for secondary metabolism genes and clusters while evaluating the potential of microbial communities for future research on potential drugs and toxin production. This study aims to (i) generate draft genomes from Lake Stechlin; (ii) to screen these genomes for new complete multi-modular enzymes from PKS and NRPS families, exploring their diversity and phylogeny.

MATERIALS AND METHODS

Sampling and Sequencing

A total of 26 metagenomic samples from Lake Stechilin, north-eastern Germany were used.

Water was collected as metagenomic samples on several occasions (April, June 2013, July 2014, August 2015) in sterile 2 L Schott bottles from Lake Stechlin (53°9′5.59N, 13°1′34.22E). All samples, except those from August 2015, were filtered through 5 μm and subsequently 0.2 μm pore-size filters. The samples collected in August 2015 were not size-fractionated and directly filtered on a 0.2 μm pore size filter. Genomic DNA was extracted

using a phenol/chloroform protocol as described in Ionescu et al. (2012) and was sent for sequencing.

Sequencing was conducted at MrDNA (Shallowater, Texas) on an Illumina Hiseq 2500, using the V3 chemistry, following, fragmentation, adaptor ligation and amplification of 50 ng genomic DNA from each sample, using the Nextera DNA Sample Preparation Kit.

Table S1 shows the general information about the 26 samples used in this study.

Environmental Draft Genomes

Briefly, all samples were pre-processed by Nesoni (https://github.com/Victorian-Bioinformatics-Consortium/nesoni) to remove low quality sequences and to trim adaptors, and afterwards assembled together using MegaHIT (default parameters) (Li et al., 2015). The reads from each sample were mapped back to these assembled contigs using BBMAP (https://sourceforge.net/projects/bbmap/) and then all data was binned using MetaBAT (Kang et al., 2015) to generate the draft genomes. The completeness and taxonomical classification were checked using CheckM (Parks et al., 2015).

Screening Secondary Metabolism Genes and Phylogenetic Analysis of NRPS and PKS Domains

DNA fasta files of the generated bins (288) were submitted to a locally installed version of Anti-SMASH (–clusterblast –smcogs –limit 1500) (Weber et al., 2015). Using in-house ruby scripts, the domains from PKS and NRPS were parsed. The PKS KS domains and NRPS C domains were submitted to NAPDOS for classification (Ziemert et al., 2012). In addition, all the KS and C domains (trimmed by NAPDOS) were submitted to BLASTP against RefSeq database (O'Leary et al., 2016), using the default parameters. The 3 best hits of each domain were extracted and added to the original multi-fasta file with the environmental domains. The full set of KS and C domains (from bins and references obtained by the blast on RefSeq database) was submitted for NAPDOS for the phylogenetic analysis. The resulting alignments and trees were exported, then trees were manually checked and annotated.

Relative Abundance of Bins in Each Sample

The reads from each sample were mapped (using BBMAP) against each bin fasta file and an in-house ruby parser script was used to calculate the relative abundance of each bin in each sample, normalizing the read counts by the number of reads of each sample. The table with the results was loaded into STAMP (Parks et al., 2014) in order to analyse the significant differences of bin abundance over the samples.

RESULTS

All the sequences generated for this study were submitted to ENA under accession numbers: PRJEB22274 and PRJEB7963.

Environmental Draft Genomes Obtained (Bins)

Metagenomic binning resulted in 288 draft environmental genomes (called bins in this study). Of these, 45 had a predicted completion level higher than 75% according to CheckM.

Table S2 shows the general information about each bin, including completeness, genome size, number of open read frames (ORFs) and taxonomical classifications (from CheckM).

Screening Secondary Metabolism Genes and Phylogenetic Analysis

By using Anti-SMASH, at least one secondary metabolite gene cluster was found in 121 of the bins, totaling 243 clusters and 2200 ORFs. From these 243 clusters, 125 (51.4%) were classified in the Terpene and 35 (14.40%) in the bacteriocin pathway. In addition, a total of 18 NRPS, 6 type I PKS and 3 hybrid PKS/NRPS clusters were found in 15 different bins (**Figure 1A**). The latest 3 obtained pathway clusters are the main focus of our study.

Figure 1B shows the taxonomical classification at phylum level for the bins showing NRPS, type I PKS and hybrid clusters. **Table S3** shows the distribution of all clusters in all bins.

A total of 43 condensation (C) domains were obtained from NRPS clusters. All these sequences were submitted to NAPDOS analysis. **Figure 2A** shows the classification of C domains into classes. Most of the sequences were classified as LCL domains (58%). This kind of domain catalyzes the formation of a peptide bond between two L-amino acids.

The screening for type I PKS resulted in 9 KS domain sequences. Most of them are classified as modular type I PKS (56%). All of them were submitted to NAPDOS and classified into 4 different classes (**Figure 2B**).

All the KS and C domains were also submitted to similarity analysis by using BLASTP against RefSeq database (**Tables S4**, **S5**) and the best 3 hits of each sequence were extracted and used for phylogenetic analyses with NAPDOS. The trees for C and KS domains are shown in **Figures 3** and **4**, respectively.

Relative Abundance of Bins on Each Sample

The relative abundance of all the bins in each sample was estimated by mapping the reads from each sample against the assembled bins. **Table S6** shows the normalized bin abundance for every sample.

Due to the differences between the filtration methods, we decided to classify the samples in 3 groups: particle associated samples (PA)–filtered on 5 μm membranes (and also samples from aggregates), free-living samples (FL)–pre-filtered through 5.0 μm membranes and subsequently filtered on 0.22 μm membranes, and non-size fractionated samples (NSF)–filtered direct on 0.22 μm membranes (without previous filtering) retaining the whole bacterial community. In total, we obtained 7 samples in the PA group, 5 in the FL group, and 14 in the NSF group.

The table with the relative abundance of the bins in all samples was loaded on STAMP and an ANOVA test was conducted followed Games-Howell *POST-HOC* test and

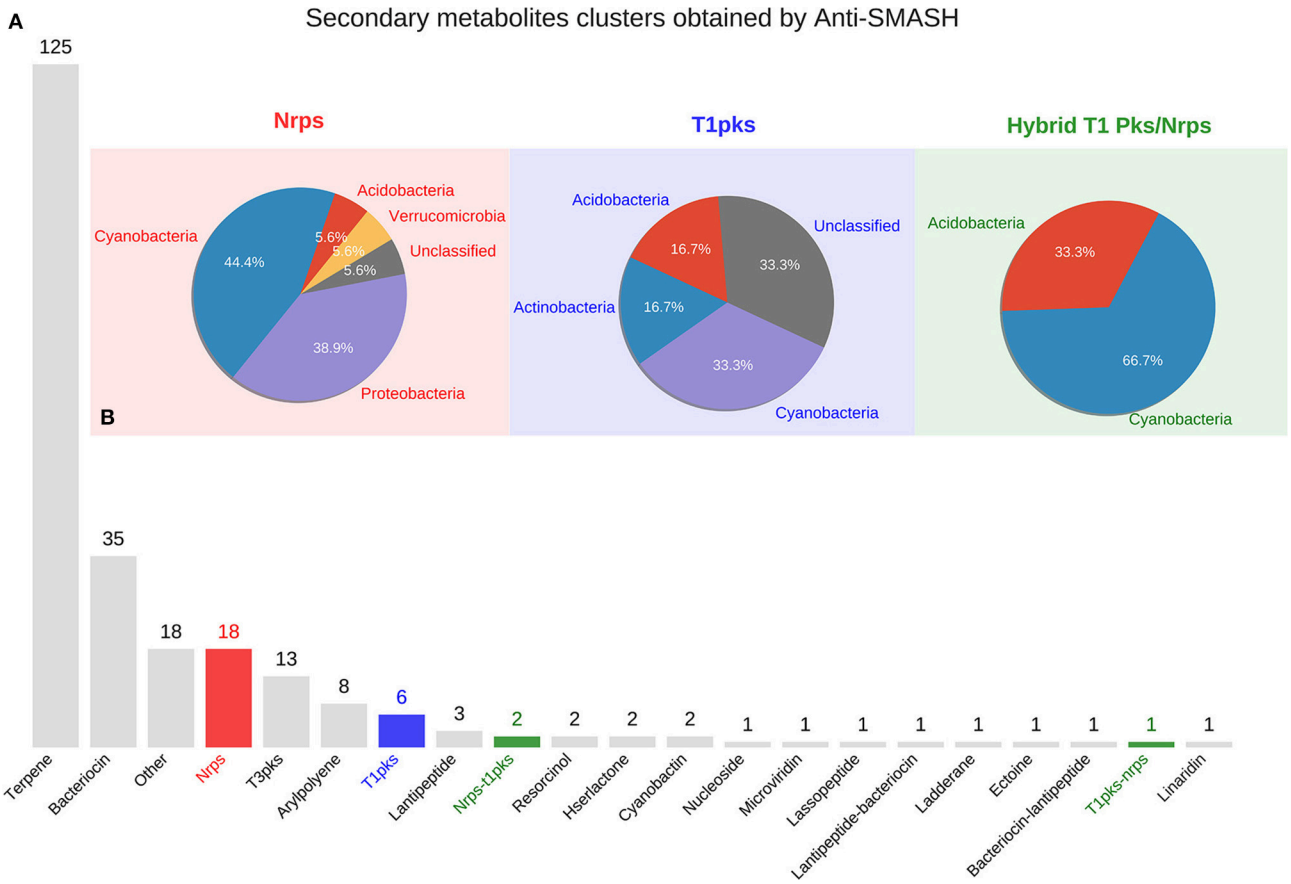


FIGURE 1 | (A) Abundance of secondary metabolite cluster types obtained with Anti-SMASH in the recovered 288 bins (environmental genomes). **(B)** Taxonomical classification of bins (Phyla) in which NRPS, PKS, and Hybrid PKS/NRPS clusters were found. Red bar and pie: NRPS; blue bar and pie: Type I PKS; green bars and pie: Hybrid clusters (NRPS-PKS and PKS-NRPS).

Benjamini-Hochberg FDR correction. **Table S7** shows 158 bins for which the difference in relative abundance was statistically significant ($p < 0.05$) between the 3 groups (FL, PA, and NSF).

From the 15 bins containing NRPS and/or type I PKS clusters, only 4 showed significant difference between the 3 groups. Bins 1 and 2 are more abundant in PA samples and bins 193 and 235 are more abundant in the FL samples.

Exploring NRPS, Type I PKS, and Hybrid Clusters from Draft Genome Bins

We highlight 3 bins (with less than 35% contamination and more than 70% completeness) out of the 15 obtained type I PKS and/or NRPS and explore their clusters.

In bin 34 (*Pseudomonas*, 98.28% completeness) it was possible to retrieve 7 clusters, including 3 NRPS clusters (**Figure 5**) and 2 bacteriocin clusters.

In cluster 2 (ctg181), multiple domains of NRPS (with the 3 minimal modules) and regulatory genes were identified, e.g., smCOG: SMCOG1057 (TetR family transcriptional regulator) (**Figure 5**, in green arrows).

All the clusters show a high similarity with *Pseudomonas* proteins. Cluster 2 has a similarity of 92% with *Pseudomonas synxantha* bg33r, conserving also the gene synteny.

The C domain sequences were submitted to NAPDOS analysis and 2 were classified as belonging to the syringomycin pathway and the LCL class, and one was classified as belonging to the microcystin pathway or and the DCL class (link an L-amino acid to a growing peptide ending with a D-amino acid).

In cluster 3 (ctg415) (**Figure 5**), in addition to the NRPS domains, the following transporter related genes were found: smCOG: SMCOG1288 (ABC transporter related protein) and SMCOG1051 (TonB-dependent siderophore receptor) (blue narrows). Nevertheless, this cluster is not complete and just one C domain was found (LCL class), which was also classified to the syringomycin pathway.

Cluster 6 (ctg857–**Figure 5**) shows many NRPS domains, regulatory factors and transporters genes, including drug resistance genes, e.g., *SMCOG1005* (drug resistance transporter, EmrB/QacA), SMCOG1044 (ABC transporter, permease protein) and SMCOG1051 (TonB-dependent siderophore receptor) (blue arrows). Two C domains from this cluster were classified as belonging to the heterocyclization class. This class catalyzes

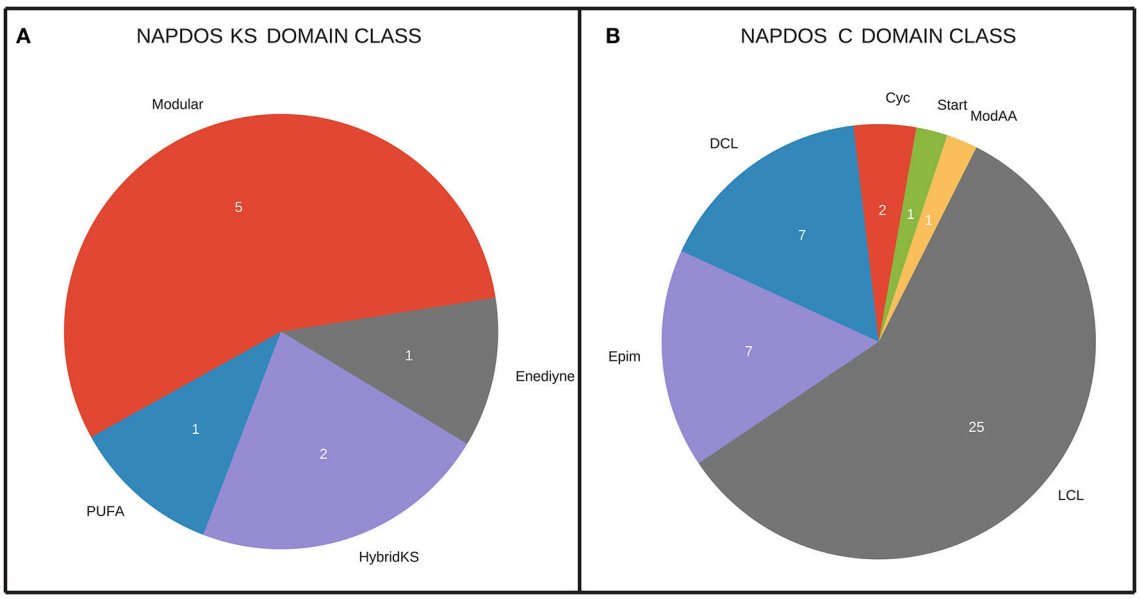


FIGURE 2 | (A) NAPDOS classification of the NRPS KS domain. Modular: possess a multidomain architecture consisting of multiple sets of modules; hybridKS: are biosynthetic assembly lines that include both PKS and NRPS components; PUFA: Polyunsaturated fatty acids (PUFAs) are long chain fatty acids containing more than one double bond, including omega-3-and omega-6- fatty acids; Enediyne: a family of biologically active natural products. The enediyne core consists of two acetylenic groups conjugated to a double bond or an incipient double bond within a nine- or ten-membered ring. **(B)** NAPDOS classification of NRPS C domain. Cyc, cyclization domains catalyze both peptide bond formation and subsequent cyclization of cysteine, serine or threonine residues; DCL, link an L-amino acid to a growing peptide ending with a D-amino acid; Epim, epimerization domains change the chirality of the last amino acid in the chain from L- to D-amino acid; LCL, catalyze formation of a peptide bond between two L-amino acids; modAA, appear to be involved in the modification of the incorporated amino acid; Start, first module of a Non-ribosomal peptide synthase (NRPS).

both peptide bond formation and subsequent cyclization of cysteine, serine or threonine residues (Di Lorenzo et al., 2008). Both domains were classified in the pyochelin pathway by NAPDOS. The phylogenetic tree (**Figure 3**) confirms both the functional and taxonomical classification (confidence value 100).

In **bin 193** (Mycobacterium, 73.37 % completeness) a type I PKS cluster was identified (ctg514) (**Figure 6**). Five PKS domains were retrieved, including the minimal core from one of the ORFs on this contig. The KS domain BLASTP result shows 82% (and 99% coverage) similarity with *Mycobacterium kansasii*. The NAPDOS analysis from the KS domain suggests that it could be a modular (epothilone pathway) or iterative type I PKS similar to the calicheamicin pathway. However, by using the phylogenetic analysis it was clustered as the iterative clade (confidence value 98.4) together with the *M. kansasii* sequence (confidence value 100) (**Figure 4**).

Two further clusters were recovered: one type III PKS and one unclassified one. All clusters show similarity with the *Mycobacterium* clusters.

Bin 131 (unclassified bacteria by CheckM) has 84.09% of completeness reported by CheckM. In this bin it was possible to find one cluster and 3 domains of type I PKS (KS, AT, and KR) (**Figure 6**). The KS domain was classified by NAPDOS as belonging to the maduropeptin and neocarzinostatin pathways. Using clusterblast inside Anti-Smash it was not possible to find any similar cluster, but using BLASTP it was possible to find similarity with the cyanobacteria *Microcystis aeruginosa* (64% identity and 99% coverage on BLASTP search). In the phylogenetic tree, it was clustered within the enediynes clade (**Figure 4**) and also with *M. aeruginosa* (confidence value 99).

In addition, there were 12 more bins with NRPS or type I PKS clusters, but with less than 70% of completeness or more than 35% contamination. The bins 1 (69.54% completeness) and 2 (16.52%) were classified as the genus *Anabaena*, showing NRPS and hybrid NRPS-type 1 PKS, respectively. Bins 6 (39.66% completeness), 7 and 8 were classified as the genus *Planktothrix* and show a high diversity of secondary metabolites: 3 NRPS clusters, one type I PKS and 2 NRPS-PKS hybrid (bin 6), 2 NRPS (bin 7) and 2 NRPS and one type I PKS (bin 8). The bin 8 also shows a microviridin cluster.

Bins 73 and 217 are classified as *Acidobacteria* showing PKS and NRPS, respectively. Bin 235 (*Burkholderiaceae* family) shows 3 NRPS clusters and bin 13 (*Comamonadaceae* family, also from *Burkholderiales* order) shows one NRPS cluster.

Bin 78 is classified as *Verrucomicrobiaceae* shows 1 NRPS cluster. Additionally, there is an unclassified *Archaea* (NRPS cluster) and one bin without any classification (bin 136, type I PKS).

The Anti-Smash results for all bins are available in the Supplemental Information (**Table S1**).

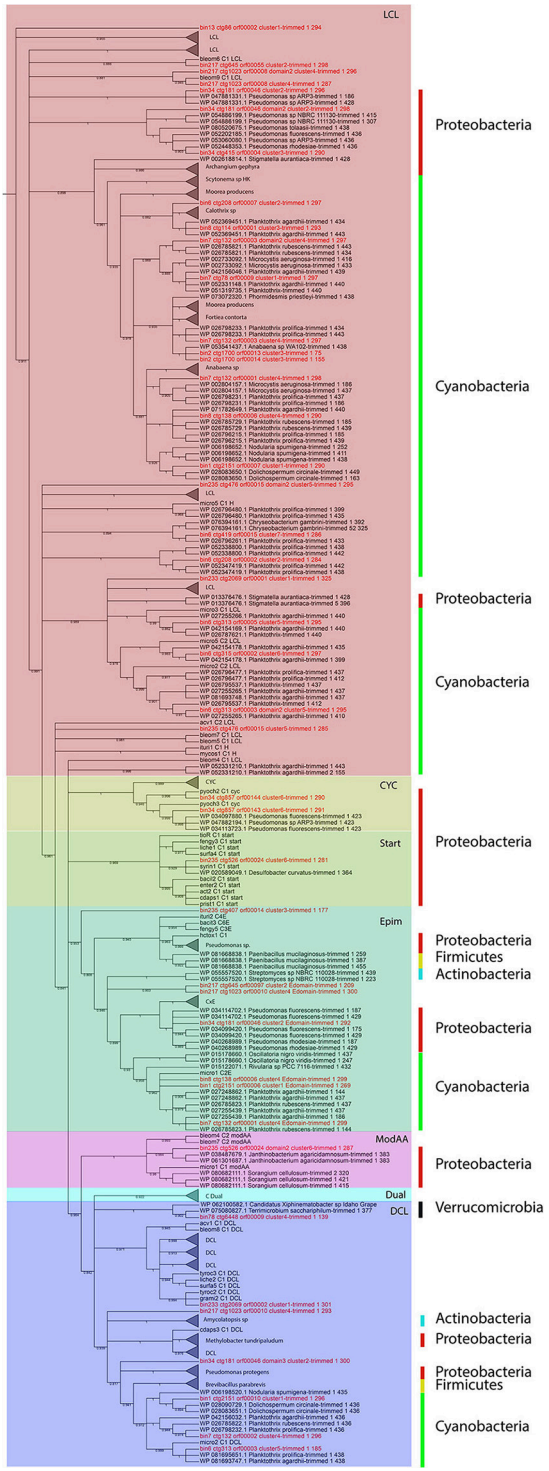


FIGURE 3 | NAPDOS phylogenetic tree of C domains (environmental domains, the top 3 blast results on RefSeq and the NAPDOS reference sequences). The shadow colors represent the domain classifications (LCL, CYC, Start domains, EPIM, ModAA, Dual, and DCL). The sidebars represent phyla (Proteobacteria, Cyanobacteria, Firmicutes, Actinobacteria, Verrucomicrobia). All the sequences from environmental bins are in red.

DISCUSSION

The field of metagenomics has generated a vast amount of data in the last decades (National Research Council US., 2007). Most of the data is poorly annotated and may show little quality control when loaded into the public databases, hence awaiting a more in-depth analysis (Gilbert et al., 2011). There are many open challenges in this field, e.g., (i) the lack of representative genomic databases from uncultivable organisms to be used in a similarity-based annotation procedure; (ii) high confidence assembly of short reads from species-rich samples; (iii) obtaining high enough coverage for every organism in the sample, including those with a low abundance, etc. (Teeling and Glöckner, 2012). Recently, some new algorithms have been proposed to overcome these limitations and to obtain partial or near complete genomes from environmental samples, e.g., MetaBat (Kang et al., 2015) and MetaWatt (Strous et al., 2012). Most of them require many samples and high coverage sequencing per sample as an input. Recently, studies have been done to recover genomes even from rare bacteria (Albertsen et al., 2013). The term Metagenomics 2.0 was introduced to describe this new generation of metagenomic analysis by Katherine McMahon (McMahon, 2015) and most of the studies using this approach have been conducted to reveal ecological interactions and networks (Sangwan et al., 2016; Vanwonterghem et al., 2016).

In this study, we recovered 288 environmental draft genomes using 26 samples from Lake Stechlin, a temperate oligo-mesotrophic lake. One of the advantages of this approach is to enable the recovery of large genomic clusters, especially the megasynthases clusters of the secondary metabolism, e.g., involved in biosynthesis of antibiotics, including its regulatory and transporter genes. Here, we have used the Anti-SMASH and NAPDOS pipelines to identify, annotate, classify and to carry out the phylogenetic analysis of a total of 243 clusters of known secondary metabolites. To our knowledge, this is the first study using the metagenomics 2.0 approach to recover megasynthases clusters. A number of previous studies had been conducted using a traditional PCR based screening (Amos et al., 2015) and shotgun metagenomic approach (Foerstner et al., 2008; Cuadrat et al., 2015) exploring the abundance and diversity of individual genes and domains, but these studies are missing the genomic context. By obtaining the entire genomic context it is possible, in future studies, to clone and to do heterologous expression for all the genes, including promoters and transporters.

The phytoplankton from Lake Stechlin has been intensively investigated since 1959, allowing researchers to follow long-term trends, for example, mass developments of cyanobacteria, found only since the late 20th, early 21st century (Salmaso and Padisák, 2007), coinciding with an increase in surface temperature (by 0.37°C per decade to an annual mean of 11.3 ± 0.5°C) (Kirillin, 2013).

The increase in water temperature has been paralleled by the occurrence of HABs, favoring potentially toxic cyanobacteria (Paerl and Huisman, 2008) such as *Microcystis*, which has been found in the Stechlin Lake since 2011 (Dadheech et al., 2014).

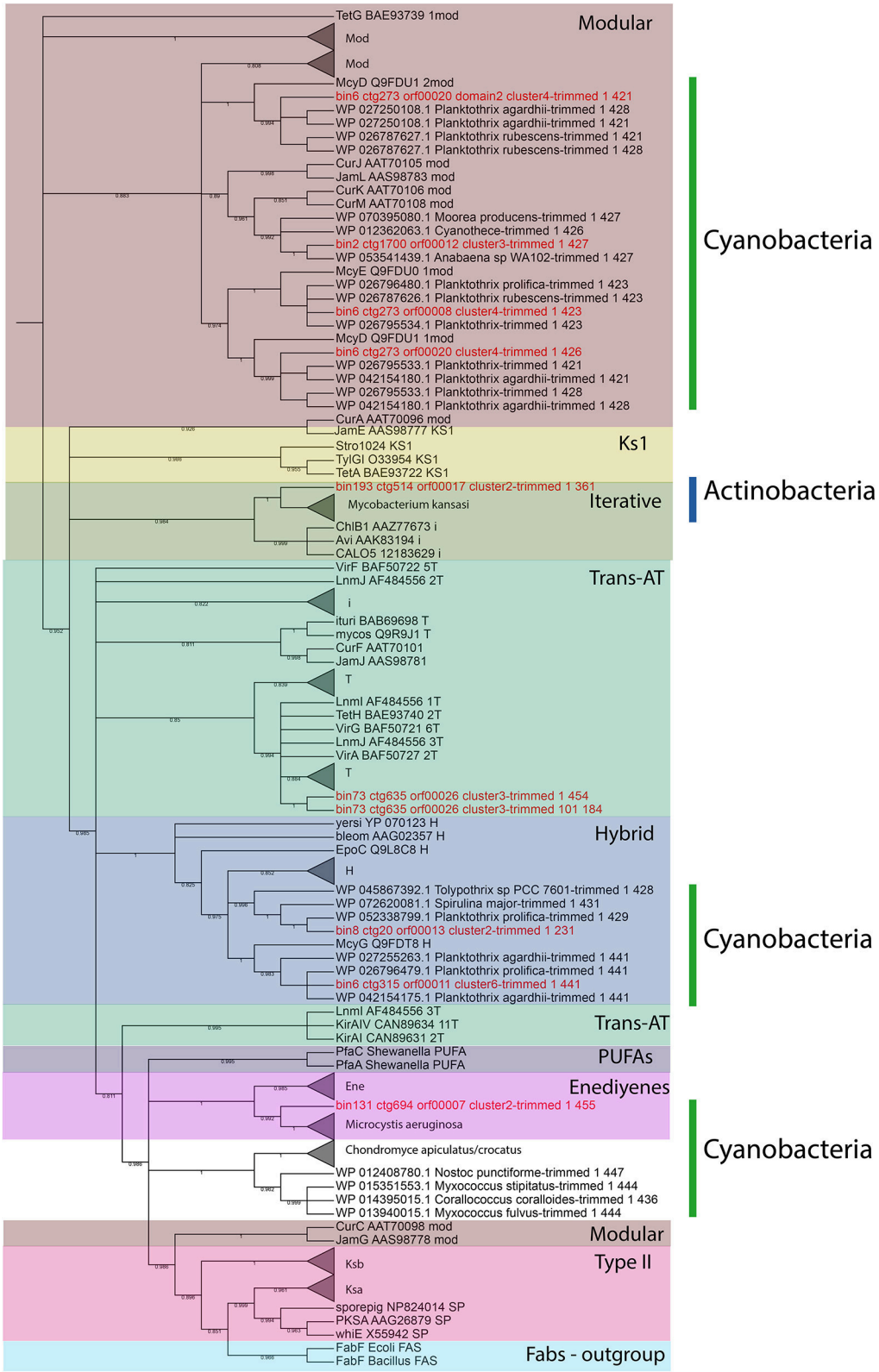


FIGURE 4 | NAPDOS tree of KS domains (environmental domains, the top 3 blast results on RefSeq and the NAPDOS reference sequences). The shadow colors represent the domain classifications (Modular, KS1, Iterative, Trans-AT, Hybrid, PUFA, Enediyenes, Type II, and Fabs, Fatty acid synthase). The sidebars represent phyla (Cyanobacteria and Actinobacteria). All the sequences from environmental bins are in red.

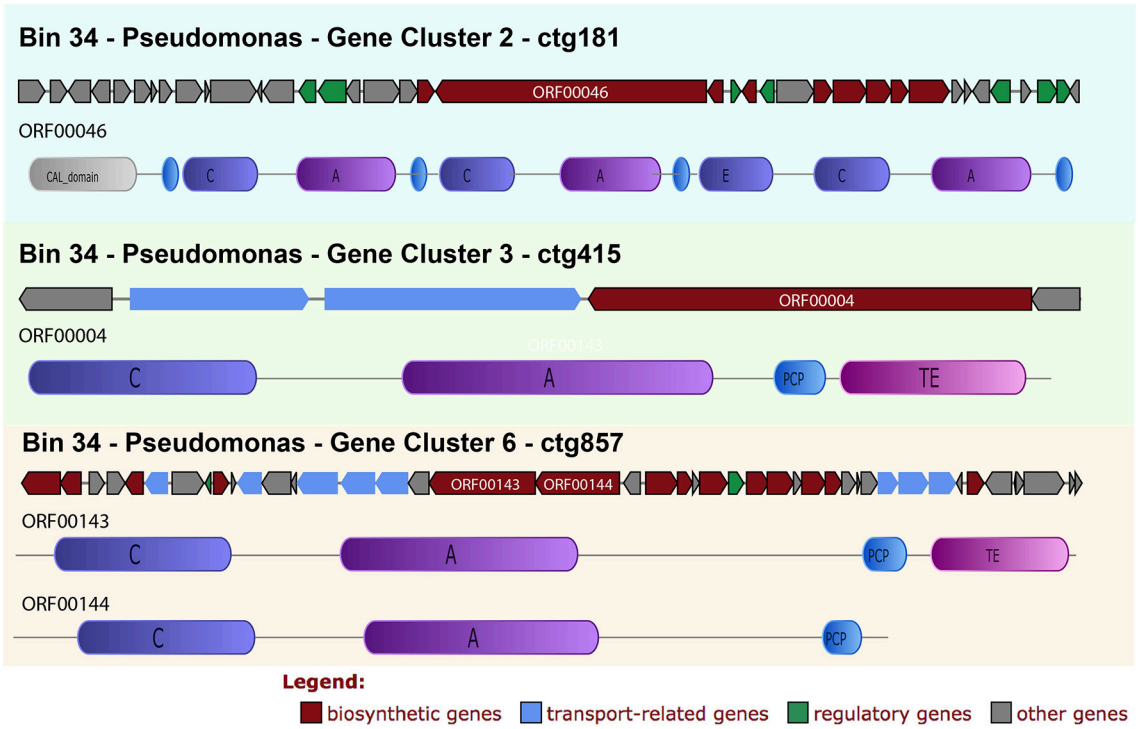


FIGURE 5 | Bin 34, NRPS clusters detailed annotation and synteny. The synteny of the clusters with a functional classification for each ORF is given. In addition, for the NPRS biosynthetic ORFS the domain annotations are given. CAL, Co-enzyme A ligase domain; C, condensation; A, adenylation; E, epimerization; TE, Termination; KR, Ketoreductase domain; ECH, Enoyl-CoA hydratase.

In the same study, for the first time, the presence of microcystin genes has been detected in the lake. Microcystins represent a family of genes coding one of the most common hepatotoxins that are produced by a wide range of cyanobacteria (Rantala et al., 2008; Ballot et al., 2010b). The genes encoding for these toxins belong to the secondary metabolism pathways, mostly the PKS, and NRPS families.

Screening the bins for secondary metabolite clusters, we see that the most abundant cluster belongs to the terpene pathway (125 clusters) (**Figure 1**). This biosynthesis pathway is well known to be present in many plant and fungi genomes, but recently it was proposed to be also widely distributed in bacterial genomes. One study revealed 262 distinct terpene synthases in the bacterial domain of life (Yamada et al., 2015). Consequently, it can represent a fertile source of new natural products, yet greatly underestimated. The second most abundant class of clusters belongs to the bacteriocin pathway (35 clusters) (**Figure 1**). bacteriocins belong to a group of ribosomal synthesized antimicrobial peptides which can kill or inhibit bacterial strains closely related or non-related to the bacteriocin producing bacteria (Yang et al., 2014). It has been suggested as a viable alternative to traditional antibiotics and can be used as narrow-spectrum antibiotics (Cotter et al., 2012). Only a few studies have been conducted to screen for bacteriocin genes by using a metatenomics approach, and solely for the host-associated microbiome (Zheng et al., 2015) or fermented food microbiome (Wieckowicz et al., 2011; Illeghems et al., 2015; Escobar-Zepeda et al., 2016). Neither of these studies was conducted for natural environments or used the metagenomics 2.0 approach.

In this study, we focused on 2 families of large modular secondary metabolite genes, type I PKS and NRPS. With our approach, it was possible to find a total of 18 NRPS, 6 type I PKS, and 3 hybrid PKS/NRPS clusters. For NRPS clusters, it was possible to recover 43 C domains, most of them (58%) from the LCL class. An LCL domain catalyzes a peptide bond between two L-amino acids (Rausch et al., 2007). A previous study also found that the LCL class was the most abundant in another aquatic environment, dominated by gram-negative bacteria (Cuadrat et al., 2015). Many studies have shown that the LCL class in aquatic environments is limited to gram-negative bacteria (Woodhouse et al., 2013). Our results further support this as we also found the LCL class only in bins of gram-negative bacteria (**Figure 3**) with the only exception of unclassified Archaea (bin 233), which should be further investigated in order to confirm the phylogenetic classification. It was also possible to recover 9 KS domains from type I PKS clusters, 56% from the modular class and 22% from the hybrid PKS/NRPS class. Those classes are larger (with many copies of each domain) than the iterative ones, increasing the chances to be recovered by metagenomic approaches. Accordingly, the NRPS and PKS clusters were more in depth analyzed, including syntheny, domain phylogeny, and partial metabolite protein structure predictions.

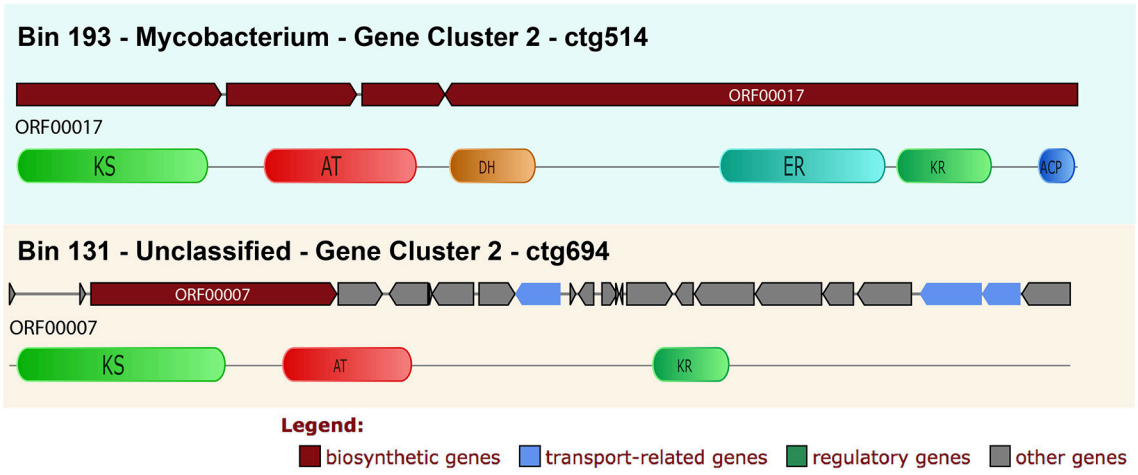


FIGURE 6 | Bin 193 and 131 type I PKS clusters detailed annotation and synteny. It is possible to see the synteny of the cluster with the functional classification for each ORF. In addition, for the PKS biosynthetic ORF the domain-specific annotations can be seen. KS, keto-synthase; AT, acyltransferase; KR, ketoreductase; E, epimerization; DH, dehydratase; ER, enoylreductase.

From the bins showing secondary metabolite genes, the most complete was from bin 34 (*Pseudomonas*). The 7 clusters on this genome vary from 8,675 base pairs (bp) to 52,516 bp in size, been only possible to be recovered due to the high completeness of the assembled genome (98.28%). The presence of a great diversity of clusters in *Pseudomonas* is expected, as many active secondary metabolites (encoded by NRPS) have been previously described in the Pseudomonas genus, ranging from antibiotics and antifungal to siderophores (Pan and Hu, 2015; Van Der Voort et al., 2015; Esmaeel et al., 2016).

From those 7 clusters in bin 34, the NRPS clusters 2 and 3 showed a high similarity with syringomycin (three domains) and microcystin (one domain) pathways. The first one is found, for example, in the *Pseudomonas syringae* (a plant pathogen) genome, as a virulence factor (syringomycin E) (Scholz-Schroeder et al., 2003), which also has antifungal activity against *Saccharomyces cerevisiae* (Stock et al., 2000). On the other hand, microcystin are a class of toxins produced by freshwater Cyanobacteria species (Dawson, 1998) and it can be produced in large quantities during massive bloom events (Bouhaddada et al., 2016). Due to the taxonomical classification of the bin and the higher number of domains similar to syringomycin, however, it is more likely that the product encoded by this cluster is functionally close to the latter pathway.

In bin 34-cluster 6, both C domains were classified as belonging to the pyochelin pathway. This peptide is a siderophore of *Pseudomonas aeruginosa* (Brandel et al., 2012). The presence of a TonB-dependent siderophore receptor in cluster 6 provides additional evidence about its functional classification. Additionally, two bacteriocin clusters and one Aryl polyene cluster were found in bin 34. Aryl polyenes are structurally similar to the well-known carotenoids with respect to their polyene systems and it was recently demonstrated that it can protect bacteria from reactive oxygen species, similarly to what is known for carotenoids (Schöner et al., 2016). These results suggest that a wide range of metabolites is encoded in this *Pseudomonas* genome, providing it an "arsenal" of secondary products, increasing the likelihood of the *Pseudomonas* species to succeed in aquatic systems.

Bin 131 (unclassified bacteria) shows a PKS cluster and 3 domains. It was classified as belonging to the enediynes pathway. These compounds are toxic to DNA and are under investigation as anti-tumor agents, with several compounds under clinical trials (Jones and Fouad, 2002). All are encoded by type I iterative PKS (Ahlert et al., 2002) and it was possible to recover the minimal core (KS, AT and KR) as well as transporter genes from the environmental genome. The most similar PKS I present in the public databases stems from *M. aeruginosa*, but only with an identity of 64%, suggesting that it is encoding for a new compound, which has not previously been described.

In bin 193 (*Mycobacterium*—sister linage *M. rhodesiae*), one of the 3 recovered clusters is a Type I PKS, similar to iterative PKS in the NAPDOS analysis (confirmed by the phylogenetic analysis). The most similar KS sequence belongs to *M. kansasii*, with 82% similarity. *M. rhodesiae* and *M. kansasii* are both non-tuberculous mycobacteria (NTM) that can be found in different environments, but both can also be opportunistic pathogens and cause a chronic pulmonary infection in immunosuppressed patients (Fedrizzi et al., 2017). The species *M. kansasii* comprises various subtypes and some are often recovered from tap water and occasionally from river or lake water (Bakuła et al., 2013; van der Wielen et al., 2013). There is still controversy about how the transmission from environment to human host occurs and also about the implications on public health. The presence of PKS in Mycobacterium genus was discovered more than a decade ago and most of the polyketides encoded by different species of this genus play a role in virulence and/or components of the extraordinarily complex mycobacterial cell

envelope (Quadri, 2014). Further studies must be done in order to investigate the potential of this bin to cause infections on humans, i.e., by screening virulence factors on the full genome.

Five bins of the phyla Cyanobacteria contained PKS and NRPS clusters. In bins 1 and 2 (Anabaena, now called "Dolichospermum"), it was possible to recover 2 NRPS and 1 hybrid NRPS-PKS, respectively. The genus *Anabaena* is known to encode several toxins, including the dangerous anatoxin-a, and to produce toxic blooms in lakes and reservoirs (Carmichael et al., 1975; Calteau et al., 2014; Brown et al., 2016; Li et al., 2016). However, the anatoxin-a is encoded by a type I PKS cluster (Méjean et al., 2014), unlike the NRPS and Hybrid clusters found in the *Anabaena* bins from this study.

On the other hand, the hepatotoxic heptapeptide of the class mycrocystin is present in many genera of Cyanobacteria, including *Anabaena*, and they are encoded by NRPS and also Hybrid NRPS-PKS clusters (Tillett et al., 2000; Rouhiainen et al., 2004; Viaggiu et al., 2004; Rastogi et al., 2015). The results of NAPDOS reveal one C domain from bin 1 classified in the pathway of Mycrocystin with e-value 6e-83.

In bins 6, 7 and 8 (*Planktothrix*), it was possible to find several type I PKS, NRPS and hybrid clusters. In bin 6 there are 3 NRPS, one type 1 PKS, and one hybrid cluster. The bin 7 shows 2 NRPS and bin 8 shows one type I PKS and 2 NRPS clusters. The genus *Planktothrix* can also be producer of anatoxin-a (Viaggiu et al., 2004) and the presence of type I PKS cluster on these bins can be alarming. However, the 3 KS domains from type I PKS from bin 6 reveal great similarity with the epothilone pathway. The NapDOS analysis and the KS domain from bin 8 suggest a high similarity to the neurotoxin jamaicamides pathway. A previous study showed in 2010 (Ballot et al., 2010a) the presence of an anatoxin-a-producing cyanobacterium in northeastern Germany Lake Stolpsee, rising concerns about the presence of these toxins in the waters of the northeastern German lakes.

The absence of anatoxin-a genes in the studied lake is in agreement with previous screening for a toxin screening in the lake (Dadheech et al., 2014).

In the bin 73 (*Acidobacteriales*) 1 PKS sequence was found. By the phylogenetic classification, its KS domain is clustered with *trans*-AT KS domains. The AT domains of *trans*-AT PKSs are not integrated into the assembly lines but expressed as free-standing polypeptides, unlike the more familiar *cis*-AT PKSs (Weissman, 2015). However, the NAPDOS result shows the AT domain of this bin in the same ORF with KS and KR domains, showing a syntheny that suggests a *cis*-AT PKS. In addition, the classification by similarity from NAPDOS suggests a polyunsaturated fatty acid (PUFA) but only with 31% of identity.

To assess the life style of the bins (free-living or particle-associated), we calculated the relative abundance of the bins in every sample. A total of 158 bins with significant difference between the 3 groups were found (**Table S7**), however from the 15 bins on which this study focused (showing NRPS and/or type I PKS clusters), only 4 bins (26.6%) were significantly differently present in the life-styles. Bins 1 and 2 (*Anabaena* genus) are more abundant in the PA group, especially on samples B7 and B9 (and also the replicates Old_b7 and Old_b9), accounting for 20–25% (bin 1) and 10–15% (bin 2) on these samples. The very high abundance of these bins on the samples can be explained by that, based on long term monitoring of the Lake Stechlin, we know that these samples were collected during the occurrence of a cyanobacterial bloom.

From the other bins containing PKS/NRPS clusters, we can see that Bins 6, 7, and 8 (*Planktothrix*), beside the lack of significant difference between FL and PA groups (p-value $> p$ 0.05), they are clearly more abundant in NSF. The possible explanation for this notion is that the NSF samples were collected during a mesocosm experiment, whereas the other samples were derived from the natural environment.

CONCLUSIONS

Using the Metagenomics 2.0 approach, we were able to recover full megasynthases sequences and their genomic context from environmental draft genomes. However, there are limitations, e.g., the genomic coverage of less abundant organisms and the possibility of chimeras. Recently, it has been demonstrated that with an increasing number of samples, it will be possible to recover individual species genomes with a high confidence (Kang et al., 2015). In the near future, with the advent of the 3rd generation sequencing, with longer reads, up to 100 kilobases, it will be possible to further improve the quality of the assemblies (Frank et al., 2016). These new approaches unlock the possibility of studying these newly recovered environmental pathways and their evolution in detail. Thus, allowing cloning and expressing these clusters will provide new natural products of great interest for the biotechnological and pharmaceutical industry. Moreover, studies have demonstrated the possibility to synthesize large functional DNA (Hutchison et al., 2016), and together with additional screening techniques, it will be possible to obtain such sequences and to synthesize the full cluster for heterologous expression, skipping the cloning and functional screening process, saving considerable time and money. In addition, the current work highlights the great potential for the discovery of new metabolically active compounds in freshwaters such as oligo-mesotrophic Lake Stechlin. Further, the study of complete or near complete genomes from uncultivated bacteria in the natural environment will enable us to better understand the multiple forms of interactions between species and how they compete for the limiting natural resources.

AUTHOR CONTRIBUTIONS

RC, DI, AD, and H-PG: Conceived and designed the experiments; RC, DI: Performed the experiments; RC, DI, and H-PG: Analyzed the data; DI and H-PG: Contributed reagents, materials, analysis tools; All authors wrote the manuscript and revised it for significant intellectual content.

FUNDING

This study was supported by the Science without Borders Program (Ciência Sem Fronteiras), CNPq. DI and H-PG were funded by German science foundation (DFG) projects Aquameth (GR1540/21-1) and Aggregates (GR1540/28-1).

ACKNOWLEDGMENTS

We thank Dr. Camila Mazzoni and all the team of Berlin Center for Genomics in Biodiversity Research (BeGenDiv) for allowing us to use the facilities and computational resources for the bioinformatics analyses. Elke Mach and the MIBI group are thanked for their technical support and fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00251/full#supplementary-material

Table S1 | Samples used in the study.

Table S2 | Taxonomical classification of bins, GC content and genome size.

Table S3 | Secondary metabolite clusters found in each bin.

Table S4 | KS domains blastP results against Refseq. Only top 3 hits are shown.

Table S5 | C domains blastP results against Refseq. Only top 3 hits are shown.

Table S6 | Relative abundance of each bin over the samples.

Table S7 | Stamp results. Only the bins where relative abundance was statistically significant different ($p < 0.05$) between the 3 groups (FL, PA, and NSF) are shown.

REFERENCES

Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., et al. (2002). The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* 297, 1173–1176. doi: 10.1126/science.1072105

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579

Amos, G. C., Borsetto, C., Laskaris, P., Krsek, M., Berry, A. E., Newsham, K. K., et al. (2015). Designing and implementing an assay for the detection of rare and divergent nrps and pks clones in European, Antarctic and Cuban soils. *PLoS ONE* 10:e0138327. doi: 10.1371/journal.pone.0138327

Austin, M. B., and Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* 20, 79–110. doi: 10.1039/b100917f

Bakuła, Z., Safianowska, A., Nowacka-Mazurek, M., Bielecki, J., and Jagielski, T. (2013). Short communication: subtyping of *Mycobacterium kansasii* by PCR-restriction enzyme analysis of the hsp65Gene. *Biomed Res. Int.* 2013:178725. doi: 10.1155/2013/178725

Ballot, A., Fastner, J., Lentz, M., and Wiedner, C. (2010a). First report of anatoxin-a-producing cyanobacterium Aphanizomenon issatschenkoi in northeastern Germany. *Toxicon* 56, 964–971. doi: 10.1016/j.toxicon.2010.06.021

Ballot, A., Fastner, J., and Wiedner, C. (2010b). Paralytic shellfish poisoning toxin-producing cyanobacterium Aphanizomenon gracile in northeast Germany. *Appl. Environ. Microbiol.* 76, 1173–1180. doi: 10.1128/AEM.02285-09

Bláhová, L., Babica, P., Adamovský, O., Kohoutek, J., Maršálek, B., and Bláha, L. (2008). Analyses of cyanobacterial toxins (microcystins, cylindrospermopsin) in the reservoirs of the Czech Republic and evaluation of health risks. *Environ. Chem. Lett.* 6, 223–227. doi: 10.1007/s10311-007-0126-x

Bláhová, L., Babica, P., Maršálková, E., Smutná, M., Maršálek, B., and Bláha, L. (2007). Concentrations and seasonal trends of extracellular microcystins in freshwaters of the Czech Republic – results of the national monitoring program. *Clean Soil Air Water* 35, 348–354. doi: 10.1002/clen.200700010

Bouhaddada, R., Nélieu, S., Nasri, H., Delarue, G., and Bouaïcha, N. (2016). High diversity of microcystins in a microcystis bloom from an Algerian lake. *Environ. Pollut.* 216, 836–844. doi: 10.1016/j.envpol.2016.06.055

Brandel, J., Humbert, N., Elhabiri, M., Schalk, I. J., Mislin, G. L. A., and Albrecht-Gary, A.-M. (2012). Pyochelin, a siderophore of *Pseudomonas aeruginosa*: physicochemical characterization of the iron(iii), copper(ii) and zinc(ii) complexes. *Dalton Trans.* 41, 2820–2834. doi: 10.1039/c1dt11804h

Brown, N. M., Mueller, R. S., Shepardson, J. W., Landry, Z. C., Morré, J. T., Maier, C. S., et al. (2016). Structural and functional analysis of the finished genome of the recently isolated toxic *Anabaena* sp. WA102. *BMC Genomics* 17:457. doi: 10.1186/s12864-016-2738-7

Calteau, A., Fewer, D. P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., et al. (2014). Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics* 15:977. doi: 10.1186/1471-2164-15-977

Cane, D. E., Walsh, C. T., and Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* 282, 63–68. doi: 10.1126/science.282.5386.63

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Carmichael, W. W., Biggs, D. F., and Gorham, P. R. (1975). Toxicology and pharmacological action of *Anabaena flos-aquae* toxin. *Science* 187, 542–544.

Cotter, P. D., Ross, R. P., and Hill, C. (2012). Bacteriocins — a viable alternative to antibiotics? *Nat. Rev. Microbiol.* 11, 95–105. doi: 10.1038/nrmicro2937

Cuadrat, R. R., Cury, J. C., and Dávila, A. M. (2015). Metagenomic analysis of upwelling-affected Brazilian coastal seawater reveals sequence domains of type I PKS and modular NRPS. *Int. J. Mol. Sci.* 16, 28285–28295. doi: 10.3390/ijms161226101

Dadheech, P. K., Selmeczy, G. B., Vasas, G., Padisák, J., Arp, W., Tapolczai, K., et al. (2014). Presence of potential toxin-producing cyanobacteria in an Oligo-Mesotrophic Lake in Baltic Lake district, Germany: an ecological, genetic and toxicological survey. *Toxins* 6, 2912–2931. doi: 10.3390/toxins 6102912

Dawson, R. M. (1998). The toxicology of microcystins. *Toxicon* 36, 953–962. doi: 10.1016/S0041-0101(97)00102-5

Di Lorenzo, M., Stork, M., Naka, H., Tolmasky, M. E., and Crosa, J. H. (2008). Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in *Vibrio anguillarum*. *Biometals* 21, 635–648. doi: 10.1007/s10534-008-9149-4

Escobar-Zepeda, A., Sanchez-Flores, A., and Quirasco Baruch, M. (2016). Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota. *Food Microbiol.* 57, 116–127. doi: 10.1016/j.fm.2016.02.004

Esmaeel, Q., Pupin, M., Kieu, N. P., Chataign,é, G., Béchet, M., Deravel, J., et al. (2016). *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. *Microbiologyopen* 5, 512–526. doi: 10.1002/mbo3.347

Fedrizzi, T., Meehan, C. J., Grottola, A., Giacobazzi, E., Serpini, G. F., Tagliazucchi, S., et al. (2017). Genomic characterization of nontuberculous mycobacteria. *Nature* 7:45258. doi: 10.1038/srep45258

Fisch, K. M. (2013). Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Adv.* 3, 18228–18247. doi: 10.1039/c3ra42661k

Foerstner, K. U., Doerks, T., Creevey, C. J., Doerks, A., and Bork, P. (2008). A computational screen for type i polyketide synthases in metagenomics shotgun data. *PLoS ONE* 3:e3515. doi: 10.1371/journal.pone.0003515

Frank, J. A., Pan, Y., Tooming-Klunderud, A., Eijsink, V. G. H., McHardy, A. C., Nederbragt, A. J., et al. (2016). Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Sci. Rep.* 6:25373. doi: 10.1038/srep25373

Gilbert, J. A., Meyer, F., and Bailey, M. J. (2011). The future of microbial metagenomics (or is ignorance bliss?). *ISME J. Int. Soc. Microb. Ecol.* 5, 777–779. doi: 10.1038/ismej.2010.178

Gokhale, R. S., Sankaranarayanan, R., and Mohanty, D. (2007). Versatility of polyketide synthases in generating metabolic diversity. *Curr. Opin. Struct. Biol.* 17, 736–743. doi: 10.1016/j.sbi.2007.08.021

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Hutchison, C. A. III., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. doi: 10.1126/science.aad6253

Illeghems, K., Weckx, S., and De Vuyst, L. (2015). Applying meta-pathway analyses through metagenomics to identify the functional properties of the major bacterial communities of a single spontaneous cocoa bean fermentation process sample. *Food Microbiol.* 50, 54–63. doi: 10.1016/j.fm.2015.03.005

Ionescu, D., Siebert, C., Polerecky, L., Munwes, Y. Y., Lott, C., Häusler, S., et al. (2012). Microbial and chemical characterization of underwater fresh water springs in the Dead Sea. *PLoS ONE* 7:e38319. doi: 10.1371/journal.pone.0038319

Jones, G. B., and Fouad, F. S. (2002). Designed enediyne antitumor agents. *Curr. Pharm. Des.* 8, 2415–2440. doi: 10.2174/1381612023392810

Jost Casper, S. (ed.). (1985). *Lake Stechlin: A Temperate Oligotrophic Lake (Monographiae Biologicae), 1st Edn*.

Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. doi: 10.7717/peerj.1165

Kirillin, G., Shatwell, T., and Kasprzak, P. (2013). Consequences of thermal pollution from a nuclear plant on lake temperature and mixing regime. *J. Hydrol.* 496, 47–56. doi: 10.1016/j.jhydrol.2013.05.023

Koglin, A., and Walsh, C. T. (2009). Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep.* 26, 987–1000. doi: 10.1039/b904543k

Komaki, H., Ichikawa, N., Hosoyama, A., Takahashi-Nakaguchi, A., Matsuzawa, T., Suzuki, K., et al. (2014). Genome based analysis of type-I polyketide synthase and nonribosomal peptide synthetase gene clusters in seven strains of five representative Nocardia species. *BMC Genomics* 15:323. doi: 10.1186/1471-2164-15-323

Lal, R., Kumari, R., Kaur, H., Khanna, R., Dhingra, N., and Tuteja, D. (2000). Regulation and manipulation of the gene clusters encoding type-I PKSs. *Trends Biotechnol.* 18, 264–274. doi: 10.1016/S0167-7799(00)01443-8

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, X., Dreher, T. W., and Li, R. (2016). An overview of diversity, occurrence, genetics and toxin production of bloom-forming Dolichospermum (Anabaena) species. *Harmful Algae* 54, 54–68. doi: 10.1016/j.hal.2015.10.015

Masschelein, J., Mattheus, W., Gao, L.-J., Moons, P., Van Houdt, R., Uytterhoeven, B., et al. (2013). A PKS/NRPS/FAS hybrid gene cluster from Serratia plymuthica RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics. *PLoS ONE* 8:e54143. doi: 10.1371/journal.pone.0054143

McMahon, K. (2015). 'Metagenomics 2.0.' *Environ. Microbiol. Rep.* 7, 38–39. doi: 10.1111/1758-2229

Méjean, A., Paci, G., Gautier, V., and Ploux, O. (2014). Biosynthesis of anatoxin-a and analogues (anatoxins) in cyanobacteria. *Toxicon* 91, 15–22. doi: 10.1016/j.toxicon.2014.07.016

Meriluoto, J., Spoof, L., and Codd, G. A. (eds.). (2017). *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis*. Chichester: Wiley.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Micallef, M. L., D'Agostino, P. M., Sharma, D., Viswanathan, R., and Moffitt, M. C. (2015). Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria. *BMC Genomics* 16:669. doi: 10.1186/s12864-015-1855-z

Minowa, Y., Araki, M., and Kanehisa, M. (2007). Comprehensive analysis of distinctive polyke tide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* 368, 1500–1517. doi: 10.1016/j.jmb.2007.02.099

Mizuno, C. M., Kimes, N. E., López-Pérez, M., Ausó, E., Rodriguez-Valera, F., and Ghai, R. (2013). A hybrid NRPS-PKS gene cluster related to the bleomycin family of antitumor antibiotics in *Alteromonas macleodii* strains. *PLoS ONE* 8:e76021. doi: 10.1371/journal.pone.0076021

National Research Council US. (2007). *Committee on Metagenomics: Challenges and Functional Applications. The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. Washington, DC: National Academies Press (US); Data Management and Bioinformatics Challenges of Metagenomics.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Paerl, H. W., and Huisman, J. (2008). Blooms like it hot. *Science* 320, 57–58. doi: 10.1126/science.1155398

Pan, H.-Q., and Hu, J.-C. (2015). Draft genome sequence of the novel strain Pseudomonas sp. 10B238 with potential ability to produce antibiotics from deep-sea sediment. *Mar. Genomics* 23, 55–57. doi: 10.1016/j.margen.2015.05.003

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494

Quadri, L. E. (2014). Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. *Crit. Rev. Biochem. Mol. Biol.* 49, 179–211. doi: 10.3109/10409238.2014.896859

Rantala, A., Rizzi, E., Castiglioni, B., de Bellis, G., and Sivonen, K (2008). Identification of hepatotoxin producing cyanobacteria by DNA-chip. *Environ. Microbiol.* 10, 653–664. doi: 10.1111/j.1462-2920.2007.01488.x

Rastogi, R. P., Madamwar, D., and Incharoensakdi, A. (2015). Bloom dynamics of cyanobacteria and their toxins: environmental health impacts and mitigation strategies. *Front. Microbiol.* 6:1254. doi: 10.3389/fmicb.2015.01254

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* 7:78. doi: 10.1186/1471-2148-7-78

Rodríguez-Valera, F. (2004). Environmental genomics, the big picture? *FEMS Microbiol. Lett.* 231, 153–158. doi: 10.1016/S0378-1097(04)00006-0

Rouhiainen, L., Vakkilainen, T., Siemer, B. L., Buikema, W., Haselkorn, R., and Sivonen, K. (2004). Genes coding for hepatotoxic heptapeptides (Microcystins) in the cyanobacterium anabaena strain 90. *Appl. Environ. Microbiol.* 70, 686–692. doi: 10.1128/AEM.70.2.686-692.2004

Salmaso, N., and Padisák, J. (2007). Morpho-functional groups and phytoplankton development in two deep lakes (Lake Garda, Italy and Lake Stechlin, Germany). *Hydrobiologia* 578, 97–112. doi: 10.1007/s10750-006-0437-0

Sangwan, N., Xia, F., and Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome* 4:8. doi: 10.1186/s40168-016-0154-5

Scholz-Schroeder, B. K., Soule, J. D., and Gross, D. C. (2003). The sypA, sypB, and sypC synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. syringae B301D. *Mol. Plant Microbe Inter.* 16, 271–280. doi: 10.1094/MPMI.2003.16.4.271

Schöner, T. A., Gassel, S., Osawa, A., Tobias, N. J., Okuno, Y., Sakakibara, Y., et al. (2016). Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *Chembiochem* 17, 247–253. doi: 10.1002/cbic.201500474

Selvin, J., Sathiyanarayanan, G., Lipton, A. N., Al-Dhabi, N. A., Valan Arasu, M., and Kiran, G. S. (2016). Ketide Synthase (KS) domain prediction and analysis of iterative type II PKS gene in marine sponge-associated actinobacteria producing biosurfactants and antimicrobial agents. *Front. Microbiol.* 7:63. doi: 10.3389/fmicb.2016.00063

Silva-Stenico, M. E., Silva, C. S. P., Lorenzi, A. S., Shishido, T. K., Etchegaray, A., Lira, S. P., et al. (2011). Non-ribosomal peptides produced by Brazilian cyanobacterial isolates with antimicrobial activity. *Microbiol. Res.* 166, 161–175. doi: 10.1016/j.micres.2010.04.002

Stock, S. D., Hama, H., Radding, J. A., Young, D. A., and Takemoto, J. Y. (2000). Syringomycin E inhibition of *Saccharomyces cerevisiae*: requirement for biosynthesis of sphingolipids with very-long-chain fatty acids and

mannose-and phosphoinositol-containing head groups. *Antimicrob. Agents Chemother.* 44, 1174–1180. doi: 10.1128/AAC.44.5.1174-1180.2000

Strous, M., Kraft, B., Bisdorf, R., and Tegetmeyer, H. E. (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* 3:410. doi: 10.3389/fmicb.2012.00410

Sun, W., Peng, C., Zhao, Y., and Li, Z. (2012). Functional gene-guided discovery of type II polyketides from culturable actinomycetes associated with soft coral *Scleronephthya* sp. *PLoS ONE* 7:e42847. doi: 10.1371/journal.pone.0042847

Teeling, H., and Glöckner, F. O. (2012). Current opportunities and challenges in microbial metagenome analysis–a bioinformatic perspective. *Brief. Bioinformatics* 13, 728–742. doi: 10.1093/bib/bbs039

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., and Neilan, B. A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem. Biol.* 7, 753–764. doi: 10.1016/S1074-5521(00)00021-1

Van Der Voort, M., Meijer, H. J. G., Schmidt, Y., Watrous, J., Dekkers, E., Mendes, R., et al. (2015). Genome mining and metabolic profiling of the rhizosphere bacterium *Pseudomonas* sp. SH-C52 for antimicrobial compounds. *Front. Microbiol.* 6:693. doi: 10.3389/fmicb.2015.00693

van der Wielen, P. W., Heijnen, L., and van der Kooij, D. (2013). Pyrosequence analysis of the hsp65 genes of nontuberculous mycobacterium communities in unchlorinated drinking water in the Netherlands. *Appl. Environ. Microbiol.* 79, 6160–6166. doi: 10.1128/AEM.01591-13

Vanwonterghem, I., Jensen, P. D., Rabaey, K., and Tyson, G. W. (2016). Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion: genome-centric resolution through deep metagenomics. *Environ. Microbiol.* 18, 3144–3158. doi: 10.1111/1462-2920.13382

Viaggiu, E., Melchiorre, S., Volpi, F., Di Corcia, A., Mancini, R., Garibaldi, L., et al. (2004). Anatoxin-a toxin in the cyanobacterium *Planktothrix rubescens* from a fishing pond in northern Italy. *Environ. Toxicol.* 19, 191–197. doi: 10.1002/tox.20011

Vincent, A. T., Derome, N., Boyle, B., Culley, A. I., and Charette, S. J. (2016). Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *J. Microbiol. Methods*. 138, 60–71. doi: 10.1016/j.mimet.2016.02.016

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, W237–W243. doi: 10.1093/nar/gkv437

Weissman, K. J. (2015). The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* 11, 660–670. doi: 10.1038/nchembio.1883

Wieckowicz, M., Schmidt, M., Sip, A., and Grajek, W. (2011). Development of a PCR-based assay for rapid detection of class IIa bacteriocin genes: detection of class IIa bacteriocins. *Lett. Appl. Microbiol.* 52, 281–289. doi: 10.1111/j.1472-765X.2010.02999.x

Woodhouse, J. N., Fan, L., Brown, M. V., Thomas, T., and Neilan, B. A. (2013). Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. *ISME J.* 7, 1842–1851. doi: 10.1038/ismej.2013.65

Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A., and Singer, S. W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2:26. doi: 10.1186/2049-2618-2-26

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K, Omura, S., Cane, D. E., et al. (2015). Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 112, 857–862. doi: 10.1073/pnas.1422108112

Yang, S.-C., Lin, C.-H., Sung, C. T., and Fang, J.-Y. (2014). Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Front. Microbiol.* 5:241. doi: 10.3389/fmicb.2014.00241

Zheng, J., Gänzle, M. G., Lin, X. B., Ruan, L., and Sun, M. (2015). Diversity and dynamics of bacteriocins from human microbiome: bacteriocins of human microbiome. *Environ. Microbiol.* 17, 2133–2143. doi: 10.1111/1462-2920.12662

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* 7:e34064. doi: 10.1371/journal.pone.0034064

Zothanpuia, Passari, A. K., Gupta, V. K., and Singh, B. P. (2016). Detection of antibiotic-resistant bacteria endowed with antimicrobial activity from a freshwater lake and their phylogenetic affiliation. *PeerJ* 4:e2103. doi: 10.7717/peerj.2103

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cuadrat, Ionescu, Dávila and Grossart. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 February 2017
doi: 10.3389/fmicb.2017.00280

Functional Diversity of Transcriptional Regulators in the Cyanobacterium *Synechocystis* sp. PCC 6803

Mengliang Shi[1,2,3], Xiaoqing Zhang[1,2,3], Guangsheng Pei[1,2,3], Lei Chen[1,2,3] and Weiwen Zhang[1,2,3,4]*

[1] Laboratory of Synthetic Microbiology, School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, [2] Key Laboratory of Systems Bioengineering – Ministry of Education, Tianjin University, Tianjin, China, [3] SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering, Tianjin, China, [4] Center for Biosafety Research and Strategy, Tianjin University, Tianjin, China

Functions of transcriptional regulators (TRs) are still poorly understood in the model cyanobacterium *Synechocystis* sp. PCC 6803. To address the issue, we constructed knockout mutants for 32 putative TR-encoding genes of *Synechocystis*, and comparatively analyzed their phenotypes under autotrophic growth condition and metabolic profiles using liquid chromatography-mass spectrometry-based metabolomics. The results showed that only four mutants of TR genes, *sll1872* (*lytR*), *slr0741* (*phoU*), *slr0395* (*ntcB*), and *slr1871* (*pirR*), showed differential growth patterns in BG11 medium when compared with the wild type; however, in spite of no growth difference observed for the remaining TR mutants, metabolomic profiling showed that they were different at the metabolite level, suggesting significant functional diversity of TRs in *Synechocystis*. In addition, an integrative metabolomic and gene families' analysis of all TR mutants led to the identification of five pairs of TR genes that each shared close relationship in both gene families and metabolomic clustering trees, suggesting possible conserved functions of these TRs during evolution. Moreover, more than a dozen pairs of TR genes with different origin and evolution were found with similar metabolomic profiles, suggesting a possible functional convergence of the TRs during genome evolution. Finally, a protein–protein network analysis was performed to predict regulatory targets of TRs, allowing inference of possible regulatory gene targets for 4 out of five pairs of TRs. This study provided new insights into the regulatory functions and evolution of TR genes in *Synechocystis*.

Keywords: function, metabolomics, *Synechocystis*, transcriptional regulators, LC-MS

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Cientificas y Tecnicas (CONICET), Argentina

Reviewed by:
Takashi Osanai, Meiji University, Japan
Paul Hudson, Royal Institute of Technology, Sweden
Johannes Asplund-Samuelsson, Royal Institute of Technology, Sweden

****Correspondence:***
Lei Chen
lchen@tju.edu.cn

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *14 December 2016*
Accepted: *09 February 2017*
Published: *21 February 2017*

Citation:
Shi M, Zhang X, Pei G, Chen L and Zhang W (2017) Functional Diversity of Transcriptional Regulators in the Cyanobacterium Synechocystis sp. PCC 6803. Front. Microbiol. 8:280. doi: 10.3389/fmicb.2017.00280

INTRODUCTION

Cyanobacteria contribute significantly to global photosynthetic productivity. It is estimated that more than half of the total primary production essential for life on earth is produced by cyanobacteria. In addition, early studies have found that cyanobacteria were able to establish competitive growth in almost any environment, at least temporarily, liquid water and sunlight, due to their strong abilities of withstanding challenges of environmental perturbations

(Badger et al., 2006). Moreover, cyanobacteria have recently attracted significant interest because of their ability to function as a "*chassis*" to produce renewable carbon neutral biofuels or bioproducts (Atsumi et al., 2009). In spite of their important ecological, environmental, and biotechnological applications, many aspects of cyanobacterial physiology remain poorly understood.

To survive the diversities of environments, abundant and dedicated regulatory systems have evolved in cyanobacteria to achieve precise controls of functional gene expression. In the model cyanobacterium *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*), a significant number of regulatory genes of various types have been identified, among which at least 40 genes were annotated as putative transcriptional regulators (TRs) (Kaneko et al., 2003). So far only a dozen of TRs have ever been functionally characterized in *Synechocystis*, and the results showed that they were involved in the regulation of a wide range of physiological functions, such as nitrite tolerance (Aichi et al., 2001), iron limitation (Michel and Pistorius, 2004), acid tolerance (Ohta et al., 2005), cadmium tolerance (Houot et al., 2007). However, a majority of TRs in the *Synechocystis* genome are still functionally unknown, which presents significant challenges not only to the basic sciences of *Synechocystis* but also to the biotechnological application of *Synechocystis* as a chassis in producing biofuels and chemicals (Atsumi et al., 2009).

To decipher regulatory function of TRs, various approaches have been previously applied. For example, sequence analysis-based identification and evolutionary analysis of DNA-binding proteins, construction of transcriptional networks including TRs and their target genes and analysis of structure and evolution of these networks (Babu et al., 2004), can be used for functional inference of TRs. Metabolomics is a method to define the diversity of low weight molecules in the cell and to display differences in small molecule abundance. When applied for analysis of cellular responses to genetic or physiological changes, it shows many advantages because metabolites are the functional entities within the cells and their concentration levels vary as a consequence of environmental changes (Zhang et al., 2010). In our previous studies, metabolomic analysis has been applied to the functional characterization of response regulators involved in acid and butanol tolerance (Ren et al., 2014; Niu et al., 2015), and TRs involved in ethanol tolerance in *Synechocystis* (Zhu et al., 2015), and the results demonstrated that it could be a powerful tool in revealing functional clues for functionally unknown regulatory genes. Towards an ultimate goal of deciphering regulatory functions of TRs in *Synechocystis*, in this study, we applied a liquid chromatography-mass spectrometry (LC-MS) based metabolomics to a comparative analysis of knockout mutants for 32 putative *Synechocystis* TR-encoding genes (Zhu et al., 2015). The results showed significant functional diversity of TRs at the metabolic level in *Synechocystis*, as well as functional diversity of TRs based on their differential clustering patterns in relationship trees resulted from TR families and metabolomic clustering analysis. This study provided interesting information on the regulatory functions and evolution of TR genes in *Synechocystis*.

MATERIALS AND METHODS

Bacterial Growth Conditions

Synechocystis sp. PCC 6803 was obtained from American Type Culture Collection (ATCC), and used as a wild type to construct single-gene knockout mutants of TR genes. A total of 32 knockout mutants of putative TR-coding genes were constructed, confirmed and described previously (Zhu et al., 2015). Briefly, for the gene target selected, three sets of primers were designed to amplify a linear DNA fragment containing the chloramphenicol resistance cassette (amplified from a plasmid pACYC184) with two flanking arms of DNA upstream and downstream of the targeted gene. The linear fused PCR amplicon was used directly for transformation into *Synechocystis* by natural transformation. The chloramphenicol resistant transformants were obtained and passed several times on fresh BG11 plates supplemented with 10 μg mL^{-1} chloramphenicol to achieve full chromosome segregation (confirmed by PCR). The mutants and the wild type were grown in the BG11 medium (pH 7.5) in 100-mL flasks each with 25 mL medium, the light intensity was approximately 50 μmol photons m^{-2}s^{-1} and the illuminating incubator was 130 rpm, the temperature was controlled at 30°C (HNY-211B Illuminating Shaker, Honour, China). All mutants were first cultivated in BG11 culture with 10 μg/ml chloromycetin for 48 h and then inoculated into BG11 culture without chloromycetin. The growth was determined by cell density measured at OD_{630} on a UV-1750 spectrophotometer (Shimadzu, Japan) every 12 h. For each mutant, three biological replicates were established independently, and each sample was measured in triplicates (Zhu et al., 2015). To confirm the growth patterns, growth experiment of every knockout mutant was repeated at least three times independently, then the growth rates of all mutants were calculated (Supplementary Table S1). Only the growth rates between different RR mutants and wild type with p-value <0.005 by t-test were considered a significant growth difference.

LC-MS Based Metabolomics Analysis

Liquid chromatography-mass spectrometry based targeted metabolomics was performed according to the protocol described previously (Wang et al., 2014). All chemicals used for LC-MS metabolomics analysis were obtained from Sigma–Aldrich (Taufkirchen, Germany). For metabolomic analysis, the wild type and the mutant cells were collected at 48 and 72 h, respectively, and each sample was prepared with three biological replicates. Due to the large amount of cultivation needed to finish the comparative experiments of 32 TRs, the samples had to be cultivated, prepared, and analyzed in five batches. A separate cultivation and analysis of the wild type as control was conducted for every batch to minimize possible batch difference. Briefly, the cells were collected by centrifugation at 7500 × g for 8 min at 4°C (Eppendorf 5430R, Hamburg, Germany), quenched, and extracted rapidly with 900 μL of 80:20 MeOH/H_2O (−80°C) and then frozen in liquid nitrogen. The samples were then frozen-thawed three times to release metabolites from the cells. The supernatant was collected after centrifugation at 15,000 × g for 5 min at −4°C and then stored at −80°C. The

remaining cell pellets were re-suspended in 500 μL of 80:20 MeOH/H_2O (−80°C), and then the above extraction process was repeated. The supernatant from the second extraction was pooled with that from the first extraction and stored at −80°C until the LC-MS analysis was conducted. LC-MS analysis was conducted on an Agilent 1260 series binary HPLC system (Agilent Technologies, Waldbronn, Germany) using a Synergi Hydro-RP (C18) 150 mm × 2.0 mm ID, 4-μm 80-Å particle column (Phenomenex, Torrance, CA, USA), coupled to an Agilent 6410 triple quadrupole mass analyzer equipped with an electrospray ionization (ESI) source. Data was acquired using Agilent Mass Hunter workstation LC/QQQ acquisition software (version B.04.01), and chromatographic peaks were subsequently integrated via Agilent Qualitative Analysis software (version B.04.00). A total of 24 metabolites were selected for LC-MS-based targeted metabolite analysis in this study. All data of metabolomic profiling was first normalized by the internal control and the cell numbers of the samples.

The 24 targeted metabolites include acetyl coenzyme A (AcCOA), adenosine 5′-diphosphate (ADP), adenosine-5′-diphosphoglucose (ADP-GCS), α-ketoglutaric acid (AKG), adenosine 5′-monophosphate (AMP), adenosine 5′-triphosphate (ATP), coenzyme A hydrate (CoA), dihydroxyacetone phosphate (DHAP), D-fructose 1,6-bisphosphate (FBP), D-fructose 6-phosphate (F6P), sodium fumarate dibasic (FUM), DL-glyceraldehyde 3-phosphate (GAP), D-glucose 6-phosphate (G6P), L-glutamic acid (GLU), α-nicotinamide adenine dinucleotide (NAD), reduced α-nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADP), reduced nicotinamide adenine dinucleotide phosphate (NADPH), uridine 5′-diphosphoglucose (UDP-GCS), oxaloacetic acid (OXA), phosphor (enol)pyruvic acid (PEP), D-(-)-3-phosphoglyceric acid (3PG), D-ribose 5-phosphate (R5P), and D-ribulose1,5-bisphosphate (UDP-GCS), uridine 5′-diphosphoglucose (RiBP).

STATISTICAL ANALYSIS

The metabolomic profiles were further normalized by comparing relative values of the mutants to the wild type, and then log2 transformed. The data were subjected to Principal Component Analysis (PCA) using software SIMCA-P 11.5 (Laiakis et al., 2010). The PCA analysis is a statistical method to find outliers in the whole set of data. Samples with p-value < 0.05 by hotelling t2 statistic were considered significantly different. For Euclidean distance calculation, we used the *dist* function in *R* software after data normalization. Only the distances larger than the upper quartile could be considered as the most affected mutants. Hierarchical clustering analysis was conducted using a *R* software (Deu-Pons et al., 2014).

TR Family Analysis

Protein sequences of all the 32 TR genes were downloaded from NCBI[1]. To define potential TR families, we used BLAST software for homology identification, only those TR with 80% aligned coverage with *E*-value < 1e-20 were consider as same families.

Protein–Protein Interaction (PPI) Network Analysis

A protein–protein interaction (PPI) dataset of *Synechocystis* was downloaded from the STRING database (http://www.string-db.org/) (Jensen et al., 2009). STRING aggregates data and predictions stemming from a wide spectrum of cell types and environmental conditions, and aims to represent the union of all possible protein–protein links. In the STRING database, several types of evidence for the association, including genomic context, high-throughput experiments, conserved co-expression and previous biological knowledge were used to calculate a single combined score for each gene in the genome. In this study, only those experimentally validated were applied to construct the PPI network to cover potential protein-protein connections, and the notes of all the proteins in this study were renamed using gene IDs (Szklarczyk et al., 2011).

RESULTS AND DISCUSSION

Comparative Growth Analysis of TR Mutants

Although bioinformatics analysis of TRs in *Synechocystis* based on the sequence similarity has been conducted previously (Huffman and Brennan, 2002; Los et al., 2010), their functional classification using experimental approaches is still insufficient. A library of single deletion mutants for 32 TR-encoding genes in *Synechocystis* was constructed and confirmed previously in our laboratory (Zhu et al., 2015), and the majority of them have not been functionally characterized. To seek more functional information for these *Synechocystis* TRs, we first measured differential growth of all TR mutants in normal BG11 medium in flask cultivation, in parallel with the wild type *Synechocystis*. While most of the mutants grew equally well as the wild type (**Supplementary Figure S2**), the comparative analysis showed that four TR mutants, △*slr0395*, △*slr1871*, △*slr0741*, and △*sll1872*, grew poorly in the BG11 medium when compared with the wild type (**Figure 1** and Supplementary Table S1), suggesting that the function of these four TRs might be related to key metabolism necessary for normal growth in the BG11 autotrophic growth medium. Among them, deletion of *slr0395* (*ntcB*) caused the most significant growth arrest, with approximately 40% of the growth compared to the wild type after cultivation of 72 h, while deletion of *slr1871* (*pirR*) resulted in the least growth defect, with only 15% of the growth compared to the wild type after 72 h cultivation.

Slr0395 has previously been annotated as nitrite-responsive transcriptional enhancer NtcB in *Synechocystis,* on the basis of the inability of the △*slr0395* mutant to rapidly accumulate the transcripts of the nitrate assimilation genes upon induction and to respond to nitrite. In the *ntcB* mutant, activities of the nitrate assimilation enzymes were 40 to 50% of the wild type level, and

[1]http://www.ncbi.nlm.nih.gov

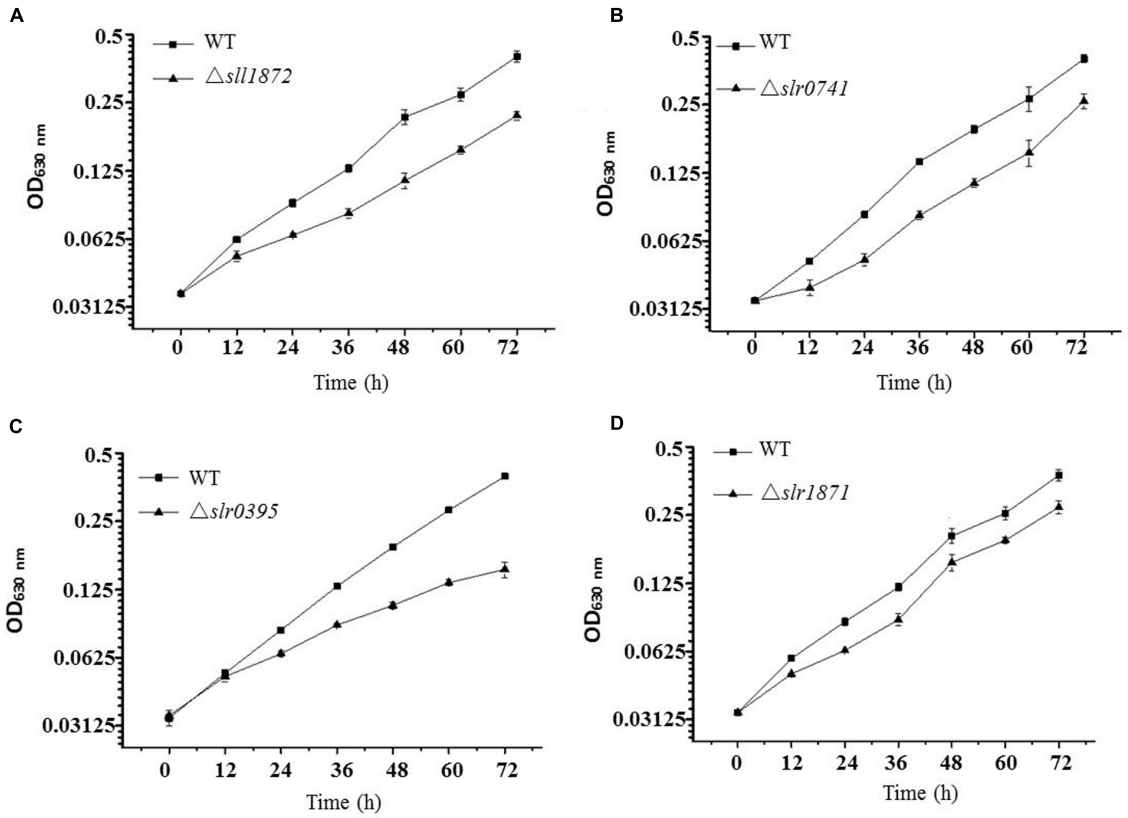


FIGURE 1 | Growth time courses of the wild type and the mutants in BG11 media. (A) The wild type and the △*sll1872* mutant. The block represents the wild type and the triangle represents the △*sll1872* mutant. **(B)** The wild type and the △*slr0741* mutant. The block represents the wild type and the triangle represents the △*slr0741* mutant. **(C)** The wild type and the △*slr0395* mutant. The block represents the wild type and the triangle represents the △*slr0395* mutant. **(D)** The wild type and the △*slr1871* mutant. The block represents the wild type and the triangle represents the △*slr1871* mutant.

the cells grew on nitrate at a rate approximately threefold lower than that of the wild type (Aichi et al., 2001).

Slr1871 was previously annotated as PirR of a LysR family, whose encoding gene is located immediately upstream of *pirAB* encoding an ortholog of pirin in the divergent direction, and DNA microarray analysis indicated that PirR repressed expression of closely located ORFs, *slr1870* and *mutS* (*sll1772*), in addition to *pirAB* and *pirR* itself (Hihara et al., 2004). Slr0741 was previously found to encode a negative regulator of the Pi regulon and its insertional inactivation in *Synechocystis* led to increase of the intracellular polyP level (Morohoshi et al., 2002); Slr0741 was also found involved in transduction of the phosphate-limitation signal in *Synechocystis* (Juntarajumnong et al., 2007), and it was also up-regulated upon ethanol stress as revealed by RNA-seq analysis (Wang et al., 2012). Currently no functional information is available for Sll1872.

Metabolomic Analysis Reveals Functional Diversity of TRs

Liquid chromatography-mass spectrometry-based metabolomics analysis has been recently used to investigate cyanobacterial metabolism due to its advantages toward chemically unstable metabolites, such as the hydrolytically unstable nucleotides (i.e., ATP, GTP, cAMP, and PEP) and the redox active nucleotides (i.e., NADPH, NADP) whose determination could be important in deciphering metabolic responses to genetic or physiological changes. Using a protocol optimized in our previous studies (Wang et al., 2014), a LC-MS-based comparative metabolomics analysis was conducted on all TR mutants and the wild type, with 24 key metabolites involved in central carbon metabolism, cellular energy charge and redox monitored in all samples at 48 and 72 h. The cell samples of 32 TR mutants and the wild type used for LC-MS-based metabolomic analysis were cultivated in BG11 media under autotrophic growth condition and collected at both 48 and 72 h, which were corresponding to earlier and latter exponential phases of cell growth. Each sample was prepared with three biological replicates. As for some, metabolite levels might change during the 8 min centrifugation procedure, a faster centrifugation method may be considered. To reduce the sampling time and maintain the metabolites as much as possible, a higher rotation rate and a lower temperature should be considered. However, with higher rotation rate, more severe damages may occur to cells. A protection agent may be considered in the future. When the rotation rate was higher than 7500 × g, the cells of *Synechocystis*

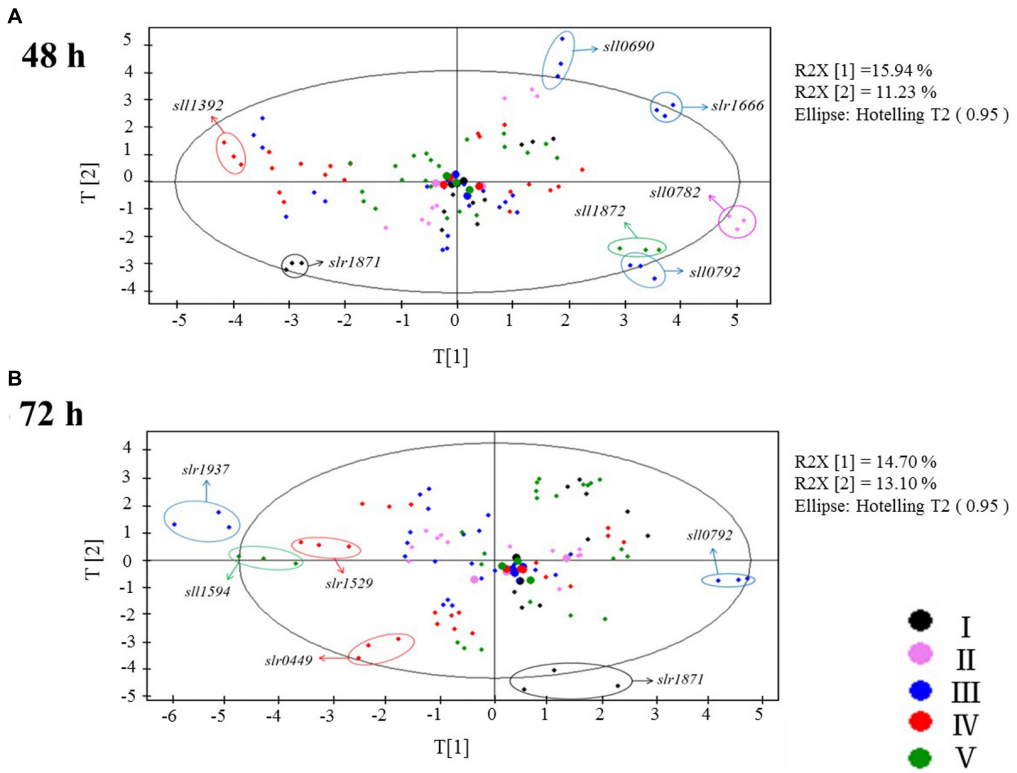


FIGURE 2 | Targeted LC-MS metabolomic analysis. (A) PCA plots of the LC-MS metabolomic profiles at 48 h. **(B)** PCA plots of the LC-MS metabolomic profiles at 72 h. I, II, III, IV, and V are experiments conducted in five separate times with different colors represent for each batch: black, purple, blue, red, green were for batch I, II, III, IV, V separately.

would be easily broken, leading to less metabolites being preserved.

Several patterns were observed in the PCA plots of the metabolomic data (**Figure 2**): (i) As a large number of cultivation was needed to finish the comparative experiments of 32 TRs, the samples had to be cultivated, prepared and analyzed in five batches. To minimize possible batch difference, a separate cultivation and analysis of the wild type as control was conducted for every batch. In the PCA plots, the big dots of five different colors representing the controls of five batches were found clustered together after data normalization, demonstrating the systematic errors resulting from the experimental design and different cultivation batches were not significant (**Figure 2**); (ii) metabolic profiles of TR mutants were in general well separated at both time points, demonstrating that the LC-MS-based methodology we utilized in this study is sensitive enough to investigate possible differences between controls and all the TR mutants; (iii) except for the four mutants that were grew poorer than the wild type, almost no growth difference was observed between the remaining 28 TR mutants and the wild type when cultivated in the BG11 medium; however, PCA analysis of metabolic profiles showed that these TR mutants were well separated from the wild type in the plots, suggesting that the deletion of these TR-encoding genes has caused significant changes to the cells at the metabolite level. At 48 h, seven TR mutants with the most significant metabolic changes from the wild type control were △*slr1871*, △*sll1392*, △*sll0690*, △*slr1666*, △*sll1872*, △*sll0782*, and △*sll0792* (**Figure 2A**, with two components proportions of 15.94 and 11.23%); while at 72 h, six TR mutants, △*slr1937*, △*sll1594*, △*slr1529*, △*sll0792*, △*slr1871*, and △*slr0449* (**Figure 2B**, with two components proportions of 14.70 and 13.10%), displayed the most significant metabolic changes, suggesting a significant functional diversity of TRs in *Synechocystis*, as revealed by the metabolic profiling of selected metabolites related to central carbohydrate metabolism; (iv) only two TR mutants, △*sll0792* and△*slr1871*, were found significantly regulated at both 48 and 72 h at the metabolite level, suggesting that time- or growth phase-dependent regulation may be involved for most of the responsive TRs; (v) a close examination of the mutant △*slr1871* showed that almost all the metabolites involved in central carbohydrate metabolism were down-regulated when compared with the wild type, consistent with the previous results that the *slr1871* (*pirR*) gene had a reduced transcript level during a light-limited linear growth when compared to the exponential growth (Foster et al., 2007). Although it has been reported that the *sll0792* gene encodes ZiaR, a Zn^{2+}-responsive repressor of *ziaA* encoding a polypeptide with sequence features of heavy metal transporting P-type ATPases in *Synechocystis* (Thelwell et al., 1998), its regulatory function on cellar metabolism has not yet been established.

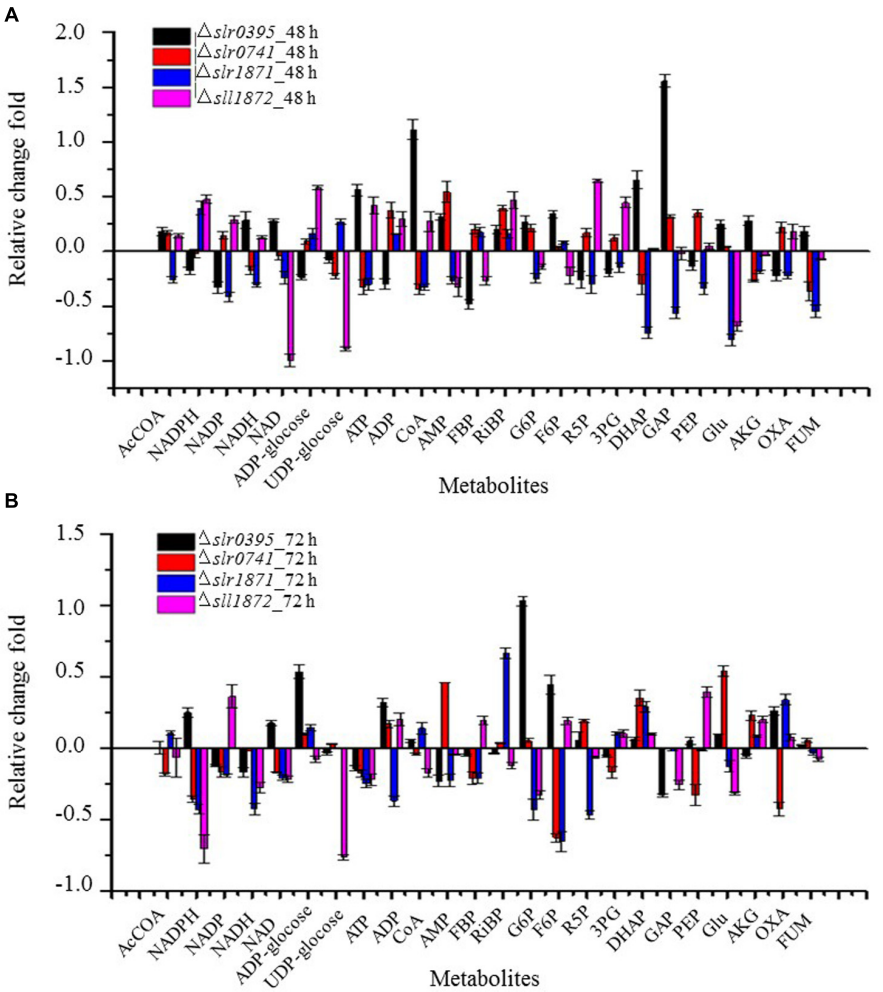


FIGURE 3 | A detailed analysis of the metabolite abundance in four mutants with differential growth. (A) calculated based on the metabolite profiles of 48 h. **(B)** calculated based on the metabolite profiles of 72 h.

To confirm the analysis with PCA, another approach, the *Euclidean* distances calculated based on the different metabolite profiles between each mutant and its control (the wild type), were also determined. In **Supplementary Figure S1**, the TR mutants were ordered based on the degree of metabolic changes when compared with the wild type in descending order. The analysis was conducted separately with metabolomic profiling data of two time points (i.e., 48 and 72 h), and showed that the top changed mutants were △*sll0690*, △*slr1871*, △*sll1670*, △*slr1666*, △*slr1245*, △*sll1957*, △*sll1594*, and △*sll1872* at 48 h; △*slr0449*, △*slr1871*, △*ssl0564*, △*sll1937*, △*slr1666*, △*slr0115*, △*slr0724*, and △*sll0690* at 72 h, respectively. The results were consistent with those of the PCA analysis.

Metabolomic Basis for the Differential Growth in Four TR Mutants

Comparative growth analysis showed that four TR mutants were grown poorly in BG11 medium when compared with the wild type. A detailed analysis of the metabolite abundance of the 24 metabolites in these mutants was then conducted (**Figure 3**). Slr0395 is involved in regulation of nitrate assimilation gene (*ntcB*) (Burnap et al., 2015), so it is expected that the deletion of the *slr0395* (*ntcB*) gene would decrease nitrogen metabolism. Accordingly, the metabolomics analysis showed Glu was up-regulated by 24.2, and 11.7% at 48 h and 72 h, respectively. In addition, abundance of CoA in the △*slr0395* was found increased.

Slr0395 and Slr1871 are LysR-type transcriptional regulator protein (LTTRs) that have been found important for regulation of the carbon concentration mechanism (CCM) in cyanobacteria (Daley et al., 2012). In Calvin–Benson cycle, F6P and G3P were converted to GAP, R5P, and then Ru5P and RiBP for CO_2 fixation (Wang et al., 2011). The metabolomic analysis showed that metabolite RiBP, F6P, GAP, and R5P associated with RiBP were down-regulated in the △*slr0395* and △*slr1871* mutants (**Figure 3**), which could be responsible for the slow growth in the mutants.

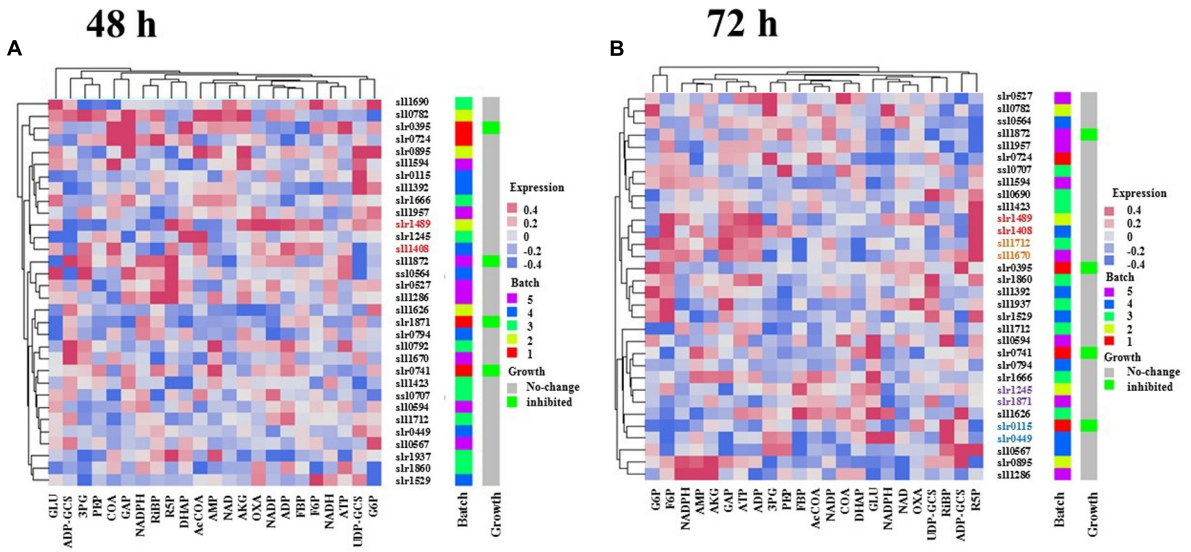


FIGURE 4 | Comparison of hierarchical clustering trees of metabolomic data. (A) Hierarchical clustering tree based on the metabolite profiles of 48 h. **(B)** Hierarchical clustering tree based on the metabolite profiles of 72 h. The same colored text for some of the gene names represents they were clustered in metabolic level and shared common target proteins.

Regulatory function of gene *sll1872* (*lytR*) has not been determined previously. According to our metabolomic analysis, several metabolites related to energy metabolism, including NADP, NADH, NAD and ATP were all down-regulated at both sampling time points, suggesting that the absence of the *sll1872* (*lytR*) may negatively affect the energy metabolism and then reduced the growth of △*sll1872* mutant (**Figure 3**).

Interestingly, although the growth of △*slr0741* was clearly slower than the wild type, the metabolomic analysis showed no obvious difference at the metabolite level between the △*slr0741* mutants and the wild type, implying that its regulatory function may be not directly related to the central carbohydrate metabolism.

Functional Conservation of TRs

The finding that a range of metabolic changes occurred in single-deletion TR mutants led to several immediate questions. First, whether the TR mutants with similar metabolic profiles have a close relationship on the evolutionary tree for their encoding genes? Second, whether the TR-encoding genes with a close evolutionary relationship have similar metabolic changes to the gene deletion? Answers to these questions could provide clues to possible function, evolution and origin of the TR genes. To seek answers to the questions, a gene families' analysis was conducted using full protein sequences of 32 TR genes. The homology analysis showed that 32 TR genes were classified into several different families, indicating different origin during the evolution of TR genes (**Figure 4** and Supplementary Table S3). Meanwhile, several pairs of TRs were found in the same gene family, suggesting possible events of gene duplication in recent evolutionary courses. Interestingly, while comparing the trees resulted from the hierarchical clustering analysis of metabolomic data and the gene families, we found that several pairs of TRs were clustered together in same gene family and the relationship tree generated using metabolomic profiles, suggesting possible functional conservation of the genes during genome evolution (**Figure 4**). At 48 h, one pair of TR genes, *slr1489* (*pchR*) and *sll1408* (*pcrR*), with similar function and evolution was identified; however, when metabolic profiles of 72 h were used, five pairs of TRs were identified. The pairs were *slr0895* (*prqR*) and *sll1286*; *slr1489* (*pchR*) and *sll1408* (*pcrR*); *sll1712* and *sll1670* (*hrcA*); *slr0115* (*rpaA*) and *slr0449* (*dnr*); *slr1871* (*pirR*) and *slr1245*. Only one pair of TRs, *slr1489* (*pchR*) and *sll1408* (*pcrR*), was identified when using metabolomic profiling data of both time points, suggesting a very conserved role that they might be playing. The difference between 48 and 72 h was probably due to the phase-dependent regulation of TRs in *Synechocystis*, which has been commonly reported in various microbes (Zhu et al., 2015).

Meanwhile, the results also showed a total of 17 pairs of TRs that were clustered together in the metabolomic trees but did not locate in the same gene family (**Figure 4**). Although experimental measurements of more metabolites are still necessary, the preliminary analysis pointed to the possibility of functional convergence of TR genes during the genome evolution, as the results showed that deletion of TR genes of different evolutionary origins caused similar metabolic responses in the mutant cells. For example, the *slr0395* (*ntcB*) and *slr0724* (*sohA*) genes were clustered together in the tree resulted from metabolic profiles of 48 h, deletion of these two TR genes caused CoA and GAP were down-regulation while ADP and ADP-glucose were up-regulated, although they belong to different gene families.

Functional Inference of TRs

Experimental and computational data from genome-scale PPI analysis has contributed significantly to the understanding of the gene function (Marcotte et al., 1999; Ikeuchi and Tabata, 2001;

Sato et al., 2007). In this study, an attempt was also made to apply PPI network of *Synechocystis* to determine possible gene targets of the five pairs of TRs identified above with similar clustering patterns in metabolomic profiles and matching gene family membership. As TRs of each pair have similar evolutionary and metabolomics patterns, it was expected that they might function through the same targets. Based on this hypothesis, we then implemented a strategy to first determine regulatory targets of each TR using PPI network analysis, and then identify the common target genes for every pair of TRs. The analysis allowed inference of possible regulatory gene targets for four out of five pairs of TRs, although no common target gene was identified for the *slr1871* (*pirR*) and *slr1245* pair (Supplementary Table S2).

(i) *sll1408* (*pcrR*) and *slr1489* (*pchR*): *sll1408* is a gene encoding a regulator protein PcrR and repressed by cold sensor Hik33 in *Synechocystis* (Suzuki et al., 2001). A previous study found that *sll1408* and *slr1489* were regulated by iron (and cadmium) and hydrogen peroxide stresses under control of the Slr1738 regulator (Houot et al., 2007). The two mutants did not show growth difference. Our analysis showed that Sll1408 and Slr1489 had four common possible target proteins: Slr0707, Slr0664, Sll0043, and Slr0594, among which Slr0707 is DNA polymerase I involved in the biological process of DNA replication (Sato et al., 2003), *slr0664* encodes a toxin–antitoxin (TA) system belonging to *rel* family (Ning et al., 2011), *sll0043* encodes a protein related to phototaxis (Shin et al., 2008), and *slr0594* was a putative membrane-spanning component of ABC transporters (Wipat et al., 1998), respectively.

(ii) *slr0895* (*prqR*) and *sll1286*: Previous studies showed that the gene *sll1286* was down-regulated by acid stress and sorbitol stress (Ohta et al., 2005; Shapiguzov et al., 2005), while *slr0895* (*prqR*) was found to be an auto-repressor regulating the adaptive responses to the oxidative stress (Cheng and He, 2014). In addition, *sll1286* and *slr0895* (*prqR*) were both encoding regulators of TetR family. As one TetR family regulator protein PfsR has been found as a key regulator of iron homeostasis in *Synechocystis*, and the PrqR (Slr0895) was an auto-repressor regulating the adaptive responses to oxidative stress (Cheng and He, 2014). No growth difference was found between these two mutants. Our analysis showed that Slr0895 and Sll1286 had two common possible target proteins: Slr1705 and Sll0782. *slr1705* (*aspA*) encodes aspartoacylase (Osanai et al., 2014), while *sll0782* is annotated as a helix–turn–helix DNA binding motif without any known function (Zhang et al., 1998). Based on COG and KEGG annotation, Slr1705 may be participating in amino acid transport and metabolism, such as alanine, aspartate and glutamate, which is consistent with the down-regulation of Glu at 72 h in the deletion mutants of *slr0895* (*prqR*) and *sll1286*.

(iii) *sll1670* (*hrcA*) and *sll1712*: *sll1670* gene encodes an ortholog of HrcA, a negative regulator of heat stress genes (Zorina et al., 2011). Sll1712 is a DNA binding protein (Chen et al., 2014) that was positively regulated by cadmium (Houot et al., 2007). The two mutants did not show growth difference. Our analysis showed that Sll1670 and Sll1712 had two common target proteins: Slr0701 and Sll0794. Slr0701 is a mercuric resistance operon regulator (Hirosawa et al., 1997), *sll0794* (*corR*) encoding a sensor gene involved in Ni^{2+}, Co^{2+}, and Zn^{2+} sensing and tolerance (García-Domínguez et al., 2000; Mehta et al., 2014), and tolerance to ethanol (Huertas et al., 2014).

(iv) *slr0115* (*rpaA*) and *slr0449* (*dnr*): *slr0449* (*dnr*) encodes a TR belonging to the Crp/Fnr family, which has been found regulated by AbrB2 (Leplat et al., 2013). Slr0115 is related to energy transfer from phycobilisomes to photosystems (Hanke et al., 2011), and deletion of *slr0115* (*rpaA*) resulted in increased efficiency of energy transfer from phycobilisomes to photosystem II relative to photosystem I (Ashby and Mullineaux, 1999). No growth difference was found between these two mutants. Our analysis showed that Slr0115 and Slr0449 had four common target proteins: Sll1196, Sll0745, Slr0884, and Sll1342, among which *sll1196* (*pfkA*) and *sll0745* (*pfkA*) encode two phosphofructokinases (Osanai et al., 2005), and participate in carbohydrate transport and metabolism, *slr0884* (*gap1*) and *sll1196* (*pfkA*) showed similar enhancement of expression through overexpression of *rre37* (*sll1330*) (Okada et al., 2015), while *sll1342* (*gap2*) encodes glyceraldehyde-3-phosphate dehydrogenase whose pathway involves in F6P, GAP, and R5P (Rowland et al., 2011; Lee et al., 2015). As a phosphofructokinase encoding gene, the expression of *sll1196* (*pfkA*) could affect the accumulation of F6P (Tabei et al., 2007), while the components GAP and FBP that were related to F6P phosphorylation were also changed.

In this study, 32 knockout mutants for putative TR-encoding genes of *Synechocystis* were constructed and comparatively analyzed via LC-MS-based metabolomics. Four mutants, *sll1872 (lytR)*, *slr0741 (phoU)*, *slr0395 (ntcB)*, and *slr1871 (pirR)*, showed differential growth patterns in BG11 medium when compared with the wild type. In the remaining TR mutants that did not show growth difference compared with the wild type, metabolomic profiling showed that they were clearly different at the metabolite level, suggesting significant functional diversity of TRs in *Synechocystis*. Finally, protein-protein interaction network analysis predicted possible regulatory targets of TRs.

AUTHOR CONTRIBUTIONS

LC and WZ conceived and designed the study. MS and XZ performed the experiments. MS, XZ, GP, LC, and WZ analyzed the data and wrote the manuscript. All authors read and approved the manuscript.

FUNDING

The research was supported by grants from the Natural Science Foundation of China (No. 31470217 and No. 21621004),

National Basic Research Program of China (National "973" program, project No. 2014CB745101), and the Tianjin Municipal Science and Technology Commission (No. 15JCZDJC32500).

ACKNOWLEDGMENT

The authors would also like to thank Mingyang Zhang and Siqiang Huang of our laboratory for their helps with constrution ofthe TR mutants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00280/full#supplementary-material

FIGURE S1 | Euclidean distances between each mutant and its control. **(A)** Calculated based on the metabolite profiles of 48 h; **(B)** Calculated based on the metabolite profiles of 72 h.

FIGURE S2 | Growth curves of 28 mutants without different growth with the wild type. The black block represents for the wild type and the red dot represents for corresponding mutant.

REFERENCES

Aichi, M., Takatani, N., and Omata, T. (2001). Role of NtcB in activation of nitrate assimilation genes in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 183, 5840–5847. doi: 10.1128/JB.183.20.5840-5847.2001

Ashby, M. K., and Mullineaux, C. W. (1999). Cyanobacterial ycf27 gene products regulate energy transfer from phycobilisomes to photosystems I and II. *FEMS Microbiol. Lett.* 181, 253–260. doi: 10.1111/j.1574-6968.1999.tb08852.x

Atsumi, S., Higashide, W., and Liao, J. C. (2009). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* 27, 1177–1180. doi: 10.1038/nbt.1586

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., and Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14, 283–291. doi: 10.1016/j.sbi.2004.05.004

Badger, M. R., Price, G. D., Long, B. M., and Woodger, F. J. (2006). The environmental plasticity and ecological genomics of the cyanobacterial CO2 concentrating mechanism. *J. Exp. Bot.* 57, 249–265. doi: 10.1093/jxb/eri286

Burnap, R. L., Hagemann, M., and Kaplan, A. (2015). Regulation of CO2 concentrating mechanism in cyanobacteria. *Life (Basel)* 5, 348–371. doi: 10.3390/life5010348

Chen, L., Zhu, Y., Song, Z., Wang, J., and Zhang, W. (2014). An orphan response regulator Sll0649 involved in cadmium tolerance and metal homeostasis in photosynthetic *Synechocystis* sp. PCC 6803. *J. Proteom.* 103, 87–102. doi: 10.1016/j.jprot.2014.03.029

Cheng, D., and He, Q. (2014). PfsR is a key regulator of iron homeostasis in *Synechocystis* PCC 6803. *PLoS ONE* 9:e101743. doi: 10.1371/journal.pone.0101743

Daley, S. M., Kappell, A. D., Carrick, M. J., and Burnap, R. L. (2012). Regulation of the cyanobacterial CO2-concentrating mechanism involves internal sensing of NADP+ and α-ketogutarate levels by transcription factor CcmR. *PLoS ONE* 7:e41286. doi: 10.1371/journal.pone.0041286

Deu-Pons, J., Schroeder, M. P., and Lopez-Bigas, N. (2014). jHeatmap: an interactive heatmap viewer for the web. *Bioinformatics* 30, 1757–1758. doi: 10.1093/bioinformatics/btu094

Foster, J. S., Singh, A. K., Rothschild, L. J., and Sherman, L. A. (2007). Growth-phase dependent differential gene expression in *Synechocystis* sp. strain PCC 6803 and regulation by a group 2 sigma factor. *Arch. Microbiol.* 187, 265–279. doi: 10.1007/s00203-006-0193-6

García-Domínguez, M., Lopez-Maury, L., Florencio, F. J., and Reyes, J. C. (2000). A gene cluster involved in metal homeostasis in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 182, 1507–1514.

Hanke, G. T., Satomi, Y., Shinmura, K., Takao, T., and Hase, T. (2011). A screen for potential ferredoxin electron transfer partners uncovers new, redox dependent interactions. *Biochim. Biophys. Acta* 1814, 366–374. doi: 10.1016/j.bbapap.2010.09.011

Hihara, Y., Muramatsu, M., Nakamura, K., and Sonoike, K. (2004). A cyanobacterial gene encoding an ortholog of Pirin is induced under stress conditions. *FEBS Lett.* 574, 101–105. doi: 10.1016/j.febslet.2004.06.102

Hirosawa, M., Isono, K., Hayes, W. S., and Borodovsky, M. (1997). Gene identification and classification in the *Synechocystis* genomic sequence by recursive gene mark analysis. *DNA Seq.* 8, 17–29. doi: 10.3109/10425179709020881

Houot, L., Floutier, M., Marteyn, B., Michaut, M., Picciocchi, A., Legrain, P., et al. (2007). Cadmium triggers an integrated reprogramming of the metabolism of *Synechocystis* PCC6803, under the control of the Slr1738 regulator. *BMC Genomics* 8:350. doi: 10.1186/1471-2164-8-350

Huertas, M. J., Lopez-Maury, L., Giner-Lamia, J., Sanchez-Riego, A. M., and Florencio, F. J. (2014). Metals in cyanobacteria: analysis of the copper, nickel, cobalt and arsenic homeostasis mechanisms. *Life* 4, 865–886. doi: 10.3390/life4040865

Huffman, J. L., and Brennan, R. G. (2002). Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* 12, 98–106. doi: 10.1016/S0959-440X(02)00295-6

Ikeuchi, M., and Tabata, S. (2001). *Synechocystis* sp. PCC 6803—a useful tool in the study of the genetics of cyanobacteria. *Photosynth. Res.* 70, 73–83. doi: 10.1023/A:1013887908680

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., et al. (2009). STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416. doi: 10.1093/nar/gkn760

Juntarajumnong, W., Hirani, T. A., Simpson, J. M., Incharoensakdi, A., and Eaton-Rye, J. J. (2007). Phosphate sensing in *Synechocystis* sp. *PCC* 6803: SphU and the SphS-SphR two-component regulatory system. *Arch. Microbiol.* 188, 389–402. doi: 10.1007/s00203-007-0259-0

Kaneko, T., Nakamura, Y., Sasamoto, S., Watanabe, A., Kohara, M., Matsumoto, M., et al. (2003). Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. *PCC* 6803. *DNA Res.* 10, 221–228.

Laiakis, E. C., Morris, G. A., Fornace, A. J. Jr., and Howie, S. R. (2010). Metabolomic analysis in severe childhood pneumonia in the Gambia, West Africa: findings from a pilot study. *PLoS ONE* 5:e12655. doi: 10.1371/journal.pone.0012655

Lee, D.-G., Kwon, J., Eom, C.-Y., Kang, Y.-M., Roh, S. W., Lee, K.-B., et al. (2015). Directed analysis of cyanobacterial membrane phosphoproteome using stained phosphoproteins and titanium-enriched phosphopeptides. *J. Microbiol.* 53, 279–287. doi: 10.1007/s12275-015-5021-8

Leplat, C., Champeimont, R., Saenkham, P., Cassier-Chauvat, C., Jean-Christophe, A., and Chauvat, F. (2013). Genome-wide transcriptome analysis of hydrogen production in the cyanobacterium *Synechocystis*: towards the identification of new players. *Int. J. Hydrogen Energy* 38, 1866–1872. doi: 10.1016/j.ijhydene.2012.11.118

Los, D. A., Zorina, A., Sinetova, M., Kryazhov, S., Mironov, K., and Zinchenko, V. V. (2010). Stress sensors and signal transducers in cyanobacteria. *Sensors* 10, 2386–2415. doi: 10.3390/s100302386

Marcotte, E. M., Pellegrini, M., Ng, H.-L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* 285, 751–753. doi: 10.1126/science.285.5428.751

Mehta, A., Lopez-Maury, L., and Florencio, F. J. (2014). Proteomic pattern alterations of the cyanobacterium *Synechocystis* sp. PCC 6803 in response to cadmium, nickel and cobalt. *J. Proteomics* 102, 98–112. doi: 10.1016/j.jprot.2014.03.002

Michel, K. P., and Pistorius, E. K. (2004). Adaptation of the photosynthetic electron transport chain in cyanobacteria to iron deficiency: the function of IdiA and IsiA. *Physiol. Plant.* 120, 36–50. doi: 10.1111/j.0031-9317.2004.0229.x

Morohoshi, T., Maruo, T., Shirai, Y., Kato, J., Ikeda, T., Takiguchi, N., et al. (2002). Accumulation of inorganic polyphosphate in *phoU* mutants of *Escherichia coli* and *Synechocystis sp*. Strain PCC6803. *Appl. Environ. Microbiol.* 68, 4107–4210. doi: 10.1128/AEM.68.8.4107-4110.2002

Ning, D., Ye, S., Liu, B., and Chang, J. (2011). The proteolytic activation of the relNEs (ssr1114/slr0664) toxin–antitoxin system by both proteases Lons and ClpP2s/Xs of *Synechocystis* sp. PCC 6803. *Curr. Microbiol.* 63, 496–502. doi: 10.1007/s00284-011-0011-5

Niu, X., Zhu, Y., Pei, G., Wu, L., Chen, L., and Zhang, W. (2015). Elucidating butanol tolerance mediated by a response regulator Sll0039 in *Synechocystis* sp. *PCC* 6803 using a metabolomic approach. *Appl. Microbiol. Biotechnol.* 99, 1845–1857. doi: 10.1007/s00253-015-6374-y

Ohta, H., Shibata, Y., Haseyama, Y., Yoshino, Y., Suzuki, T., Kagasawa, T., et al. (2005). Identification of genes expressed in response to acid stress in *Synechocystis* sp. PCC 6803 using DNA microarrays. *Photosynth. Res.* 84, 225–230. doi: 10.1007/s11120-004-7761-x

Okada, K., Horii, E., Nagashima, Y., Mitsui, M., Matsuura, H., Fujiwara, S., et al. (2015). Genes for a series of proteins that are involved in glucose catabolism are upregulated by the Hik8-cascade in *Synechocystis* sp. PCC 6803. *Planta* 241, 1453–1462. doi: 10.1007/s00425-015-2270-z

Osanai, T., Kanesaki, Y., Nakano, T., Takahashi, H., Asayama, M., Shirai, M., et al. (2005). Positive regulation of sugar catabolic pathways in the cyanobacterium *Synechocystis* sp. PCC 6803 by the group 2 sigma factor sigE. *J. Biol. Chem.* 280, 30653–30659. doi: 10.1074/jbc.M505043200

Osanai, T., Oikawa, A., Iijima, H., Kuwahara, A., Asayama, M., Tanaka, K., et al. (2014). Metabolomic analysis reveals rewiring of *Synechocystis* sp. *PCC* 6803 primary metabolism by ntcA overexpression. *Environ. Microbiol.* 16, 3304–3317. doi: 10.1111/1462-2920.12554

Ren, Q., Shi, M., Chen, L., Wang, J., and Zhang, W. (2014). Integrated proteomic and metabolomic characterization of a novel two-component response regulator Slr1909 involved in acid tolerance in *Synechocystis* sp. PCC 6803. *J. Proteomics* 109, 76–89. doi: 10.1016/j.jprot.2014.06.021

Rowland, J. G., Simon, W. J., Prakash, J. S., and Slabas, A. R. (2011). Proteomics reveals a role for the RNA helicase crhR in the modulation of multiple metabolic pathways during cold acclimation of *Synechocystis* sp. PCC6803. *J. Proteome Res.* 10, 3674–3689. doi: 10.1021/pr200299t

Sato, N., Terasawa, K., Miyajima, K., and Kabeya, Y. (2003). Organization, developmental dynamics, and evolution of plastid nucleoids. *Int. Rev. Cytol.* 232, 217–262. doi: 10.1016/S0074-7696(03)32006-6

Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007). A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* 14, 207–216. doi: 10.1093/dnares/dsm021

Shapiguzov, A., Lyukevich, A. A., Allakhverdiev, S. I., Sergeyenko, T. V., Suzuki, I., Murata, N., et al. (2005). Osmotic shrinkage of cells of *Synechocystis* sp. *PCC* 6803 by water efflux via aquaporins regulates osmostress-inducible gene expression. *Microbiology* 151(Pt 2), 447–455. doi: 10.1099/mic.0.27530-0

Shin, B.-J., Oh, J., Kang, S., Chung, Y.-H., Park, Y. M., Kim, Y. H., et al. (2008). Cyanobacterial hybrid kinase Sll0043 regulates phototaxis by suppressing pilin and twitching motility protein. *J. Microbiol.* 46, 300–308. doi: 10.1007/s12275-007-0212-6

Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M., and Murata, N. (2001). Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*. *Mol. Microbiol.* 40, 235–244. doi: 10.1046/j.1365-2958.2001.02379.x

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39, D561–D568. doi: 10.1093/nar/gkq973

Tabei, Y., Okada, K., and Tsuzuki, M. (2007). Sll1330 controls the expression of glycolytic genes in *Synechocystis* sp. *PCC* 6803. *Biochem. Biophys. Res. Commun.* 355, 1045–1050. doi: 10.1016/j.bbrc.2007.02.065

Thelwell, C., Robinson, N. J., and Turner-Cavet, J. S. (1998). An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10728–10733. doi: 10.1073/pnas.95.18.10728

Wang, D., Zhang, Y., Pohlmann, E. L., Li, J., and Roberts, G. P. (2011). The poor growth of *Rhodospirillum rubrum* mutants lacking RubisCO is due to the accumulation of ribulose-1, 5-bisphosphate. *J. Bacteriol.* 193, 3293–3303. doi: 10.1128/JB.00265-11

Wang, J., Chen, L., Huang, S., Liu, J., Ren, X., Tian, X., et al. (2012). RNA-seq based identification and mutant validation of gene targets related to ethanol resistance in cyanobacterial *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* 5:89. doi: 10.1186/1754-6834-5-89

Wang, Y., Shi, M., Niu, X., Zhang, X., Gao, L., Chen, L., et al. (2014). Metabolomic basis of laboratory evolution of butanol tolerance in photosynthetic *Synechocystis* sp. PCC 6803. *Microb. Cell Fact.* 13, 151. doi: 10.1186/s12934-014-0151-y

Wipat, A., Brignell, S. C., Guy, B. J., Rose, M., Emmerson, P. T., and Harwood, C. R. (1998). The yvsA-yvqA (293-289) region of the *Bacillus subtilis* chromosome containing genes involved in metal ion uptake and a putative sigma factor. *Microbiology* 144, 1593–1600. doi: 10.1099/00221287-144-6-1593

Zhang, C.-C., Gonzalez, L., and Phalip, V. (1998). Survey, analysis and genetic organization of genes encoding eukaryotic-like signaling proteins on a cyanobacterial genome. *Nucleic Acids Res.* 26, 3619–3625. doi: 10.1093/nar/26.16.3619

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology* 156, 287–301. doi: 10.1099/mic.0.034793-0

Zhu, Y., Pei, G., Niu, X., Shi, M., Zhang, M., Chen, L., et al. (2015). Metabolomic analysis reveals functional overlapping of three signal transduction proteins in regulating ethanol tolerance in cyanobacterium *Synechocystis* sp. PCC 6803. *Mol. Biosyst.* 11, 770–782. doi: 10.1039/c4mb00651h

Zorina, A., Stepanchenko, N., Novikova, G. V., Sinetova, M., Panichkin, V. B., Moshkov, I. E., et al. (2011). Eukaryotic-like Ser/Thr protein kinases SpkC/F/K are involved in phosphorylation of GroES in the cyanobacterium *Synechocystis*. *DNA Res.* 18, 137–151. doi: 10.1093/dnares/dsr006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Shi, Zhang, Pei, Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 February 2018
doi: 10.3389/fmicb.2018.00295

Diversity of Gene Clusters for Polyketide and Nonribosomal Peptide Biosynthesis Revealed by Metagenomic Analysis of the Yellow Sea Sediment

*Yongjun Wei, Lei Zhang, Zhihua Zhou and Xing Yan**

Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (CAS), Shanghai, China

Polyketides (PKs) and nonribosomal peptides (NRPs) are widely applied as drugs in use today, and one potential source for novel PKs and NRPs is the marine sediment microbes. However, the diversities of microbes and their PKs and NRPs biosynthetic genes in the marine sediment are rarely reported. In this study, 16S rRNA gene fragments of the Yellow Sea sediment were analyzed, demonstrating that *Proteobacteria* and *Bacteroidetes* accounted for 62% of all the bacterial species and *Actinobacteria* bacteria which were seen as the typical PKs and NRPs producers only accounted for 0.82% of all the bacterial species. At the same time, PKs and NRPs diversities were evaluated based on the diversity of gene fragments of type I polyketide synthase (PKS) ketosynthase domain (KS), nonribosomal peptide synthetase (NRPS) adenylation domain (AD), and dTDP-glucose-4,6-dehydratase (dTGD). The results showed that AD genes and dTGD genes were abundant and some of them had less than 50% identities with known ones; By contrast, only few KS genes were identified and most of them had more than 60% identities with known KS genes. Moreover, one 70,000-fosmid clone library was further constructed to screen for fosmid clones harboring PKS or NRPS gene clusters of the Yellow Sea sediment. Nine selected fosmid clones harboring KS or AD were sequenced, and three of the clones were assigned to *Proteobacteria*. Though only few *Actinobacteria* 16S rRNA gene sequences were detected in the microbial community, five of the screened fosmid clones were assigned to *Actinobacteria*. Further assembly of the 9 fosmid clones resulted in 11 contigs harboring PKS, NRPS or hybrid NPRS-PKS gene clusters. These gene clusters showed less than 60% identities with the known ones and might synthesize novel natural products. Taken together, we revealed the diversity of microbes in the Yellow Sea sediments and found that most of the microbes were uncultured. Besides, evaluation of PKS and NRPS biosynthetic gene clusters suggested that the marine sediment might have the ability to synthesize novel natural products and more NRPS gene clusters than PKS gene clusters distributed in this environment.

Keywords: PKS-I and NRPS diversity, gene cluster, biosynthesis, metagenomics, marine sediment

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Paul Race, University of Bristol, United Kingdom
Gang Liu, Chinese Academy of Sciences (CAS), China

****Correspondence:***
Xing Yan
yanxing@sibs.ac.cn

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *26 December 2017*
Accepted: *08 February 2018*
Published: *27 February 2018*

Citation:
Wei Y, Zhang L, Zhou Z and Yan X (2018) Diversity of Gene Clusters for Polyketide and Nonribosomal Peptide Biosynthesis Revealed by Metagenomic Analysis of the Yellow Sea Sediment. Front. Microbiol. 9:295. doi: 10.3389/fmicb.2018.00295

INTRODUCTION

Natural products of polyketides (PKs) and nonribosomal peptides (NRPs) are secondary metabolites of microbes, which can help microbes adapt to environment and resist to stressful natural conditions. Until now, more than 23,000 natural products of PKs and NRPs had been identified and characterized, and they had been widely use as antibiotic and antitumor medicine (Bérdy, 2005; Walsh, 2007; Demain and Sanchez, 2009; Newman and Cragg, 2012; Katz and Baltz, 2016). Additionally, many previously identified PKs and NRPs were recovered from isolated bacteria of terrestrial environments (Handelsman et al., 1998; Daniel, 2004), especially from the phylum of *Actinobacteria* (Bérdy, 2005; Fenical and Jensen, 2006; Bull and Stach, 2007). However, nowadays, PKs and NRPs recovered from easily cultured microbes are often proven to be the same ones previously identified, suggesting the possibility of getting novel PKs and NRPs by traditional cultivation-dependent method dramatically decreases (Tulp and Bohlin, 2005; Xiong et al., 2015).

The oceans, which cover more than 70% of the earth, represent a rich source of valuable novel natural products (Molinski et al., 2009). More than 1,000 novel compounds had been identified from the ocean in the past few years (Blunt et al., 2013, 2014, 2015). It was estimated that *Actinobacteria* was the main source of bioactive PKs and NRPs in the ocean (Fiedler et al., 2005; Jensen et al., 2005). 16S rRNA analysis showed that *Actinobacteria* bacteria were not the most abundant phyla in the marine sediment, suggesting there might be other potential PKs and NRPs producer which can produce novel PKs and NRPs (Zhu et al., 2013). Moreover, 99% of the bacteria were recalcitrant to cultivate, indicating that uncultured marine microbes might be the potential reservoir to discover novel natural products (Torsvik et al., 1996; Whitman et al., 1998; Rappé and Giovannoni, 2003). Besides, different natural product biosynthetic genes are distributed in phylogenetic similar microbial communities (Reddy et al., 2012), showing the unrevealed marine environments are the attractive starting points to recover novel bioactive compounds (Brady et al., 2002; Piel, 2011; Wilson and Piel, 2013).

PKs are mainly synthesized by type I polyketide synthase (PKS-I) and type II polyketide synthase (PKS-II) gene clusters, and gene fragment of PKS-I ketosynthase (KS) domain and PKS-II KSα domain are often used for PKs diversity evaluation. In the meanwhile, NRPs are synthesized by nonribosomal peptide synthetase (NRPS), and NPRS adenylation (AD) domain is used for NRPs diversity evaluation (Reddy et al., 2012). Nowadays, culture-independent metagenomic methods have successfully used for discovering of novel natural product biosynthetic gene clusters from diverse environments, and can be used to evaluate natural product biosynthetic gene diversity in the marine sediment (Brady et al., 2009; Wilson and Piel, 2013). Deep sequencing of NRPS and PKS gene fragments in the marine sponges confirmed that only a small amount of the recovered genes assigned to *Actinobacteria*, suggesting that PKs and NRPs recovered from the marine environment would be different from previously known ones mainly produced by *Actinobacteria* (Woodhouse et al., 2013). Moreover, the dTDP-glucose-4,6-dehydratase (dTGD) which can glycosylate natural products by 6-deoxyhexose (6DOH) are usually parts of natural product biosynthetic gene clusters and involved in secondary metabolism, and can be used to evaluate 6DOH-modified NRPs and PKs diversities in the environments (Thibodeaux et al., 2007; Bruender et al., 2010; Chen et al., 2011).

In this study, gene fragments of 16S rRNA, NRPS AD domain, PKS-I KS domain and dTGD were sequenced to evaluate bacteria and their natural product biosynthetic gene diversities in the Yellow Sea sediment. In order to recover natural product biosynthetic genes, one fosmid clone library were constructed using same DNA extracted from the Yellow Sea sediment. Nine selected fosmid clones harboring natural product biosynthetic genes were recovered and their natural product biosynthetic gene clusters were analyzed.

MATERIALS AND METHODS

Sample Collection, DNA Isolation and the Fosmid Clone Library Construction

The marine sediment samples were collected at the depth of 50–100 m in the summer of 2010 from several adjacent sites of undisturbed environment in the Yellow Sea sediment close to Rizhao, Shangdong province, China (Xiong et al., 2015). They were stored in sterilized plastic bags after being taken from the deep sea and transported to the laboratory at 4°C. As there might be some hazard microbes in the marine samples, the extraction experiment was carried out according to safety procedures of our institute. The sediment samples were mixed together, and total DNA was extracted from mixed sediment samples by a previously described protocol (Zhou et al., 1996; Geng et al., 2012). 10 ml PBS buffer was added to the mixed sediment samples and large particulates were removed by briefly centrifuging at 200 × g for 5 min. Cell pellets were obtained by further centrifugation (9,000 × g, 5 min), and 7 ml extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM EDTA [pH 8.0], 1.5 M NaCl, 100 mM Na_3PO_4,1% cetyl trimethyl-ammonium bromide [wt/vol], 1% sodium dodecyl sulfate [wt/vol], 1 mg/ml proteinase K) were added to suspend the cell pellets. The suspension was incubated at 55°C for 20 min and then 70°C for 10 min. The crude lysate was centrifuged at 17,000 × g for 10 min, and the cellular debris was pelleted. The supernatant of the crude lysate was moved to new centrifuge tubes and extracted with phenol/chloroform/isoamyl alcohol (25:24:1) for twice, followed by extraction with chloroform/isoamyl alcohol (24:1) for one time. The crude DNA was precipitated from the supernatant with the addition of 0.6 volumes of isopropyl alcohol and collected by centrifugation (5,400 × g, 10 min). The pelleted DNA was washed with 70% ethanol and resuspended with 200 μl TE buffer (10 mM Tris, 1 mM EDTA [pH 8.0]) containing 10 μg ribonuclease solution (RNase).

The high molecular-weight DNA (about 40 kb) was separated by CHEF electrophessis (Bio-rad, USA) and electroeluted from the agarose. Some DNA was blunt ended, ligated into the

fosmid vectors, packaged into lambda phage and transfected into *Escherichia coli* EPI300 to construct a fosmid clone library according to the Copy-Control™ HTP Fosmid Library Production Kit manual (Epicentre, San Diego, CA, USA). The fosmid clone library contained more than 70,000 clones and all of them were preserved at 384-well plates. Besides, some DNA was used to amplify conserved domains of the natural product biosynthetic gene fragments and 16S rRNA gene fragments.

PCR Amplification of Gene Fragments from Marine Sediment Samples

Primers designed to recognize conserved regions in PKS-I ketosynthase (KS), NRPS adenylation (AD), PKS-II ketosynthase alpha (KSα) and dTGD genes were selected to investigate the natural product biosynthetic gene diversities in the marine sediment (**Table 1**; Metsä-Ketelä et al., 1999; Du et al., 2004; Ginolhac et al., 2004; Ayuso-Sacido and Genilloud, 2005). The 16S rRNA gene V1-V3 regions were amplified with primers 27f/P2 (Dong et al., 2011). For dTGD and 16S rRNA gene fragments, one adaptor (5′-CGTATCGCCTCCCTCGCGCCATCAG-3′) was added to the forward primer dTGDF and 27f (Parameswaran et al., 2007). To sequence different genes in one run and fetch the genes from the sequencing results, specific barcodes were added to each primer as described in **Table 1**.

For all the genes, LA Taq polymerase (Takara Biotechnology, Dalian) was used as the amplification polymerase and gradient annealing temperature was tried to amplify the corresponding gene fragments from the marine sediment. However, no PKS-II KSα sequences were amplified from the marine sediment. Another three primer pairs of K1(5′-TSAAGTCSAACATCGGBCA)/M6R(5′-CGCAGGTTSCSGTACCAGTA), (5′-CCSCAGSAGCGCSTSTTSCTSGA)/(5′-GTSCCSGTSCCGTGSGTSTCSA), and (5′-CCSCAGSAGCGCSTSCTSCTSGA)/(5′-GTSCCSGTSCCGTGSGCCTCSA) had been tried for KS gene amplification, but failed (No targeted gene fragment for K1/M6R, and very weaker fragments for the other two primer pairs) (Courtois et al., 2003; Ayuso-Sacido and Genilloud, 2005). The optimum 25 μl PCR reaction mixture used for gene amplification contained 5 ng marine sediment DNA, 0.5 μM each primer, 200 μM each deoxynucleoside triphosphate (dNTP), 1*GC Buffer I (Takara Biotechnology, Dalian) and 0.2 U LA Taq polymerase (Takara Biotechnology, Dalian). For KSLF/R and A3/A7R primers, a PCR protocol with an initial denaturation (5 min, 94°C), followed by 30 cycles consisting of 30 s at 94°C, 30 s at 59°C and 45 s at 72°C, and finally one cycle of 10 min at 72°C was used. For dTGDF/dTGDR and 27f/P2 primers, a PCR protocol with an initial denaturation (5 min, 94°C), followed by 25 cycles consisting of 30 s at 94°C, 30 s at 55°C and 45 s at 72°C, and finally one cycle of 10 min at 72°C was used.

454 Pyrosequencing and Data Processing of 16S rRNA and Natural Product Biosynthetic Gene Fragments

PCR products of KS and AD gene fragments were gel purified using Axygen MinElute columns (Axygen Scientific Inc., USA) and equivalently mixed. Then two different 454-sequencing adaptors (Forward adaptor, 5′-CGTATCGCCTCCCTCGCGCCATCAG-3′ and reverse adaptor, 5′-CTATGCGCCTTGCCAGCCCGCTCAG-3′) were incorporated to the forward and reverse primers separately for parallel pyrosequencing by Roche 454 GS-FLX Titanium. For dTGD and 16S rRNA gene fragments, they were directly sequenced by Roche 454 GS-FLX Titanium. All the raw reads were deposited in the NCBI Sequence Read Archive (SRA) database under accession number of PRJNA403937.

Reads of these 4 gene fragments more than 400 bp were extracted by Acacia software with each barcode on the forward primers (Bragg et al., 2012). For KS and AD gene fragments, all the retrieved reads were trimmed to 400 bp from the forward primer site; for dTGD and 16S rRNA gene fragments, only reads with full-length (the reads should contain 27f/P2 or dTGDF/dTGDR primers) were selected for further analysis. Potential chimeric sequences of KS, AD and dTGD reads were filtered using Chimera filter in USEARCH with the commands of UCHIME denovo and ref (KS_REF, AD_REF and dTGD-REF, described below; Edgar, 2010; Edgar et al., 2011). Then KS reads were compared to the KS_REF databases, and any reads that did not align to a reference over a coverage of 90% and *E*-value of $<10e^{-10}$ were removed. For AD reads and dTGD reads were more divergent than KS reads, a cutoff *E*-value of $10e^{-5}$ was used in AD_REF and dTGD-REF databases searches. Finally, reads sorted by USEARCH were used for OTU classification (**Table 3**).

TABLE 1 | Primers used for amplification of conserved domains from natural product gene clusters.

Candidate genes used for amplification	Predicted length (bp)	Primer name	Primer sequences (5′ → 3′)	Barcode
Type I polyketide synthase KS domain fragments	700	KSLF	CCSCAGSAGCGCSTSYTSCTSGA	ACTCGTCT
		KSLR	GTSCCSGTSCCGTGSGYSTCSA	ACCTGTCT
Type II polyketide synthases KSα gene fragments	672	KSα-F	TSGCSTGCTTCGAYGCSATC	
		KSα-R	TGGAANCCGCCGAABCCGCT	
NRPS Adenylation domain fragments	795	A3	GCSTACSYSATSTACACSTCSGG	ATGCATGC
		A7R	SASGTCVCCSGTSCGGTAS	ATATATGC
dTDP-glucose-4,6-dehydratase gene fragment	530	dTGDF	GSGGSGSSGCSGGSTTCATSGG	TAGTAGTC
		dTGDR	GGGWRCTGGYRSGGSCCGTAGTTG	TATACATA
16S rRNA gene fragment	450	27f	AGATAAGAGTTTGATCMTGGCTCAG	ATCAGATC
		P2	ATTACCGCGGCTGCTGG	TGACACTA

Construction of Reference Sequence Databases for KS, AD, and dTGD Genes (KS_REF, AD_REF, and dTGD_REF)

Among all the databases collected information of the natural product biosynthetic genes, ClusterMine360 is an updated database which collects diverse microbial PKS and NRPS biosynthetic genes (Conway and Boddy, 2013). In order to generate a reference database of known AD and KS sequences, we downloaded all the gene clusters and domains from ClusterMine360 database. Then a primer-mapping search on all the domains was performed, only those sequences matched with KSLF/R or A3/A7R primers (allow for 3 mismatches) were selected and the sequences between the forward and reverse primers were retrieved. For all the sequences deposited in ClusterMine360 database were microbial PKS/NRPS biosynthetic genes, it was believed that the sequences retrieved from the database were authentic microbial KS and AD gene fragments. Though we used the same method to identify dTGD genes, only 7 sequences were retrieved from the ClusterMine360 database. To generate dTGD-REF with more dTGD gene sequences, another 51 reference functional dTGD sequences described before were retrieved from the NCBI database and added to the dTGD-REF (Chen et al., 2011). The KS_REF, AD_REF, and dTGD-REF databases contain 331 KS, 193 AD, and 58 dTGD gene sequences, respectively.

16S rRNA Gene Phylogenetic Analysis

The retrieved full-length 16S rRNA gene reads (with forward and reverse primers) were processed with QIIME analysis and all the obtained 16S rRNA gene reads were classified into operational taxonomic units (OTUs) with uclust method at 97% similarity (Caporaso et al., 2010; Kuczynski et al., 2011). Then the chimeric sequences were identified with ChimeraSlayer program. After removing of chimeric sequences, all the 16S rRNA gene reads and a representative sequence of each OTU were used for the RDP-based phylum level classification, respectively. OTUs with more than 50 reads were considered as abundant OTUs and blasted for closest isolates from RDP database in this study.

Assignment of Natural Product Biosynthetic Gene Fragments to Similarity-Based OTUs

The retrieved reads of KS, AD, and dTGD reads were aligned with MAFFT alignments and clustered at appropriate identity into OTUs using USEARCH (Chen et al., 2011; Reddy et al., 2012). In order to construct phylogenetic tree, representative reads of OTU in each dataset, natural product biosynthetic gene fragments extracted from the contigs (describe below) and the corresponding sequences in the reference databases were aligned with MAFFT alignments. Circular phylogenetic trees of KS, AD, and dTGD genes were constructed using MEGA 5.2 with Neighbor-joining method and edited with FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/; Tamura et al., 2011). For AD genes, only AD OTUs contained more than 3 reads were used for AD tree construction.

Type I PKSs can typically be classified into two types, acyltransferase (AT)-less type I PKSs and canonical type I PKSs. In order to identify the evolutionary lineage of recovered KS gene fragments, the representative reads of each KS OTUs and 6 KS gene fragment identified from the fosmid contigs were translated into protein sequences using meta-prodigal (Hyatt et al., 2010). Another 12 reference KS gene sequences (9 AT-less type I PKSs KS gene fragments and 3 canonical type I PKSs KS gene fragments) were downloaded (Lohman et al., 2015). All the KS gene fragments were aligned with MAFFT alignments and used for phylogenetic tree construction using Maximum Likelihood method.

Screening, Sequencing and Annotation of Fosmid Clones Harboring Natural Product Biosynthetic Gene Fragments

Besides investigating the diversity of natural product biosynthetic genes, the primers of KSLF/R and A3/A7R were used to screen fosmid clones harboring the natural product biosynthesis genes. The PCR mixture (20 μl) used for clone screening was composed with 0.2 μM dNTPs (Takara Biotechnology, Dalian), 0.5 μM of each primer, 5 ng DNA or 2 μl of overnight culture as template, 5% dimethyl sulfoxide (DMSO), 0.5 U rTaq polymerase and the recommend rTaq buffer (Takara Biotechnology, Dalian). The screening PCR protocol was initially with one cycle of 94°C for 5 min, followed by 40 cycles of 94°C for 30 s, 59°C for 30 s and 72°C for 1 min, finally with one cycle of 10 min at 72°C. Fosmid DNA extracted from each 384-well plate was used as template to screen out plates harboring natural product biosynthetic gene fragments. The overnight culture collected from each column or row of the positive plates were mixed and 2 μl of them was used as PCR template to screen for the positive fosmid clones harboring KS or AD domain. Then 8 fosmid clones contained AD genes and 1 fosmid contained KS genes were sequenced using 454 pyrosequencing and 12 contigs were assembled from these 9 fosmid clones using Newbler software version 2.6. Some KS domains and AD domains were extracted from the sequences by primer-mapping search of the sequences as described above and used for phylogenetic tree construction (allow 4 mismatches in each primer). The taxonomy of 12 contigs was determined with PhyloPythiaS analysis and annotated with fgenesb and BlastX (Patil et al., 2012). All the contigs were submitted to antiSMASH for further PKS and NRPS gene cluster annotation (Weber et al., 2015). The 12 contigs were deposited in Genbank and the accession numbers of them were MF964193-MF964204.

RESULTS

Bacterial Diversity in the Marine Sediment

Bacterial diversity in the marine sediment was investigated using 16S rRNA gene fragments. A total of 7905 16S rRNA gene fragments were extracted from the original 454 pyrosequencing reads by Acacia software, and 6675 reads predicted to be full-length (with P2 and 27f primer in the reads) were used for further analysis. After removing of 100 chimera sequences, 6575

reads were classified into 1335 OTUs at 97% identity. The Good's coverage of the bacterial 16S rRNA genes was 89.08%, indicating that most bacteria of the marine sediment sample had been detected in this study. Based on analysis results of the 6575 16S rRNA reads and the representative reads of the 1355 OTUs, Proteobacteria were the most dominant phylum, which constituted 40.7% of the total 16S rRNA reads and 41.6% of all the representative OTU reads, respectively (**Figure 1**); followed by Bacteroidetes bacteria which composed 34.9% of total 16S rRNA reads and 20.2% of all the representative OTU reads, respectively. Only 0.3% of total 16S rRNA reads and 0.8% of the representative OTU were assigned as Actinobacteria bacteria which was believed to synthesize more than 45% of the known microbial PKs and NRPs in nature (Bérdy, 2005).

Of all the OTUs, 17 most abundant OTUs constituted 30.84% of the total 6575 16S rRNA gene reads (**Table 2**). Among them, OTU1242 constituted 8.29% and another 4 OTUs constituted more than 2% of the total reads. Interestingly, three of the 5 most abundant OTUs, OTU1242, OTU1041, and OTU1140, were assigned to Bacteroidetes, not assigned to the most abundant phylum of Proteobacteria in the sample. Most of the 17 abundant OTUs had less than 97% identity with their closest isolates, hinting that these abundant species in the marine sediment might be uncultured (**Table 2**). A total of 613 *Actinobacteria* stains had been isolated from the same marine sediment and 105 16S rRNA sequences had been released in previous study (Xiong et al., 2015). However, among all the 105 16S rRNA sequences, only three16S rRNA sequences (The Genbank accession numbers are JQ924069, JQ924085, and JQ924089, 100% identity of these 3 sequences) showed more than 95% identities with one OTU (containing 2 reads, and they showed 99.1 and 100% identities with these 3 sequences, respectively) identified in this study, suggesting that *Actinobacteria* bacteria identified by culture independent method were different from previously identified culturable *Actinobacrteria* in the same sample.

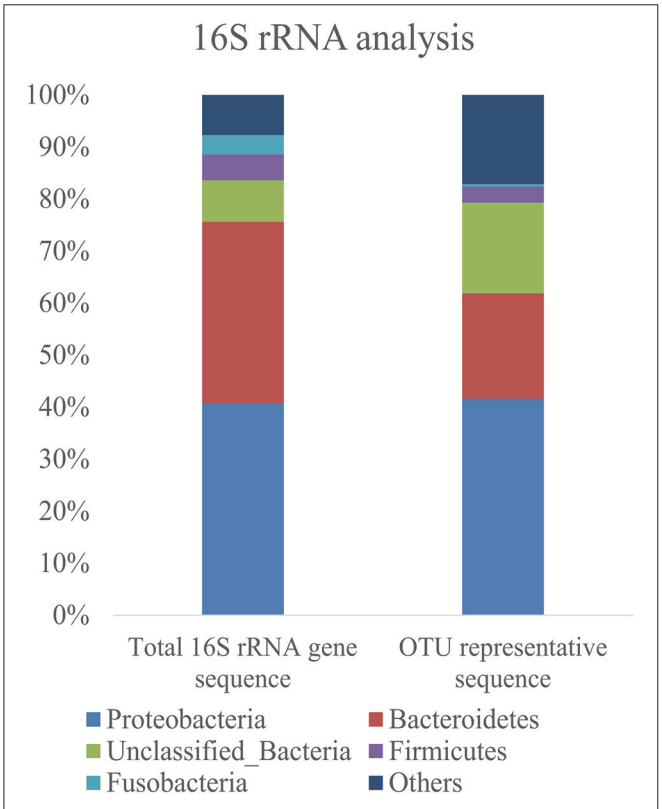


FIGURE 1 | Bar graphs show the frequency of bacteria phyla identified from the Yellow Sea sediment samples. The distribution of total 16S rRNA gene sequences and representative sequences of each OTU was shown.

Phylogenetic Analyses of KS Domain, AD Domain, and dTGD Gene Fragments

After processing of the retrieved reads with Acacia software, 9995 KS reads, 4042 AD reads and 602 dTGD reads were obtained (**Table 3**). The KS and AD reads were classified into OTUs at 90% identity, and the dTGD reads were classified into OTUs at 80% similarity (Reddy et al., 2012). The KS, AD and dTGD reads can be classified into 27 OTUs, 1087 OTUs, and 279 OTUs at their selected identities, respectively (**Table 3**), showing NRPS and dTGD diversity might be more abundant than that of PKS in the microbial community of the marine sediment.

The KS and AD gene fragments were clustered with their corresponding known reference gene fragments in the phylogenetic tree, however, all of them showed less than 80% identities with the reference KS genes, suggesting they were part of novel PKS gene cluster (**Figures 2**, **3**). In the phylogenetic trees, 7 KS representative reads (A total of 27 representative KS reads in the tree) and 66 AD representative reads (A total of 308 representative AD reads in the tree) were clustered with corresponding reference gene fragments of Actinobacteria, showing they were some potential natural product biosynthetic gene clusters of Actinobacteria bacteria (**Figures 2**, **3**). Only few dTGD reads were clustered with known reference Actinobacteria dTGD gene fragments, suggesting that most dTGD genes of the marine sediment sample might not derive from Actinobacteria and were different from the known ones (**Figure 4**). Further analyses of the KS gene fragment evolutionary lineage suggested that 11 KS representative reads were clustered with known AT-less type I PKS KSs and the other 16 KS representative reads were clustered with known canonical type I PKS KSs (Figure S2).

Screening and Sequencing of Fosmid Clones Harboring KS or AD Domains

The primers of KSLF/R and A3/A7R were used to screen fomsid clones harboring natural product biosynthetic genes from the constructed fosmid clone library. Similar to the diversity analysis, AD domains were more abundant than KS domains. A total of 43 PCR-positive 384-well plates containing AD genes and 28 PCR-positive 384-well plates containing KS gene fragments were identified from all the 188 384-well plates.

For AD domains were more abundant than KS domains, eight AD gene-containing and one KS gene-containing fosmid clones were selected for further sequencing to reveal natural

TABLE 2 | The 17 most abundant OTUs in Yellow Sea sediment samples and their closest named isolates.

OTU ID	Percentage (%)	Closest named isolates	Accession number	Identity (%)
OTU1242	8.29	*Gillisia* sp. NP17	EU196340.1	96.52
OTU225	2.59	*Bizionia paragorgiae* str. KMM 6029	NR_025827.1	99.57
OTU1041	2.59	*Alviniconcha* sp. gill symbiont	AB189712.1	93.7
OTU1140	2.46	*Polaribacter* sp. str. UDC421	HM032000.1	96.54
OTU46	2.08	*Ilyobacter psychrophilus* str. FQ50	AJ877255.1	94.2
OTU1354	1.66	*Sulfurovum* sp. str. NBC37-1	AP009179.1	96.46
OTU1166	1.38	*Pelobacter venetianus*	U41562.1	97.33
OTU863	1.19	*Sphingomonas* sp. BAC84	EU131006.1	93.42
OTU232	1.17	*Propionigenium maris* str. DSL-1	Y16799.1	94.2
OTU828	1.16	*Sedimentibacter* sp. str. IMPC3	EF189918.1	81.98
OTU44	1.00	*Pelobacter venetianus*	U41562.1	96.89
OTU1246	1.00	*Geosporobacter subterrenus* str. VNs68	DQ643978.1	92.67
OTU872	0.96	*Pelobacter venetianus*	U41562.1	97.36
OTU472	0.90	*Oligobrachia mashikoi* endosymbiont D	AB271122.1	97.4
OTU469	0.87	*Brumimicrobium* sp. str. P99	EU195945.1	97.84
OTU15	0.79	*Alviniconcha* sp. gill symbiont	AB189712.1	95.49
OTU708	0.76	*Sulfurimonas autotrophica* str. OK5	AB088432.1	95.22

TABLE 3 | Number of reads remained after each processing steps and number of OTUs classified based on selected identities.

	PKS KS domains	NRPS AD domains	dTGD genes
454 Reads fetched by Acacia	12,295	6,148	1,698
Reads >400 bp	10,943	4,899	1,221(full length)
No chimeras (de novo)	10,943	4,895	1,221
No chimeras (reference data)	10,943	4,895	1,221
Ref-homologs (BLAST)	9,995	4,042	602
OTUs	27 (90% identity)	1,087 (90% identity)	279 (80% identity)
Good's coverage	99.93%	85.75%	68.72%

product biosynthetic gene clusters in the marine sediment. Due to repetitive sequences in 3 fosmid clones, a total of 12 contigs named YFC1 to YFC12 were obtained from these 9 fosmid clones (**Table 4**). PhyloPythiaS analysis of these 12 contigs showed that five fosmid clones were assigned to Actinobacteria, three fosmid clones were assigned to Proteobacteria and one fosmid clone was assigned to unknown bacteria (**Table 4**; Patil et al., 2012).

Natural Product Biosynthetic Gene Analysis of the Contigs

Annotation of the 12 contigs showed that 11 contigs contained natural product biosynthetic gene clusters, among which, four contigs harbored PKS/NRPS hybrids gene clusters, one contig harbored PKS gene clusters, and six contigs harbored NRPS gene clusters (**Table 4** and **Figure 5**). Especially, one glycosyl transferase gene clustered with natural product biosynthetic genes was identified in contig YFC3 and one dTGD gene fragment YFC_dTGD(3787…4830) was identified in YFC11 (Figures S1A,C). Moreover, two contigs of YFC8 and YFC11 might have full-length NRPS gene clusters, because genes of natural product biosynthetic pathway usually clustered in the genome and there are some nonnatural product biosynthetic genes flanking around the NRPS gene clusters. Some of another nine contigs were predicted to harbor several completed PKS or NRPS modules, but not completed natural product gene clusters (Figures S1B,C). All the PKS or PKS/NRPS hydrids contigs harbored AT domains in their PKS modules, suggesting these contigs harbored canonical type I PKSs (Figure S2).

Six KS gene fragments, 12 AD gene fragments and one dTGD gene fragment were identified in the 11 contigs using the primer-mapping search. Among them, one AD gene fragment YFC3_AD(2358…3031) was the same as the most abundant AD gene fragment (AD1_7989size_270) identified by 454 pyrosequencing (**Figure 3**), but other AD, KS and dTGD gene fragments identified in the contigs were different from the corresponding ones obtained by 454 pyrosequencing. The dTGD genes in YFC11 (YFC11_dTGD) were clustered with natural product biosynthetic genes, suggesting it might be one dTGD gene used for natural product glycosylation. The cluster harboring YFC11_dTGD and other obtained dTGD genes were different from the known dTGD genes, suggesting this cluster might represent one novel dTGD clade used for natural product glycosylation (**Figure 4**). Besides, five KS gene fragments of YFC1 clustered together in the phylogenetic tree (**Figure 2**). In the meanwhile, some AD gene fragments derived from same contigs were not clustered together in the phylogenetic tree, such as three AD gene fragments of YFC3_AD(2358…3031), YFC3_AD(5775…6470) and YFC3_AD(9202…9859) distributed in 3 different clades.

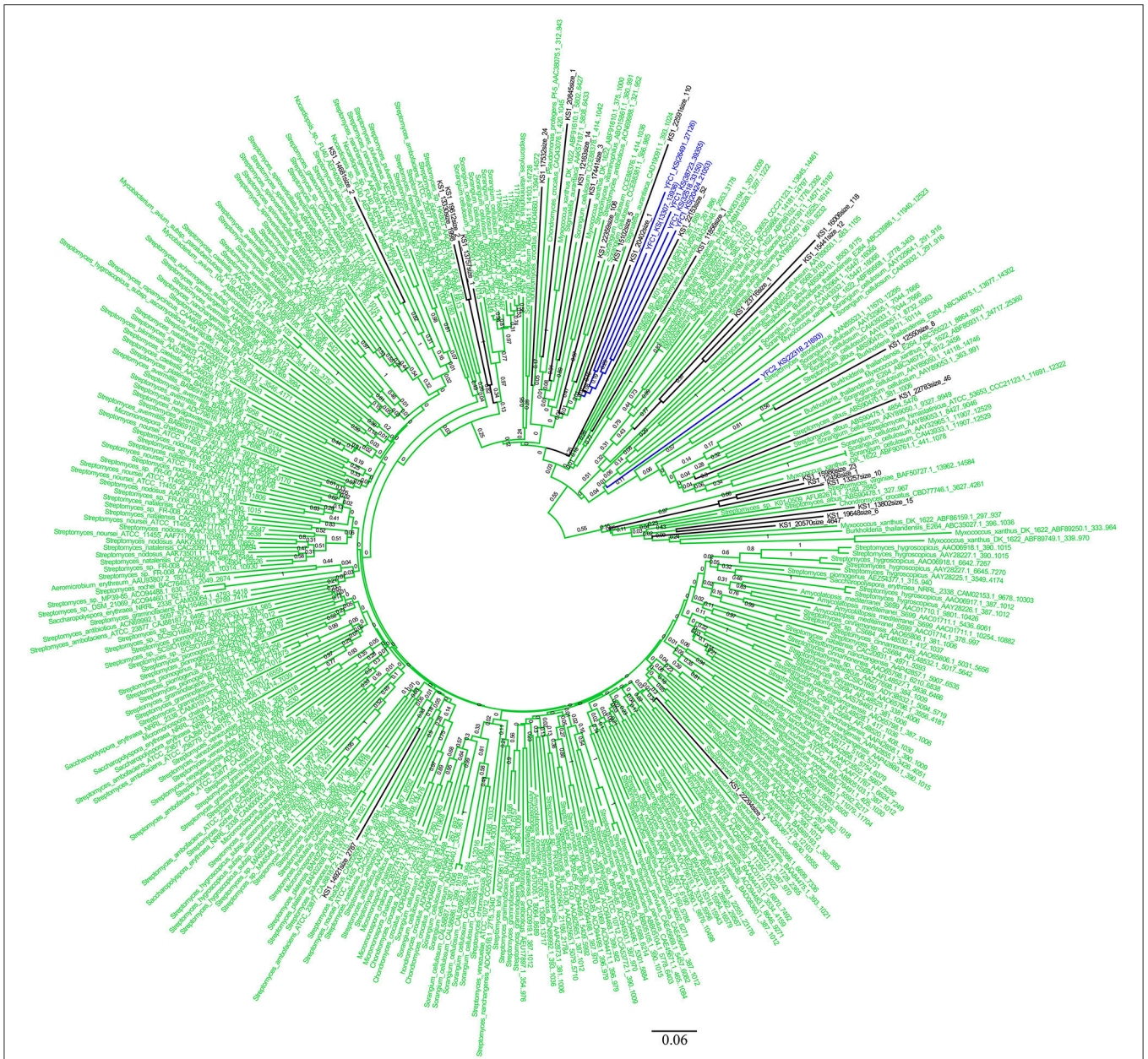


FIGURE 2 | Phylogenetic tree of KS representative sequences of each OTU obtained from Yellow Sea sediment and the reference functional characterized KS sequences. All the representative sequences of each OTU clustered at 90% identities were marked with black. All the sequences identified in fosmid contigs were marked with blue. The functionally characterized KS retrieved from ClusterMine360 database were marked with green.

Moreover, all the 6 identified KS gene fragments were clustered with canonical type I PKS KS domains, further showing that the contigs of YFC1 and YFC2 harbored canonical type I PKSs (Figure S2).

DISCUSSION

In this study, the 16S rRNA gene analysis indicated that *Proteobacteria* was the most dominant phylum in the Yellow Sea sediment, which was similar with bacteria diversity distributed in the East China Sea (south of the Yellow Sea) and the South China Sea (south of the East China Sea) where *Proteobacteria* was the most dominant phylum (Lu et al., 2011; Zhu et al., 2013). NRPS and PKS gene clusters are widely distributed in *Proteobacteria* (Wang et al., 2014). Moreover, 25% of the tested bacteria of the East China Sea possessed biological activities and some of them can produce novel natural product compounds (Lu et al., 2011), hinting natural product biosynthetic gene clusters in the marine sediment would be abundant and used for novel compounds biosynthesis. Only 11 OTUs (0.8% of

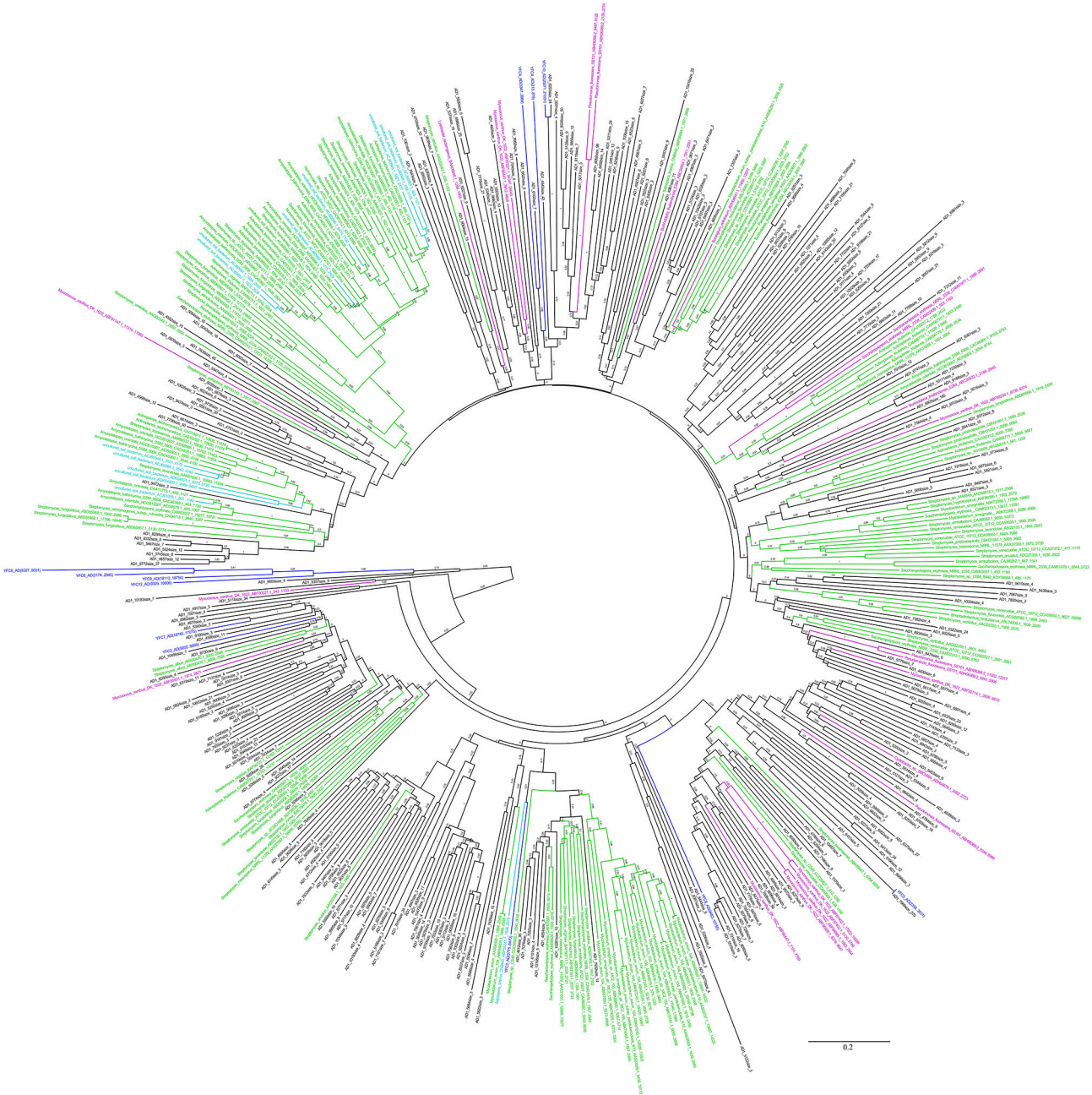


FIGURE 3 | Phylogenetic tree of AD representative sequences of each OTU obtained from Yellow Sea sediment and the reference functional characterized AD sequences. All the representative sequences of each OTU clustered at 90% identities were marked with black. All the sequences identified in fosmid contigs were marked with blue. The functionally characterized AD of *Actinobacteria* retrieved from ClusterMine360 database were marked with green. The functionally characterized AD of non-*Actinobacteria* bacteria retrieved from ClusterMine360 database were marked with pink. The functionally characterized AD of uncultured bacteria retrieved from ClusterMine360 database were marked with cyan.

all the OTUs) were assigned to *Actinobacteria*, but they were different from the isolated *Actinobacteria* strains from the same sample, suggesting that metagenomic method revealed some *Actinobacteria* bacteria which were different from the isolated ones (Xiong et al., 2015). However, five of the nine recovered fosmid clones were assigned to *Actinobacteria*, this may due to the fact that A3F/A7R and KSLF/R primers were designed with known *Actinobacteria* genes and it was easier to obtain *Actinobacteria* KS and AD genes than other bacteria KS and AD genes (Ginolhac et al., 2004; Ayuso-Sacido and Genilloud, 2005). Moreover, three of the sequenced nine fosmid clones were assigned to *Proteobacteria*, further showing

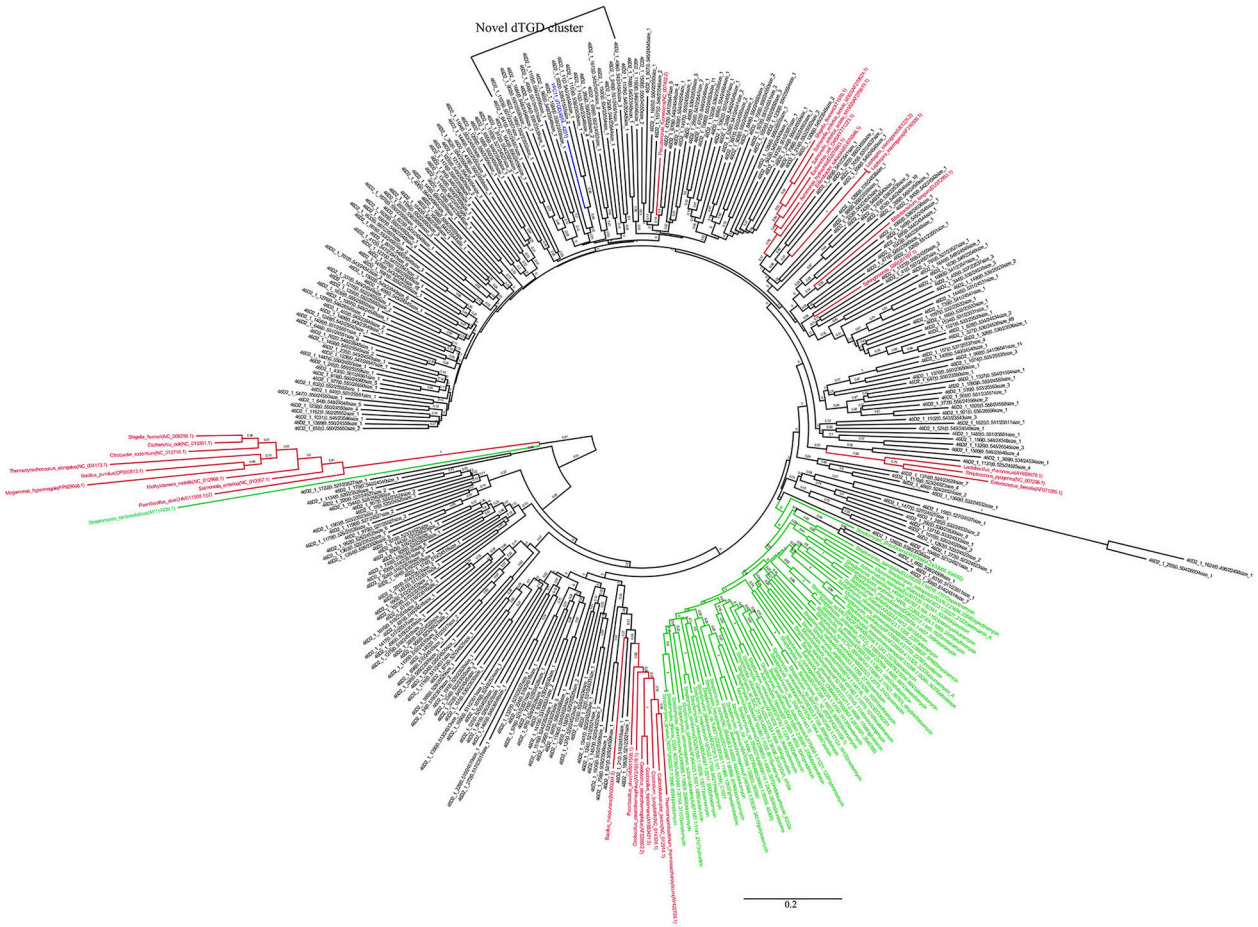


FIGURE 4 | Phylogenetic tree of dTGD representative sequences of each OTU obtained from Yellow Sea sediment and the reference functional characterized sequences. All the representative sequences of each OTU clustered at 80% identities were marked with black. All the sequences identified in fosmid contigs were marked with blue. The functionally characterized dTGD identified for natural product biosynthesis were marked with green. The reference sequence of dTGD identified for primary metabolism were marked with red.

Proteobacteria was the most abundant bacteria and was the potential natural product biosynthetic producers in the Yellow Sea sediment.

Computationally screening of the NCBI-NT database with the primers KSα F/R showed that all of the identified KSα genes were from phylum *Actinobacteria* (Reddy et al., 2012). Though different PCR conditions had been tried, no KSα gene fragments were amplified from the marine sediment, the reason might be that *Actinobacteria* bacteria were rare in Yellow Sea sediment and few PKS-II KSα domains distributed in this sample. More than half of the AD and KS genes in Uniprot database were from *Actinobateria* (Minowa et al., 2007; Reddy et al., 2012). Only part of the AD representative reads of each OTU were clustered with known *Actinobacteria*-derived reference genes and *Actinobacteria* represents only a small proportion of all the bacteria in the Yellow Sea sediment, suggesting some NRPS gene clusters which were different from the known ones distributed in the samples. The most abundant AD read was the same as one AD gene identified in the fosmid contigs, showing AD gene fragment diversity evaluated by PCR amplification might cover most of the abundant NRPS in the samples. Due to the fact that no corresponding KS reads were same with the KS genes identified in the fosmid contigs, hinting PCR amplification bias introduced by the primers or other factors may lead us to underestimate the true diversity of PKS-I in the marine sediment. Besides, more than one KS or AD domains were distributed in one natural product biosynthetic gene clusters, such as YFC3 contains 3 AD domains (**Figure 3**), suggesting metagenomic sequencing of the microbial community should be tried to help evaluate KS or AD diversity.

The recovered high dTGD diversity and most representative dTGD reads were not clustered with the reference dTGD genes, suggesting bacteria in the marine sediment might encode diverse natural products in the Yellow Sea sediment tend to be glycosylated and there would be some novel 6DOH-modified natural products (**Figure 4**). One dTGD gene clustered with predicted natural product biosynthetic genes identified in YFC11 showed only 67.2% amino acid identity with its nearest known dTGD genes, further implying that dTGD in the Yellow Sea had

TABLE 4 | Twelve fosmid contigs assembled from 9 fosmid clones harboring KS or AD gene fragments.

Fosmid contigs ID	Sequences used for assembly	Length (bp)	Natural product biosynthetic domains identified	Taxonomic binned by PhyloPythiaS	Similar known gene cluster predicted by antiSMASH
YFC1	Fosmid45M19	40,128	KS	Syntrophobacterales	Ajudazol_biosynthetic_gene_cluster (53% of genes show similarity)
YFC2	Fosmid60N12	33,750	AD and KS	Actinobacteria (class)	N/A
YFC3	Fosmid78P12	32,837	AD	Coriobacteriaceae	Gramicidin_biosynthetic_gene_cluster (16% of genes show similarity)
YFC4	Fosmid164P7	32,576	AD	Coriobacteriaceae	Jagaricin_biosynthetic_gene_cluster (13% of genes show similarity)
YFC5	Fosmid79B24R	5,655		Actinomycetales	N/A
YFC6	Fosmid65F11F	9,478	AD and KS	Actinobacteria (class)	Microsclerodermins_biosynthetic_gene_cluster (14% of genes show similarity)
YFC7	Fosmid65F11R	28,082	AD and KS	Actinobacteria (class)	Pellasoren_biosynthetic_gene_cluster (25% of genes show similarity)
YFC8	Fosmid60H14	27,713	AD	Bacteria	N/A
YFC9	Fosmid79B24F	27,636	AD and KS	Actinobacteria (class)	N/A
YFC10	Fosmid163L15	26,927	AD	Deltaproteobacteria	N/A
YFC11	Fosmid8H1R	24,345	AD	Proteobacteria	Sphingan_polysaccharide_biosynthetic_gene_cluster (13% of genes show similarity)
YFC12	Fosmid8H1F	13,106	AD	Oceanospirillales	N/A

N/A shows no identity with known natural product biosynthetic gene clusters.

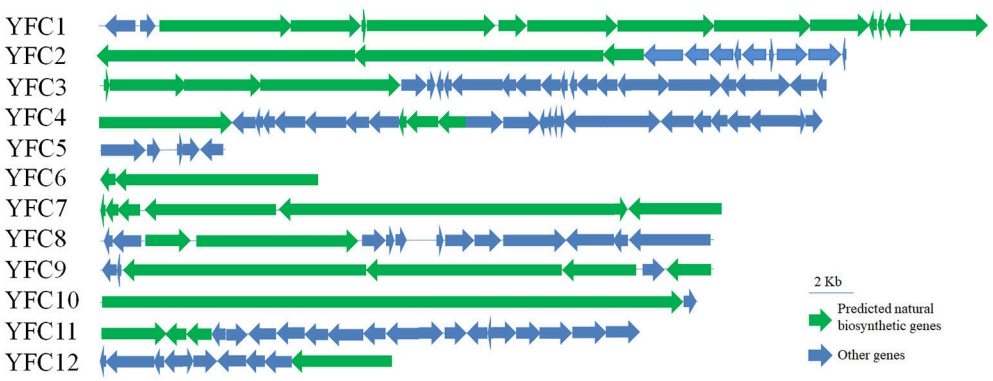


FIGURE 5 | Gene organization of the 12 fosmid contigs harboring natural product biosynthetic gene clusters were shown. The scale bar of 2 kb was marked.

potential to help synthesize novel natural products (**Figure 4**). Moreover, one glycosyl transferase (GT) gene was identified to be clustered with one natural product biosynthetic gene cluster, suggesting there might be some glycosylated natural products in the Yellow Sea sediment.

Most natural product biosynthetic genes in the recovered fosmid clones showed low identities with the known natural product biosynthetic genes clusters (**Table 4**), suggesting there might be abundant novel natural product biosynthetic gene clusters in marine sediment of the Yellow Sea. Though only two potential natural product gene clusters were completely recovered, these gene clusters give us insights into the potential natural products of the Yellow Sea sediment (Figure S3). Moreover, AT-less and canonical type I PKS KS gene fragments were identified in the metagenomics analysis (Figure S2), suggesting different kinds of PKs were available in the marine sediment. With the development of synthetic biology, rational design of natural product biosynthetic gene clusters with recovered natural product gene modules in this study and express them in appropriate chassis cells would be promising way to produce novel natural products (Menzella et al., 2005; Winter and Tang, 2012; Cobb et al., 2013; Montiel et al., 2015).

The bacteria community in the marine sediment of the Yellow Sea is highly diverse and the natural product biosynthetic genes are different from the ones identified in soils (Reddy et al., 2012), suggesting the ocean might be a rich source for novel natural products discovery. Moreover, our study suggests that the culture independent metagenomic method not only shows bacterial and natural product diversity of the Yellow Sea sediment, but also helps reveal novel PKS and NRPS gene clusters.

AUTHOR CONTRIBUTIONS

Conception and design: XY, ZZ; Data acquisition: YW, LZ, XY; Analysis and interpretation of the data: YW, LZ; Drafting of the article: YW, LZ; Critical revision of the manuscript: XY, ZZ.

ACKNOWLEDGMENTS

This work was financially supported by the programs from the Chinese Academy of Sciences (No. ZSYS-016), the National Basic Research Program of China (973 program, 2015CB755703) and the Chinese Academy of Sciences (No. KFJ-SW-STS-164). We thank Dr. Yong Wang and Dr. Zhiqiang Xiong for providing marine sediment samples and valuable scientific discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00295/full#supplementary-material

REFERENCES

Ayuso-Sacido, A., and Genilloud, O. (2005). New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb. Ecol.* 49, 10–24. doi: 10.1007/s00248-004-0249-6

Bérdy, J. (2005). Bioactive microbial metabolites. *J. Antibiot.* 58, 1–26. doi: 10.1038/ja.2005.1

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H., and Prinsep, M. R. (2013). Marine natural products. *Nat. Prod. Rep.* 30, 237–323. doi: 10.1039/C2NP20112G

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H., and Prinsep, M. R. (2014). Marine natural products. *Nat. Prod. Rep.* 31, 160–258. doi: 10.1039/c3np70117d

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H., and Prinsep, M. R. (2015). Marine natural products. *Nat. Prod. Rep.* 32, 116–211. doi: 10.1039/C4NP00144C

Brady, S. F., Chao, C. J., and Clardy, J. (2002). New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* 124, 9968–9969. doi: 10.1021/ja0268985

Brady, S. F., Simmons, L., Kim, J. H., and Schmidt, E. W. (2009). Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat. Prod. Rep.* 26, 1488–1503. doi: 10.1039/b817078a

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G. W. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* 9, 425–426. doi: 10.1038/nmeth.1990

Bruender, N. A., Thoden, J. B., and Holden, H. M. (2010). X-ray structure of KijD3, a key enzyme involved in the biosynthesis of D-kijanose. *Biochemistry* 49, 3517–3524. doi: 10.1021/bi100318v

Bull, A. T., and Stach, J. E. (2007). Marine actinobacteria: new opportunities for natural product search and discovery. *Trends Microbiol.* 15, 491–499. doi: 10.1016/j.tim.2007.10.004

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Chen, F., Lin, L., Wang, L., Tan, Y., Zhou, H., Wang, Y., et al. (2011). Distribution of dTDP-glucose-4, 6-dehydratase gene and diversity of potential glycosylated natural products in marine sediment-derived bacteria. *Appl. Microbiol. Biotechnol.* 90, 1347–1359. doi: 10.1007/s00253-011-3112-y

Cobb, R. E., Ning, J. C., and Zhao, H. (2013). DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* 41, 469-477. doi: 10.1007/s10295-013-1358-3

Conway, K. R., and Boddy, C. N. (2013). ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res.* 41, D402–D407. doi: 10.1093/nar/gks993

Courtois, S., Cappellano, C. M., Ball, M., Francou, F. X., Normand, P., Helynck, G., et al. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl. Environ. Microbiol.* 69, 49–55. doi: 10.1128/AEM.69.1.49-55.2003

Daniel, R. (2004). The soil metagenome–a rich resource for the discovery of novel natural products. *Curr. Opin. Biotechnol.* 15, 199–204. doi: 10.1016/j.copbio.2004.04.005

Demain, A. L., and Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *J. Antibiot.* 62, 5–16. doi: 10.1038/ja.2008.16

Dong, Q., Nelson, D. E., Toh, E., Diao, L., Gao, X., Fortenberry, J. D., et al. (2011). The microbial communities in male first catch urine are highly similar to those in paired urethral swab specimens. *PLoS ONE* 6:e19709. doi: 10.1371/journal.pone.0019709

Du, Y., Li, T., Wang, Y. G., and Xia, H. (2004). Identification and functional analysis of dTDP-glucose-4, 6-dehydratase gene and its linked gene cluster in an aminoglycoside antibiotics producer of *Streptomyces tenebrarius* H6. *Curr. Microbiol.* 49, 99–107. doi: 10.1007/s00284-004-4212-z

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Fenical, W., and Jensen, P. R. (2006). Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* 2, 666–673. doi: 10.1038/nchembio841

Fiedler, H.-P., Bruntner, C., Bull, A. T., Ward, A. C., Goodfellow, M., Potterat, O., et al. (2005). Marine actinomycetes as a source of novel secondary metabolites. *Antonie Van Leeuwenhoek* 87, 37–42. doi: 10.1007/s10482-004-6538-8

Geng, A., Zou, G., Yan, X., Wang, Q., Zhang, J., Liu, F., et al. (2012). Expression and characterization of a novel metagenome-derived cellulase Exo2b and its application to improve cellulase activity in *Trichoderma reesei*. *Appl. Microbiol. Biotechnol.* 96, 951–962. doi: 10.1007/s00253-012-3873-y

Ginolhac, A., Jarrin, C., Gillet, B., Robe, P., Pujic, P., Tuphile, K., et al. (2004). Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* 70, 5522–5527. doi: 10.1128/AEM.70.9.5522-5527.2004

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–R249. doi: 10.1016/S1074-5521(98)90108-9

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jensen, P. R., Mincer, T. J., Williams, P. G., and Fenical, W. (2005). Marine actinomycete diversity and natural product discovery. *Antonie Van Leeuwenhoek* 87, 43–48. doi: 10.1007/s10482-004-6540-1

Katz, L., and Baltz, R. H. (2016). Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* 43, 155–176. doi: 10.1007/s10295-015-1723-5

Kuczynski, J., Stombaugh, J., Walters, W. A., González, A., Caporaso, J. G., and Knight, R. (2011). Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr. Protoc. Bioinformatics* Chapter 10:Unit 10.7. doi: 10.1002/0471250953.bi1007s36

Lohman, J. R., Ma, M., Osipiuk, J., Nocek, B., Kim, Y., Chang, C., et al. (2015). Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12693–12698. doi: 10.1073/pnas.1515460112

Lu, X., Liu, X., Long, C., Wang, G., Gao, Y., Liu, J., et al. (2011). A preliminary study of the microbial resources and their biological activities of the East China Sea. *Evid. Based Compl. Alt.* 2011:806485. doi: 10.1155/2011/806485

Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., et al. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* 23, 1171–1176. doi: 10.1038/nbt1128

Metsä-Ketelä, M., Salo, V., Halo, L., Hautala, A., Hakala, J., Mäntsälä, P., et al. (1999). An efficient approach for screening minimal PKS genes from Streptomyces. *FEMS Microbiol. Lett.* 180, 1–6. doi: 10.1111/j.1574-6968.1999.tb08770.x

Minowa, Y., Araki, M., and Kanehisa, M. (2007). Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* 368, 1500–1517. doi: 10.1016/j.jmb.2007.02.099

Molinski, T. F., Dalisay, D. S., Lievens, S. L., and Saludes, J. P. (2009). Drug development from marine natural products. *Nat. Rev. Drug Discov.* 8, 69–85. doi: 10.1038/nrd2487

Montiel, D., Kang, H.-S., Chang, F.-Y., Charlop-Powers, Z., and Brady, S. F. (2015). Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8953–8958. doi: 10.1073/pnas.1507606112

Newman, D. J., and Cragg, G. M. (2012). Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* 75, 311–335. doi: 10.1021/np200906s

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., et al. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* 35:e130. doi: 10.1093/nar/gkm760

Patil, K. R., Roune, L., and McHardy, A. C. (2012). The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS ONE* 7:e38581. doi: 10.1371/journal.pone.0038581

Piel, J. (2011). Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol.* 65, 431–453. doi: 10.1146/annurev-micro-090110-102805

Rappé, M. S., and Giovannoni, S. J. (2003). The uncultured microbial majority. *Annu. Rev. Microbiol.* 57, 369–394. doi: 10.1146/annurev.micro.57.030502.090759

Reddy, B. V., Kallifidas, D., Kim, J. H., Charlop-Powers, Z., Feng, Z., and Brady, S. F. (2012). Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl. Environ. Microbiol.* 78, 3744–3752. doi: 10.1128/AEM.00102-12

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Thibodeaux, C. J., Melançon, C. E., and Liu, H. W. (2007). Unusual sugar biosynthesis and natural product glycodiversification. *Nature* 446, 1008–1016. doi: 10.1038/nature05814

Torsvik, V., Sørheim, R., and Goksøyr, J. (1996). Total bacterial diversity in soil and sediment communities—a review. *J. Ind. Microbiol. Biotechnol.* 17, 170–178. doi: 10.1007/BF01574690

Tulp, M., and Bohlin, L. (2005). Rediscovery of known natural compounds: nuisance or goldmine? *Bioorg. Med. Chem.* 13, 5274–5282. doi: 10.1016/j.bmc.2005.05.067

Walsh, C. T. (2007). The chemical versatility of natural-product assembly lines. *Acc. Chem. Res.* 41, 4–10. doi: 10.1021/ar7000414

Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9259–9264. doi: 10.1073/pnas.1401734111

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, 237–243. doi: 10.1093/nar/gkv437

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583. doi: 10.1073/pnas.95.12.6578

Wilson, M. C., and Piel, J. (2013). Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* 20, 636–647. doi: 10.1016/j.chembiol.2013.04.011

Winter, J. M., and Tang, Y. (2012). Synthetic biological approaches to natural product biosynthesis. *Curr. Opin. Biotechnol.* 23, 736–743. doi: 10.1016/j.copbio.2011.12.016

Woodhouse, J. N., Fan, L., Brown, M. V., Thomas, T., and Neilan, B. A. (2013). Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. *ISME J.* 7, 1842–1851. doi: 10.1038/ismej.2013.65

Xiong, Z.-Q., Liu, Q.-X., Pan, Z.-L., Zhao, N., Feng, Z.-X., and Wang, Y. (2015). Diversity and bioprospecting of culturable actinomycetes from marine sediment of the Yellow Sea, China. *Arch. Microbiol.* 197, 299–309. doi: 10.1007/s00203-014-1059-y

Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–322.

Zhu, D., Tanabe, S.-H., Yang, C., Zhang, W., and Sun, J. (2013). Bacterial community composition of South China sea sediments through pyrosequencing-based analysis of 16S rRNA genes. *PLoS ONE* 8:e78501. doi: 10.1371/journal.pone.0078501

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wei, Zhang, Zhou and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 July 2017
doi: 10.3389/fmicb.2017.01304

Genomic Reconstruction of Carbohydrate Utilization Capacities in Microbial-Mat Derived Consortia

Semen A. Leyn[1,2], Yukari Maezato[3], Margaret F. Romine[3] and Dmitry A. Rodionov[1,2]**

[1] Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, CA, United States, [2] A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia, [3] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, United States

OPEN ACCESS

Edited by:
Florence Abram, NUI Galway, Ireland

Reviewed by:
Boyang Ji, Chalmers University of Technology, Sweden
Timothy Casselli, University of North Dakota, United States
Biswarup Sen, Tianjin University, China
Daniela Medeot, National University of Río Cuarto, Argentina

***Correspondence:**
Dmitry A. Rodionov
rodionov@sbpdiscovery.org
Semen A. Leyn
semen.leyn@gmail.com

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *23 March 2017*
Accepted: *28 June 2017*
Published: *13 July 2017*

Citation:
Leyn SA, Maezato Y, Romine MF and Rodionov DA (2017) Genomic Reconstruction of Carbohydrate Utilization Capacities in Microbial-Mat Derived Consortia. Front. Microbiol. 8:1304. doi: 10.3389/fmicb.2017.01304

Two nearly identical unicyanobacterial consortia (UCC) were previously isolated from benthic microbial mats that occur in a heliothermal saline lake in northern Washington State. Carbohydrates are a primary source of carbon and energy for most heterotrophic bacteria. Since CO_2 is the only carbon source provided, the cyanobacterium must provide a source of carbon to the heterotrophs. Available genomic sequences for all members of the UCC provide opportunity to investigate the metabolic routes of carbon transfer between autotroph and heterotrophs. Here, we applied a subsystem-based comparative genomics approach to reconstruct carbohydrate utilization pathways and identify glycohydrolytic enzymes, carbohydrate transporters and pathway-specific transcriptional regulators in 17 heterotrophic members of the UCC. The reconstructed metabolic pathways include 800 genes, near a one-fourth of which encode enzymes, transporters and regulators with newly assigned metabolic functions resulting in discovery of novel functional variants of carbohydrate utilization pathways. The *in silico* analysis revealed the utilization capabilities for 40 carbohydrates and their derivatives. Two *Halomonas* species demonstrated the largest number of sugar catabolic pathways. Trehalose, sucrose, maltose, glucose, and beta-glucosides are the most commonly utilized saccharides in this community. Reconstructed regulons for global regulators HexR and CceR include central carbohydrate metabolism genes in the members of Gammaproteobacteria and Alphaproteobacteria, respectively. Genomics analyses were supplemented by experimental characterization of metabolic phenotypes in four isolates derived from the consortia. Measurements of isolate growth on the defined medium supplied with individual carbohydrates confirmed most of the predicted catabolic phenotypes. Not all consortia members use carbohydrates and only a few use complex polysaccharides suggesting a hierarchical carbon flow from cyanobacteria to each heterotroph. In summary, the genomics-based identification of carbohydrate utilization capabilities provides a basis for future experimental studies of carbon flow in UCC.

Keywords: microbial community, carbohydrate utilization, comparative genomics, metabolic reconstruction, transcriptional regulation

INTRODUCTION

Knowledge of molecular interactions that occur between bacteria in microbial communities is critical for understanding of both environmental and human-associated ecological niche formation. Because natural microbial communities are typically complex in terms of species diversity and function, simplified models are necessary to study the basis of their behavior. Recently two nearly identical unicyanobacterial consortia (UCC) were derived from a photosynthetic microbial mat from Hot Lake, Washington (Cole et al., 2014) and their member genomes assembled from metagenome and isolate DNA sequence (Nelson et al., 2015). Since the single cyanobacterial member of each consortium is also the only autotroph, their heterotrophic cohorts depend on the cyanobactium for organic carbon when CO_2 is the only external source of carbon provided. Previous experimental data suggest the presence of numerous metabolic interactions between the heterotrophs and the cyanobacterium, which makes these consortia excellent models for defining the metabolic interactions potential in the community (Cole, 1982; Carpenter and Foster, 2002; Seymour et al., 2010; Beliaev et al., 2014).

Cyanobacteria excrete numerous organic compounds including fermentation products (Stal and Moezelaar, 1997), osmolytes (Hagemann, 2011), polysaccharides, proteins, nucleic acids, and lipids (Decho, 1990; Chiovitti et al., 2003; Flemming and Wingender, 2010). The UCC cyanobacteria form sheaths made up of exopolysaccharides (EPSs), high-molecular-mass heteropolymers composed of various sugars and their derivatives. Electron microscopy of the UCC show that heterotrophic bacteria are attached to these sheaths, which suggests that their constituents could be used as carbon source for the heterotrophic members (Cole et al., 2014). Cyanobacterial EPSs are characterized by their complex structure with high diversity of monosaccharides found as their building blocks. Up to 75% of EPSs are heteropolysaccharides composed of least six different types of monosaccharides (Pereira et al., 2009). The most common carbohydrates found in EPSs of Cyanobacteria are glucose, galactose, mannose, fructose, fucose, rhamnose, xylose, arabinose, as well as glucuronic and galacturonic acids (Pereira et al., 2009). The composition of EPSs produced by the UCC cyanobacteria is currently unknown, however, the previous metabolic analysis of UCC composition identified an abundance of putative osmolytes such as glycerol, gluconate, glucosylglycerol, glucosylglycerate, sucrose, and trehalose (Cole et al., 2014).

A highly diverse array of metabolic pathways for utilization of carbohydrates has been previously described for heterotrophic bacteria (Yang et al., 2006; Gu et al., 2010; Leyn et al., 2012; Rodionova et al., 2012, 2013a,b; Zhang et al., 2012). The sugar utilization networks in bacteria are represented by a large number of species-to-species variations in carbohydrate hydrolases, uptake transporters, transcriptional regulators and enzymes catalyzing the catabolism of monosaccharides. A subsystems-based comparative genomics approach allows us to substantially enhance the accuracy of genomic annotations, to infer functions of previously unknown gene families and to describe metabolic pathways and associated transcriptional networks in several diverse bacterial taxa (Osterman and Overbeek, 2003; Overbeek et al., 2005; Rodionov, 2007). The subsystems approach was highly efficient for prediction of novel sugar catabolic pathways that are often comprised of co-localized and co-regulated genes. The applicability and efficacy of this *in silico* approach was shown by our previous reconstructions of sugar utilization networks in *Bacteroides*, *Shewanella*, and *Thermotoga* genera (Rodionov et al., 2010, 2013; Ravcheev et al., 2013) and by similar works of others (Warda et al., 2016).

Recently, we applied the integrated subsystems-based approach to reconstruct vitamin cofactor biosynthesis pathways and associated transporter capabilities in the 19 organisms that comprise the two model UCC derived from microbial mat from Hot Lake in Washington (Romine et al., 2017) and to predict cofactor exchange among consortial members. In this work, we focused on identification of carbohydrate utilization abilities of the heterotrophic members of these consortia so that we could predict that types of carbohydrates exchanged with the cyanobacterium. Using the bioinformatics approach we systematically mapped peripheral carbohydrate utilization pathways and the central carbohydrate metabolism (CCM) in a group of 17 UCC heterotrophs with sequenced genomes. The reconstructed carbohydrate catabolic network allowed us to annotate a large number of catabolic enzymes, and to infer associated catabolic pathways. In particularly, we identified novel pathway variants such as the predicted pathway for mannoheptulose utilization. In addition, we identified potential transporters and regulators involved in the uptake and sensing of the utilized carbohydrates. The obtained carbohydrate catabolic phenotypes were assessed experimentally using Api50 tests and/or growth of selected UCC isolates on defined media with individual carbohydrates as carbon and energy sources. The combined *in silico* analyses and *in vivo* experiments revealed a large and diverse set of carbohydrate utilization pathways unevenly distributed across the majority of the heterotrophic UCC organisms.

MATERIALS AND METHODS

Bioinformatics Analysis of UCC Genomes

Assembled genomes of 19 UCC members were obtained from the U.S. Department of Energy, Joint Genome Institute (DOE-JGI). The annotated genomic sequences were downloaded from the Integrated Microbial Genome (IMG) expert review database (Chen et al., 2016). In addition, the genomic assemblies can be accessed in the European Nucleotide Archive[1]. UCC is composed of a combination of the species-resolved metagenome bins and isolate genome sequences for organisms that were previously cultivated axenically (Nelson et al., 2015) (**Table 1**). Completeness of genomic content for most of the analyzed metagenomic bins, as previously estimated by presence/absence of 100 conserved single-copy genes, was at least 98%, with the exception of one member, bin09, whose estimated coverage

[1]http://www.ebi.ac.uk/ena

TABLE 1 | Genomic properties and overview of carbohydrate utilization capabilities of analyzed unicyanobacterial consortia (UCC) organisms.

UCC assembly/isolate genome	Alias	Total genes	GHs[1]	CU genes[2]	CU pathways[3]
Bacteroidetes					
Bacteroidetes bin01	Bin01	2691	34 (8)	32 (4)	4
Algoriphagus marincola HL-49	HL-49	3820	59 (2)	73 (11)	9
Gammaproteobacteria					
Halomonas sp. HL-48	HL-48	3463	12 (0)	105 (19)	19
Halomonas sp. HL-93	HL-93	3928	11 (0)	150 (28)	25
Aliidiomarina calidilacus HL-53	HL-53	2602	14 (1)	7 (3)	2
Marinobacter excellens HL-55	HL-55	3717	14 (1)	4 (0)	0
Marinobacter sp. HL-58	HL-58	3948	28 (1)	45 (3)	6
Alphaproteobacteria					
Roseibaca calidilacus HL-91	HL-91	3313	22 (1)	46 (13)	8
Rhodobacteriaceae bin07	Bin07	3338	16 (0)	8 (6)	1
Rhodobacteriaceae bin08	Bin08	3542	34 (0)	164 (41)	18
Rhodobacteriaceae bin09	Bin09	3817	27 (0)	49 (15)	11
Rhodobacteriaceae bin12	Bin12	3644	15 (0)	7 (0)	0
Rhodobacteriaceae bin18	Bin18	3357	24 (0)	75 (19)	12
Oceanicaulis bin04	Bin04	2667	9 (0)	0 (0)	0
Salinivirga fredricksonii HL-109	HL-109	3832	21 (0)	16 (5)	4
Erythrobacter sp. HL-111	HL-111	2862	20 (0)	2 (0)	0
Porphyrobacter sp. HL-46	HL-46	3057	14 (0)	15 (4)	4
Cyanobacteria					
Phormidium OSCR	OSCR	4426	30 (0)	n /a	n /a
Phormidesmis priestleyi ANA	ANA	4911	37 (0)	n /a	n /a

[1]*Number of predicted glycosyl hydrolases (GHs) per genome. Number of predicted extracellular GHs is given in parenthesis. The details on prediction of GHs and bioinformatics assignment of their cellular localization are given in Supplementary Table S1.* [2]*Number of genes involved in the reconstructed carbohydrate utilization (CU) pathways including enzymes, uptake transporters and transcriptional regulators. Number of genes with newly assigned functions is given in parenthesis. The details of all functional assignments are provided in Supplementary Table S2.* [3]*Number of reconstructed carbohydrate utilization pathways reflects an estimate of a number of different carbohydrates utilized through the reconstructed pathways. Incomplete CU pathways with missing carbohydrate transporters are not counted. Detailed distribution of CU pathways is given in* ***Figure 1****.*

is 88% (Nelson et al., 2015). In addition to the cyanobacterial members, UCC contain 17 heterotrophic members including 10 Alphaproteobacteria, five Gammaproteobacteria, and two species from the *Bacteroidetes* phylum.

Glycoside hydrolases (GHs) were predicted by analyzing the deduced proteome sequence from all 19 UCC organisms on the dbCAN server (Yin et al., 2012). Cellular localizations of GHs were predicted as previously described (Romine, 2011). Briefly, genomes were assessed for the presence of secretion systems to identify those capable of secreting proteins and then deduced GH sequences analyzed with the following web-tools: SignalP with sensitive parameters (0.5 SignalP-noTM/0.42 SignalP-TM) (Petersen et al., 2011); LipoP (Juncker et al., 2003); TatP (Bendtsen et al., 2005b); SecretomeP (Bendtsen et al., 2005a); TMHMM (Krogh et al., 2001); PRED-TMBB2 (Tsirigos et al., 2016); PSORTb (Yu et al., 2010); and SOSUIGramN (Imai et al., 2008). Predictions of localization were additionally improved based on the presence of location-informative domains and the assumption that orthologous GHs should have same subcellular localization. Identification of orthologs in closely related genomes was performed using IMG. Functional annotations of predicted GHs were manually curated with input from the UniProt database (Boutet et al., 2016) and the RAST annotation server (Overbeek et al., 2014).

Genomic Reconstruction of Metabolic Pathways and Regulons

The UCC genomes were previously annotated via the following two pipelines: (i) the DOE-JGI Microbial Genome Annotation Pipeline (Huntemann et al., 2015), and (ii) the RAST server (Overbeek et al., 2014). First, we obtained the set of genes that are potentially involved in the carbohydrate metabolism in UCC genomes by filtering the RAST-based gene annotations and subsystem assignments, and by adding the predicted sets of functionally annotated GHs. The initial gene set was further expanded by potential carbohydrate metabolism genes according to their KEGG Orthology (KO) annotations (Kanehisa et al., 2016). Finally, we added genes from specific protein families associated with carbohydrate metabolism in the Pfam database according to their Gene Ontology terms (Finn et al., 2016). The expanded set of genes potentially involved in carbohydrate metabolism was further analyzed using manual inspection and the genome context techniques. We used the following three genome context techniques to functionally link a set of genes to a single pathway: (i) clustering of genes on the chromosome (operons), (ii) co-regulation of genes by a common regulator (regulons), and (iii) co-occurrence of genes in a set of related genomes (Overbeek et al., 2007; Rodionov, 2007; Haft, 2015).

Reconstruction of carbohydrate utilization pathways in 17 heterotrophic UCC members was performed using the subsystem-based comparative genomics approach combined with genomic reconstruction of carbohydrate-specific transcription factor (TF) regulons and identification of candidate carbohydrate-specific transporters as previously described (Rodionov et al., 2010, 2013; Ravcheev et al., 2013). Typical metabolic reconstruction workflow included: (i) analysis of gene neighborhood conservation across closely related microbial genomes using the Gene Ortholog Neighborhood tool in IMG; (ii) BLAST searches for functionally characterized orthologs in SwissProt/UniProt; (iii) reconstruction of local TF regulons to identify additional co-regulated gene loci; (iv) metabolic subsystem analysis for closely related genomes in the SEED database (Overbeek et al., 2014). Many of the initially identified gene candidates whose functional roles were deemed unrelated to carbohydrate utilization (e.g., involved in biosynthetic pathways) were rejected. The refined functional annotations for genes involved in the reconstructed pathways are provided in Supplementary Table S2.

For reconstruction of novel TF regulons, we used the bioinformatics technique based on identification and comparative analysis of candidate TF-binding sites in closely related genomes (Rodionov, 2007) and implemented in the RegPredict software (Novichkov et al., 2010). This approach includes the following steps: (i) search for orthologous groups of the studied TFs in other reference genomes; (ii) selection of conserved orthologous gene loci containing the studied TFs; (iii) prediction of candidate TF binding motifs with palindromic or tandem repeat structures; (iv) construction of positional weight matrices (PWMs) for identified DNA motif and its application for identification of additional sites and regulon members in each TF-containing genome. Scores of candidate sites were calculated as the sum of positional nucleotide weights. The threshold for site scores was defined as the lowest score observed in the training set. The reconstructed regulons included the conserved regulatory interactions in at least two other genomes with TF binding sites above threshold. The CceR and HexR regulons were analyzed using the previously constructed PWMs from the RegPrecise database (Novichkov et al., 2013). Weblogo package (Crooks et al., 2004) was used to build sequence logos for the derived DNA-binding motifs. The reconstructed CceR and HexR regulons are described in Supplementary Table S3. Other identified sugar catabolic regulons and their candidate TF binding sites are provided in Supplementary Table S2.

Phenotypic Analysis of Heterotrophic UCC Isolates

An ultimate validation of the genomics-based metabolic reconstructions was attained by experimental testing of growth phenotypes. Four UCC isolates including the *Halomonas* sp. HL-48 and HL-93 strains, *Roseibaca calidilacus* HL-91 and *Marinobacter* sp. HL-58 were tested for their ability to grow on a panel of various carbon sources as a sole carbon and energy source. Cells were grown in Hot Lake Heterotroph (HLH) medium (Cole et al., 2014), containing 10 mM TES pH 8.0, 400 mM $MgSO_4$, 80 mM Na_2SO_4, 20 mM KCl, 1 mM $NaHCO_3$, 5 mM NH_4Cl, and supplemented with 5 mM of a specific carbohydrate as a sole carbon source. For initial validation of genomic prediction of carbohydrate utilization we used bioMerieux^{TM} Api^{TM}50 CH carbohydrate fermentation strips. The Api^{TM}50 CH strip contains 49 wells with different carbohydrates in each well and one negative control well (no carbon source). The HLH media for starter cultures was supplemented with 5 mM glycerol (for HL-91) or 5 mM sucrose (for HL-48, HL-93, and HL-58). The mid-log phase grown cells were further washed three times with HLH medium without any carbon source to eliminate carry over of carbon substrates from starter media prior to inoculation and washed cells were used as inoculums in Api50 CH strip. Utilization of carbon sources were indicated by the growth on the well. For each strain, two independent repetitions were performed. The incubation time for Api50 CH strip measurements was 3 days (for HL-48, HL-93, and HL-58) and 7 days (for HL-91). Growth phenotype of Api50 CH strip results were further validated by growth measurements using selected carbohydrates. An optical density (OD_{600}) was measured to monitor cell growth during 60–100 h using a plate reader instrument Norden Lab Professional-Bioscreen. 250 μL culture volumes in the 100 well Bioscreen plate, and each growth experiments were performed in 10 replicates.

RESULTS AND DISCUSSION

Glycoside Hydrolases

To estimate sugar degradation capabilities of UCC members, we identified sets of carbohydrate active glycosyl hydrolases (GHs) that are involved in breakdown of oligosaccharides (and polysaccharides) into monosaccharides. Overall, 441 proteins containing at least one GH domain were found unevenly distributed in the studied UCC genomes (**Table 1**). The majority of the identified GHs have a predicted cellular localization either in the cytoplasm (for 217 GHs) or in the periplasm (for 178 GHs), with an additional 29 GHs found in the inner membrane (Supplementary Table S1).

The type two generalized protein secretion system was found in only four heterotrophs; *Aliidiomarina calidilacus* HL-53, *Marinobacter* sp. HL-58, *Marinobacter excellens* HL-55, and *Oceanicaulis* bin04 but are predicted to only secrete a single GH per genome, except bin04 which has no predicted extracellular GH (Supplementary Table S1). HL-53 also encodes a single outer membrane GH. The type IX protein secretion system was found in *Bacteroidetes* bin01 and *Algoriphagus marincola* HL-49 and is predicted to be responsible for secreting eight and two GHs, respectively. *R. calidilacus* HL-91 and *Rhodobacteriaceae* bin12 and bin18 encode a single autotransporter that secretes an orthologous cellulase. Collectively these results suggest that at least eight heterotrophs are able to degrade polysaccharides and that the remaining heterotrophs may rely on them for production of mono- and disaccharides that can be transported into the cell for further degradation or that they utilize other forms of carbon to support their carbon and energy needs.

Using similarity searches against the Uniprot database and metabolic reconstructions via the comparative genomics techniques (see below) we analyzed potential function of 374

GHs identified in the heterotrophic UCC organisms. As result, we tentatively assigned substrate specificity and metabolic pathway to 338 GHs (Supplementary Table S1). Of these, 125 of the GH enzymes are membrane-bound or periplasmic transglycosylases that are involved in peptidoglycan metabolism. An additional 53 GHs are putatively involved in biosynthesis of trehalose, maltose, or glycogen. The remaining 160 GHs with assigned functional roles are potentially involved in carbohydrate utilization pathways, at that 112 and 29 of them are potentially located in the cytoplasm and the periplasm, respectively. The remaining functionally annotated GHs with catabolic functions are distributed between the periplasm, the inner and outer membranes and the extracellular milieu. Among 12 extracellular GHs there are six β- and five α-glucosidases involved the glucan and maltodextrin utilization, as well as a probable chitinase. Half of these secreted GHs were from *Bacteroidetes* bin01, suggesting it is an important UCC member contributing to initial breakdown of polysaccharides.

Peripheral Carbohydrate Utilization

We applied subsystem-based comparative genomics approach to reconstruct peripheral carbohydrate utilization pathways in the 17 heterotrophic UCC members. Our analysis revealed highly diverse capabilities of UCC organisms to utilize carbohydrates and their derivatives (**Figure 1**). Overall, we identified pathways for utilization of six hexoses (glucose, galactose, fructose, mannose, fucose, rhamnose), two amino sugars (*N*-acetylgalactosamine and *N*-acetylglucosamine), two pentoses (arabinose and xylose), 10 sugar acids and diacids (see below) and six sugar alcohols including inositol, arabinitol, mannitol, sorbitol, erythritol, and glycerol. In addition to monosaccharides, we reconstructed catabolic pathways for several oligosaccharides including α- and β-glucosides, α- and β-galactosides, maltose, sucrose, and trehalose. Finally, we predicted a novel putative pathway for utilization of mannoheptulose (a heptose).

Most heterotrophic UCC members were predicted to catabolize at least one carbohydrate (**Table 1**). *M. excellens* HL-55, *Erythrobacter* sp. HL-111, *Rhodobacteriaceae* bin12, and *Oceanicaulis* bin04 lack any complete carbohydrate utilization pathway although some of them possess some catabolic genes in the absence of predicted carbohydrate-specific transporters. UCC members with a limited number of carbohydrate utilization pathways include *A. calidilacus* HL-53 (predicted to utilize glucose and β-glucosides) and *Rhodobacteriaceae* bin07 (only has sorbitol utilization pathway). In contrast, two *Halomonas* species (HL-48 and HL-93) and *Rhodobacteriaceae* bin08 have the largest numbers of identified carbohydrate utilization genes and pathways. For instance, HL-93 has the predicted capabilities to utilize 25 carbohydrates and their derivatives, whereas HL-48 and bin08 have 18–19 individual pathways.

The peripheral carbohydrate utilization pathways include between eight proteins in HL-111 to up to 150 proteins in HL-93 (**Table 1**). The complete list of 798 proteins involved in the reconstructed pathways across 16 organisms along with their deduced functional annotations is provided in Supplementary Table S2. Nearly half of these proteins constitute metabolic enzymes including nearly 100 of GHs with assigned catabolic pathway. The set of 285 annotated proteins are components of almost 100 carbohydrate transport systems. The obtained metabolic reconstruction includes 88 DNA-binding TFs that presumably control the reconstructed carbohydrate catabolic pathway genes. Using the metabolic reconstruction approach, we predicted specific functional assignments for 171 proteins, whose functions were previously unknown or annotated only at the level of general class (**Table 1** and Supplementary Table S2). Below, we describe the key novel aspects of the reconstructed catabolic pathways in UCC organisms in more details.

L-Arabinose and L-Arabinonate Utilization

In both studied *Halomonas* species we found a new gene locus potentially involved in the utilization of L-arabinose and L-arabinonate (**Figure 2**). It encodes proteins that are orthologous to (i) the ABC-type arabinose uptake transporter system AraFHG and (ii) the AraA, AraC, and AraE enzymes from the oxidative arabinose degradation pathway in *Azospirillum brasiliense* (Watanabe et al., 2006a,b). Based on genome context and distant homology analysis we have identified candidates for the missing 2-keto-3-deoxy-L-arabonate dehydratase (AraD) and a second isozyme of arabinose-1-dehydrogenase (AraY). A member of aldose 1-epimerase family encoded in the *ara* gene cluster was previously assigned the functional role L-arabinose mutarotase (AraM), which interconverts alpha and beta anomers of L-arabinose. The *ara* gene locus in *Halomonas* encodes two novel TFs from the LysR and GntR families (named AraR and AraR2, respectively). Reconstruction of their cognate regulons using these and other *Halomonas* genomes has revealed two different DNA motifs (Supplementary Table S2). AraR presumably controls the divergently transcribed *araMFGHCY* and *araR* genes, whereas AraR2 is predicted to co-regulate the *araD-araT-araR2-araA* operon, the *araE* gene and several other genes encoding a novel transporter from the tripartite ATP-independent periplasmic (TRAP) transporter family and a hypothetical lactonase, which was assigned the missing arabinolactonase function (named AraB). Nearly all known TRAP-family transporters have specificities to organic acids (Vetting et al., 2015), suggesting the novel AraR2-regulated transporter is specific to L-arabinonate, an intermediate of the L-arabinose catabolism. In summary, the *ara* gene locus represents interconnection of two regulatory systems controlling a shared catabolic pathway for utilization of L-arabinose and L-arabinonate.

L-Fucose, L-Fuconate, and L-Galactonate Utilization

The oxidative pathway for utilization of L-fucose, where L-fuconate is an intermediate, was shown in *Xanthomonas campestris* (Yew et al., 2006). We observed loci containing genes from this pathway in both *Halomonas* species (HL-48 and HL-93), *A. marincola* HL-49 and *Rhodobacteriaceae* bin08 (**Figure 3**). HL-48 possesses genes encoding the last two steps of this pathway, COG0179 and COG1028. However, the absence of genes for the first two steps of the fucose pathway suggests that HL-48 can only utilize the L-fuconate intermediate. In contrast, HL-93 and bin08 have the complete L-fucose utilization

Utilized carbohydrates / **UCC genomes**	Bact.		γ-Proteobacteria					α-Proteobacteria									
	bin01	HL-49	HL-48	HL-93	HL-53	HL-55	HL-58	HL-91	bin07	bin08	bin09	bin12	bin18	bin04	HL-109	HL-111	HL-46	Total
Hexoses																		
D-glucose	+		+	+	+		+	+					+				+	8
D-mannose				+														1
D-fructose		+	+	+			+			+								5
D-galactose		+[12]	+[1]	+[1]													+	3
L-fucose		+		+						+								3
L-rhamnose										+[3]								1
N-acetylgalactosamine										+[1]								1
N-acetylglucosamine		+						+*			+	+[12]	+[2]					3
Galactosides (α, β)										+	+[1]		+[1]				+	4
α-D-glucosides								+		+	+		+					4
β-D-glucosides	+	+			+			+		+	+		+				+	8
Maltose/maltodextrin	+	+	+	+		+[2]	+	+		+	+		+		+			10
Sucrose	+	+	+	+		+[2]	+	+		+	+		+		+			10
Trehalose			+	+			+	+		+	+		+		+			8
Pentoses																		
L-arabinose			+	+														2
D-xylose		+	+[3]	+[3]						+	+							5
Heptoses																		
Mannoheptulose										+								1
Sugar acids																		
L-arabinonate			+	+														2
D-galactarate			+	+														2
D-glucarate			+	+														2
D-glucuronate		+		+				+[2]		+			+[1]					4
D-galacturonate		+		+														2
L-gulonate			+[1]	+[1]														2
D-galactonate															+			1
L-galactonate			+	+														2
L-fuconate			+	+														2
D-gluconate			+	+														2
Sugar alcohols																		
Inositol				+[1]						+			+[1]					3
D-arabinitol			+	+						+	+		+					5
D-mannitol			+	+						+	+		+					5
D-sorbitol			+	+					+	+	+		+					6
D-erythritol				+														1
Glycerol			+	+			+	+	+[2]	+	+[2]	+[2]	+[2]		+[2]	+[2]	+[2]	5
Total	4	9	19	25	2	0	6	8	1	18	11	0	12	0	4	0	4	

FIGURE 1 | Predicted carbohydrates utilization capability of heterotrophic unicyanobacterial consortia (UCC) organisms. Aliases for analyzed UCC genomes are described in **Table 1**. The ability of UCC species to grow on a panel of sugar substrates was predicted based on the presence of the respective reconstructed pathways and carbohydrate-specific transporters in their genomes. Superscript numbers indicate that: (1) the reconstructed pathway contains a missing enzyme; (2) the catabolic pathway is present but carbohydrate-specific transporter is missing; (3) the catabolic pathway includes uncharacterized enzymes with unclear biochemistry. Asterisk indicates the predicted ability to utilize *N*-acetylmuramic acid (the ether of lactic acid and *N*-acetylglucosamine), which is based on the presence of MurQ etherase and the complete *N*-acetylglucosamine catabolic pathway.

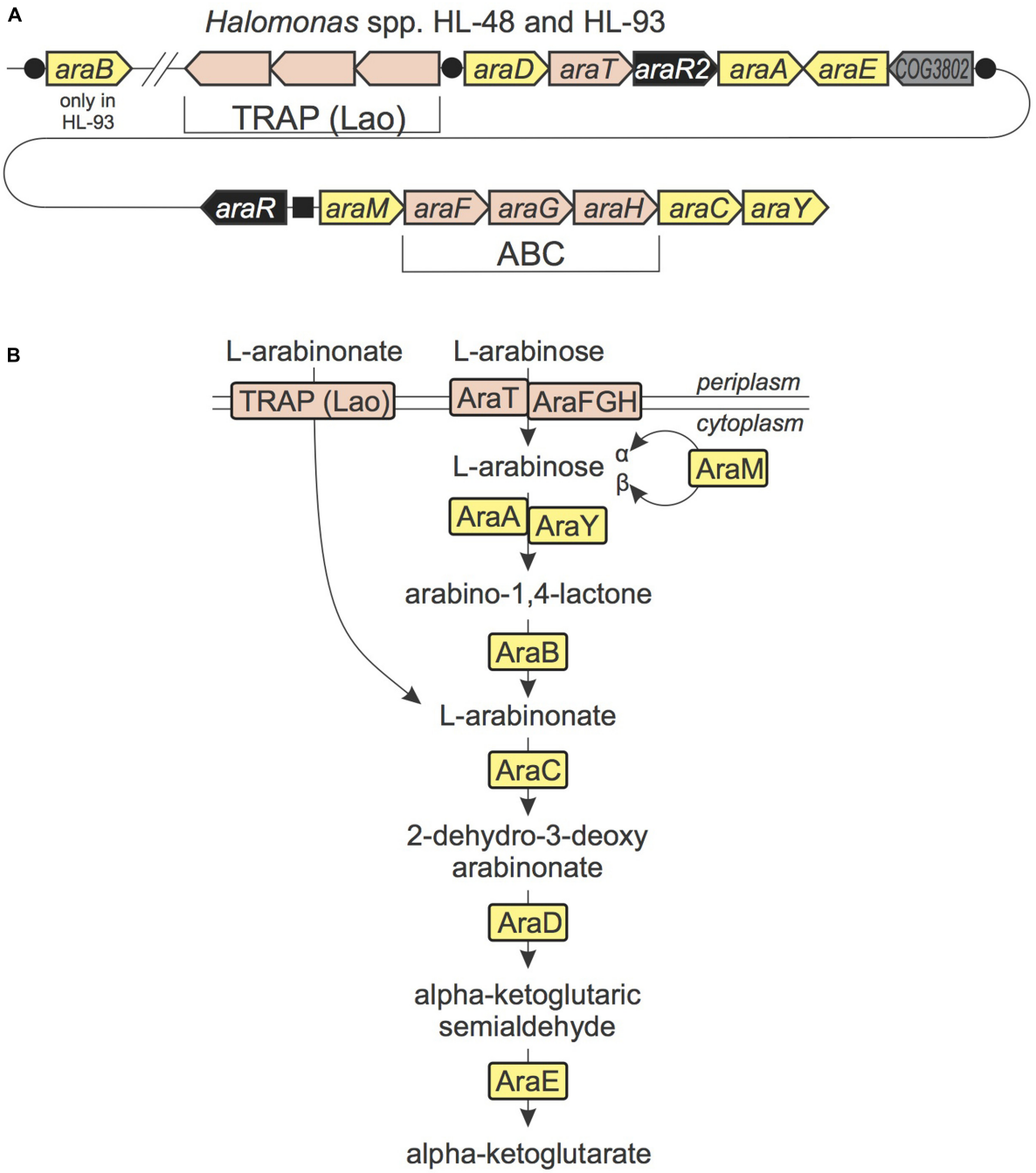


FIGURE 2 | L-Arabinose and L-arabinonate utilization in *Halomonas* species. **(A)** The conserved locus in HL-48 and HL-93 genomes containing genes from L-arabinose and L-arabinonate utilization pathways. Genes are colored by their general functional roles: enzymes (yellow), transcription factors (black), transporters (pink), uncharacterized genes (gray). Candidate TF binding sites are shown as black circles (for AraR2) and a black square (for AraR). **(B)** Scheme of L-arabinose and L-arabinonate utilization pathways in *Halomonas* spp. Enzymes are colored by matching colors as in **(A)**.

pathway, whereas in HL-49 L-fuconate dehydratase FucD is missing and L-fucose dehydrogenase is substituted with a novel non-orthologous dehydrogenase (named $FucO^{II}$). The four UCC genomes have different transporters encoded in the *fuc* loci. HL-49 has a gene encoding an ortholog of fucose permease FucP from *X. campestris*. HL-93 and bin08 have two non-orthologous ABC systems that we predict to be involved in uptake of L-fucose.

The L-fucose/L-fuconate utilization gene loci in both *Halomonas* species contain the *lgoD* gene encoding L-galactonate-5-dehydrogenase, which is the signature gene

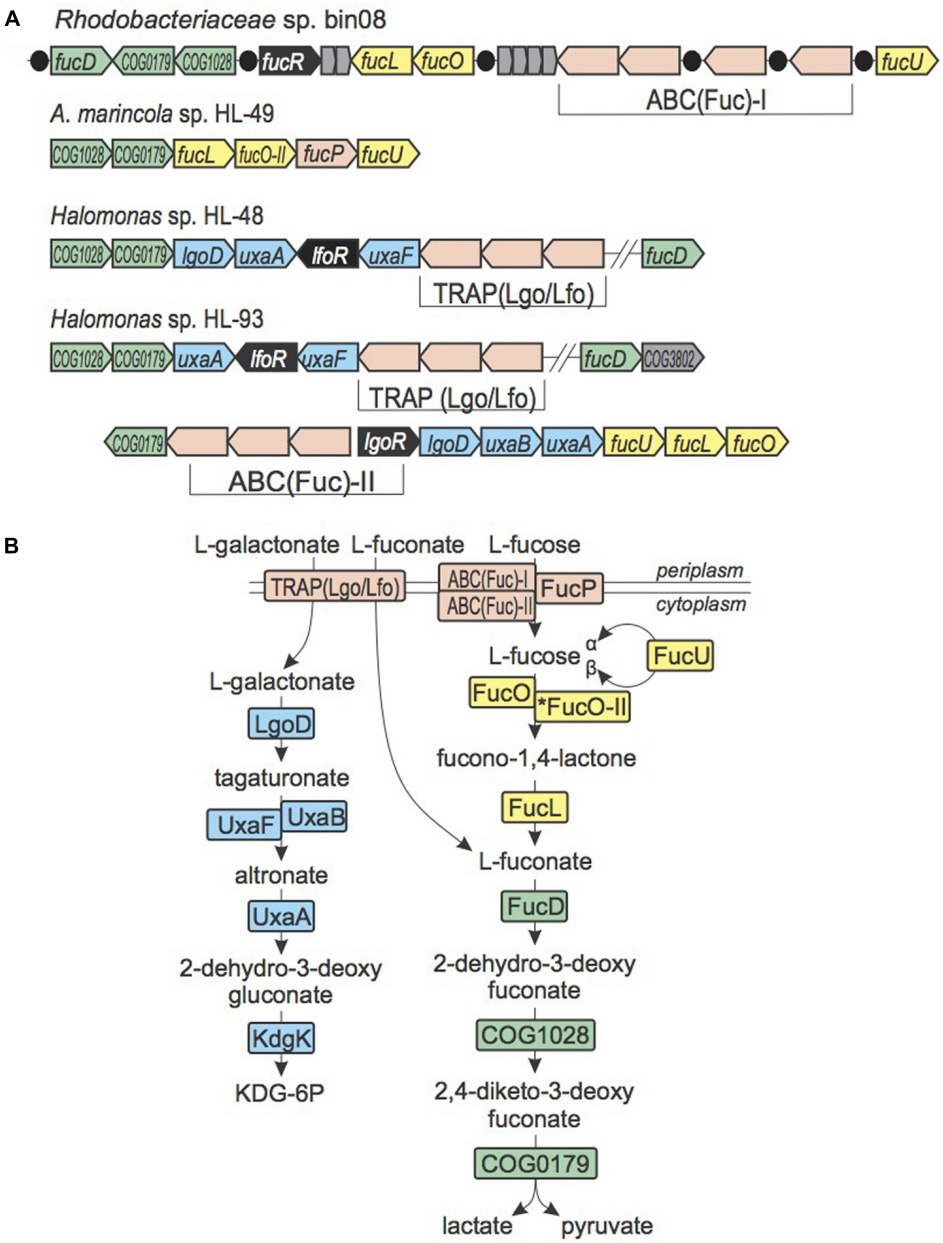


FIGURE 3 | L-Fucose, L-fuconate, and L-galactonate utilization in UCC bacteria. **(A)** Loci containing L-fucose, L-fuconate, and L-galactonate utilization genes. Genes are colored by their general functional roles: transcription factors (black), transporters (pink), uncharacterized genes (gray), enzymes from L-galactonate utilization pathway (blue), enzymes from the L-fucose utilization pathway (yellow), enzymes from the L-fuconate utilization pathway (green). Predicted DNA binding sites of FucR are shown as black circles. **(B)** Scheme of L-fucose, L-fuconate, and L-galactonate utilization pathways. Enzymes are colored by matching colors as in **(A)**.

of L-galactonate utilization (Kuivanen and Richard, 2014), the *uxaA* and *uxaB* (or *uxaF*) genes involved in the downstream steps of L-galactonate catabolism, as well as a novel TRAP-family transporter operon (**Figure 3**). A similar L-galactonate catabolic gene locus in *Chromohalobacter salexigens* (Csal_1738-1731) encodes the same set of catabolic enzymes and a non-orthologous TRAP transporter, which was previously characterized to have a dual specificity toward L-fuconate and L-galactonate (Vetting et al., 2015). Based on these observations, we propose that the novel TRAP system encoded within the L-fucose/L-fuconate/L-galactonate gene loci in *Halomonas* species is involved in the utilization of both L-fuconate and L-galactonate and thus was named Lgo/Lfo. This example introduces an interesting case of chromosomal co-localization (and, likely, co-regulation) of genes that are involved in the shared carbohydrate utilization pathways.

A

Name	Isomerase pathway							DHG pathway					Diacids		L-gulonate			Transporters					GH	Regulators
	UxaC	UxuB	UxuA	UxaB	UxaA	KdgK	KdgA (Eda)	Udh	Gli	Gci	KdgD	KgsD	GarD	GudD	RspB	RspD	RspA	UxuPQM	UxuPQM-II	ABC(GlcA)	TctABC	GulPQM	GluA/LfaA	
HL-49	+	+	+	+	+	+	+												+				+	UxuR2
HL-48			+				(+)				+	+	+	+	+	+	+				+	+		GguR,RspR
HL-93			+				(+)	+	+	+	++	+	+	+	+	+	+	+				+	+		GguR,RspR
HL-91	+	+	+			+	(+)																+	-
bin08	+	+	+			+	(+)													+				UxuR
bin18	+	+	+			+	?													+				UxuR

B

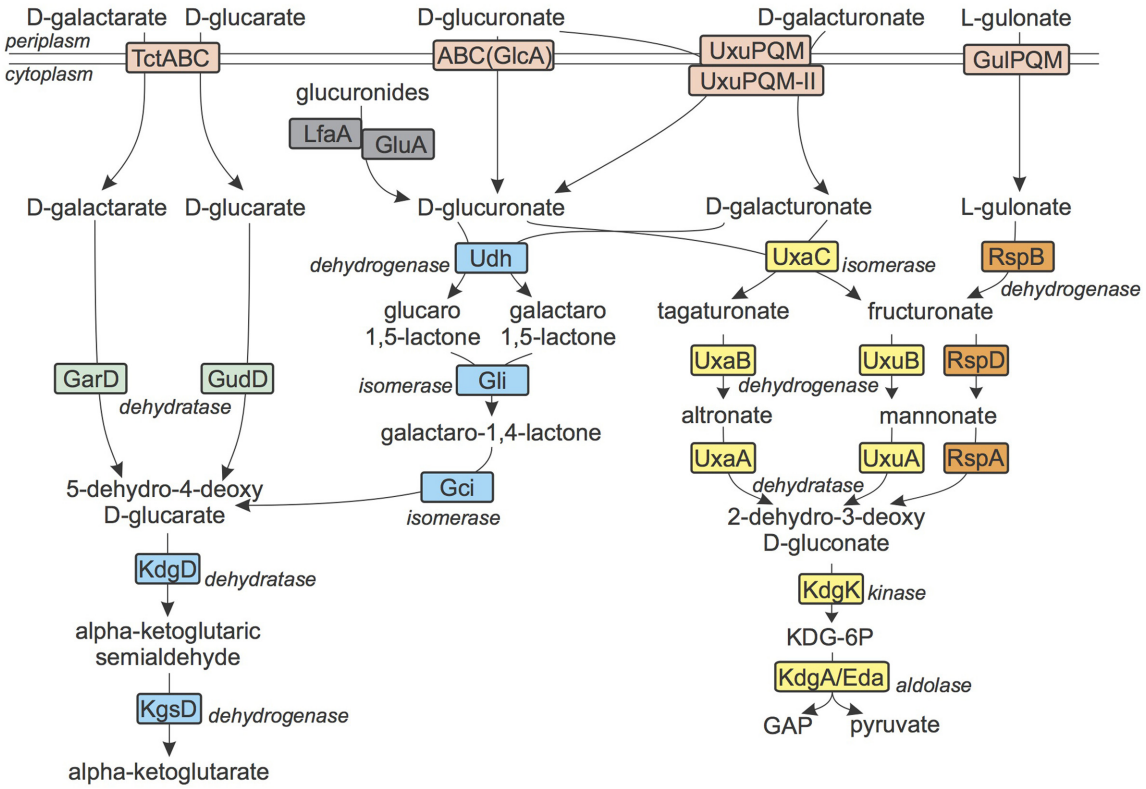


FIGURE 4 | D-Glucuronate, D-galacturonate, D-glucarate, D-galactarate, and L-gulonate utilization in UCC bacteria. **(A)** Distribution of enzymes from the sugar acid utilization pathways in UCC genomes. Presence of enzymes is marked by "+" for each gene with a given function. Each pathway colored by distinct color. Enzymes from central carbohydrate metabolism (CCM) are marked by "(+)." Missing enzymes are marked by "?" **(B)** Scheme of sugar acid utilization pathways in UCC genomes. Enzymes are colored by matching colors as in **(A)**.

Utilization of Hexuronic Acids, Hexose Diacids, and L-Gulonate

D-Galacturonate and D-glucuronate are hexuronic acids that are commonly found in pectins, proteoglycans and glucuronans. D-Galactarate and D-glucarate are the ring opened hexose diacids (or aldaric acids) that serve as a growth substrate to many microorganisms. Both *Halomonas* species (HL-48 and HL-93) contain a gene cluster encoding enzymes involved in the galactarate/glucarate utilization, as well as a novel predicted transporter from the tripartite tricarboxylate transporter (TTT) family (named TctABC) and a novel GntR-family TF (termed GguR). The TctABC transporter is predicted to be involved in galactarate/glucarate uptake. The reconstructed GguR regulon in both *Halomonas* genomes includes the galactarate/glucarate utilization operon, whereas HL-93 has an additional GguR-regulated operon, which encodes the glucoronate/galacturonate utilization enzymes Udh, Gli, and Gci, as well as an ortholog of the known hexuronate transporter UxuPQM. We concluded that HL-93 (but not HL-48) has an additional capability to utilize galacturonate and glucoronate using the pathway partially shared with the galactarate/glucarate pathway (**Figure 4**).

The *R. calidilacus* HL-91 and *Rhodobacteriaceae* bin08 and bin18 genomes encode a different variant of the glucuronate

utilization pathway, which starts from the UxaC isomerase and continues through 2-dehydro-3-deoxygluconate (KDG) and its phosphorylated derivative, KDG-6P. The corresponding glucoronate utilization gene loci include a novel ABC-family transporter operon, which was predicted to be co-regulated with the glucuronate utilization genes by a novel GntR-family regulator (UxuR). However, the corresponding transporter and regulator are missing in HL-91, thus the mechanism of glucuronate uptake in yet unknown in this organism. We speculate that it can take up glucuronides that are hydrolyzed in the cytoplasm by the LfaA glucosidase, thus providing glucoronate to feed the catabolic pathway.

Algoriphagus marincola HL-49 has two separate loci with galacturonate and glucuronate utilization genes. Both of these gene loci are controlled by a novel LacI-family TF (UxuR2) and encode a novel or TRAP-family transporter (UxuPQM^{II}) and an uncharacterized glycosyl hydrolase from the GH109 family. UxuPQM^{II} is distantly related to the previously characterized UxuPQM transporters in various Proteobacteria (Vetting et al., 2015), however, it belongs to a distinct orthologous group of TRAP transporters that are mostly present in the *Bacteroidetes* phylum. We propose that UxuPQM^{II} also has the dual specificity for both hexuronic acids it is co-regulated with both galacturonate and glucuronate utilization genes in HL-49.

In both studied *Halomonas* genomes, we identified a conserved locus encoding proteins homologous to catabolic enzymes, a TRAP-family transporter and a GntR-family regulator that were previously characterized as a part of the L-gulonate utilization pathway in *C. salexigens* (Wichelecki et al., 2014). Thus, we predict that HL-48 and HL-93 are able to utilize L-gulonate.

D-Galactose, D-Galactosides, and D-Galactonate Utilization

In *Escherichia coli* and other Enterobacteria, D-galactose is utilized via the Leloir pathway, which involves galactokinase GalK, galactose-1-phosphate uridylyltransferase GalT and UDP-glucose 4-epimerase GalE (Holden et al., 2003). The *Porphyrobacter* sp. HL-46 and *A. marincola* HL-49 genomes contain the Leloir pathway genes, suggesting there are able to utilize D-galactose (**Figure 5**). The galactose gene locus in HL-46 contains a predicted galactose permease from the SSS family, which is orthologous to the GalP^{II} transporter previously identified in the galactose catabolic gene loci in *Shewanella* spp. (Rodionov et al., 2010). Additionally, the *gal* locus in HL-46 contain two genes encoding cytoplasmic galactosidases, RafA and BgaL, suggesting galactose-containing oligosaccharides may serve as additional inputs to the galactose catabolic pathway. In contrast, the galactose utilization pathway in HL-49 is incomplete with both the GalT uridylyltransferase and a galactose-specific transporter missing.

A different variant of D-galactose catabolic pathway, which is known as the DeLey-Doudoroff pathway, was identified in three *Rhodobacteriaceae* isolates, namely bin08, bin09 and bin18 (**Figure 5**). In this pathway, D-galactose is first oxidized to D-galactonate, which is then converted to pyruvate and GAP through the subsequent action of a dehydratase, a kinase and an aldolase (Wong and Yao, 1994). In addition to the DeLey-Doudoroff pathway enzymes, the galactose utilization gene loci in these three *Rhodobacteriaceae* genomes include genes encoding a cytoplasmic α-galactosidase, an unknown TF from the IclR family and a novel ABC-type transporter. This novel ABC transport system belongs to the Carbohydrate Uptake Transporter-1 (CUT1) family that mostly known to transporter di- and oligo-saccharides, according to the TCDB database (Saier et al., 2014). Thus we propose that this novel transport system is involved in uptake of α-galactosides and that the *Rhodobacteriaceae* spp. are able to utilize as α-galactosides rather than D-galactose.

In *Salinivirga fredricksonii* HL-109, the DeLey-Doudoroff pathway locus is missing the galactose dehydrogenase and lactonase that are required for conversion of D-galactose to D-galactonate. The *dgoK* gene in HL-109 is clustered with genes encoding a novel IclR-family TF (termed GalR) and a novel TTT-family transporter. Known transporters from the TTT family are specific to tricarboxylate and sugar acids (Saier et al., 2014). We predicted that a TTT-family transporter from the incomplete galactose catabolic gene locus is involved in D-galactonate uptake. We also propose that the GalR TF encoded in the same locus senses D-galactonate as an effector. Overall, HL-109 is the only UCC member that is able to utilize this sugar acid.

The DeLey-Doudoroff pathway genes for D-galactonate catabolism are present in *Halomonas* HL-48 and HL-93, as well as in several other reference *Halomonas* genomes. The corresponding *dgo* gene loci contain a hypothetical sugar lactone lactonase from (COG3386, named DgoL), which can serve as a non-orthologous gene displacement for D-galactono-1,4-lactone lactonase GalA. The reconstructed DgoR regulon in the *Halomonas* genomes includes additional candidate co-regulated gene encoding a SSS-family transporter with predicted galactose specificity (GalP^{II}). Although orthologs of known D-galactose dehydrogenase (GalD) are missing in *Halomonas* spp., we tentatively assigned them the galactose utilization capability, which is supported by growth phenotype testing (see below). Further similarity searches revealed one possible candidate for the missing GalD reaction in *Halomonas* spp. – a D-xylose dehydrogenase (XylD) from the xylose utilization gene cluster (it has 47% identity with the characterized GalD enzyme from *Rhizobium meliloti*). Thus, we tentatively propose that XylD in *Halomonas* spp. has specificity to multiple substrates including D-galactose and D-xylose.

Novel Carbohydrate Utilization Pathway Variants

The reconstructed peripheral pathways in heterotrophic UCC members contain 171 novel genes distinguishing them from those previously described in model species. These include 22 genes encoding novel enzymes with assigned function, 107 genes encoding components of novel sugar transporters and 42 novel sugar-specific transcriptional regulators. Most common are numerous cases of non-orthologous gene displacement, when a novel functional role is encoded by a gene that is not orthologous to any of the previously known genes of the same function. Several predicted non-orthologous enzymes

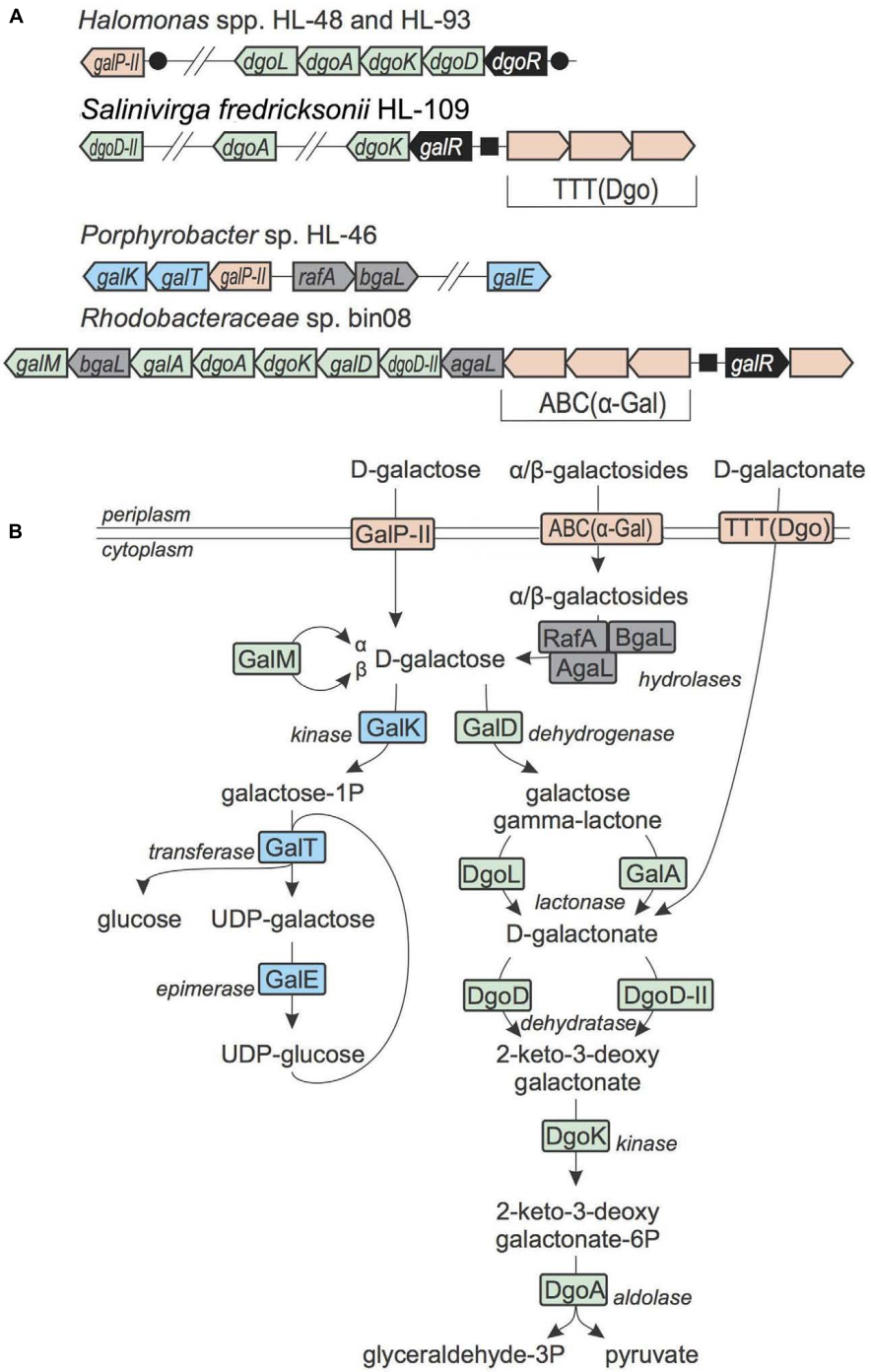


FIGURE 5 | D-Galactose, galactoides, and D-galactonate utilization in UCC bacteria. **(A)** Loci containing D-galactose, galactosides, and D-galactonate utilization genes. Genes are colored by their general functional roles: transcription factors (black), transporters (pink), the Leloir pathway enzymes (blue), the DeLey-Doudoroff pathway enzymes (green), glycosyl hydrolases (gray). Candidate DNA binding sites are shown as circles (for DgoR) and squares (for GalR). **(B)** Scheme of D-galactose, galactosides, and D-galactonate utilization pathways. Enzymes are colored by matching colors as in **(A)**.

involved in utilization of arabinose (AraB, AraD), fucose ($FucO^{II}$), galactose ($DgoD^{II}$, DgoL), and L-galactonate (UxaF) are described in details above. Other proposed cases of non-orthologous enzymes include a novel *N*-acetylglucosamine kinase (COG1070) in bin09, the putative sorbitol dehydrogenase $SorD^{II}$ in bin07, and the predicted fructokinase MtlZ involved in utilization of arabinitol, sorbitol, and mannitol in both *Halomonas* spp.

Carbohydrate uptake transporters constitute the largest group of newly functionally assigned genes in UCC genomes. Most of these genes encode components of 27 multicomponent transport systems from the ABC, TRAP, and TTT families (Supplementary Table S2). Among 18 novel ABC systems most are predicted to transport hexoses (*N*-acetylglucosamine, *N*-acetylgalactosamine, fucose), oligosaccharides (α-/β-galactosides, α-glucosides, fructooligosaccharides), as well as glucoronate, sorbitol, erythritol, and mannoheptulose. All five newly predicted TRAP systems and three TTT-family transporters are specific to sugar acids (hexuronates, arabinonate, galactonate, fuconate) and hexose diacids (glucarate, galactarate). We also identified 12 novel single-component sugar permeases located in the inner membrane (AraT, BglT, $BglT^{II}$, $GalP^{II}$, FruT, MalP, COG2211), three TonB-dependent outer membrane transporters (with predicted specificities to sucrose and β-glucosides) and two novel outer membrane porins (possibly involved in uptake of glucose and glycerol).

Transcriptional regulation is another highly variable aspect of the sugar utilization pathways in UCC genomes. Indeed, 42 of the 88 TFs tentatively associated this UCC sugar utilization pathways are non-orthologous to their counterparts previously characterized in other bacteria, as captured in the RegPrecise database (Novichkov et al., 2013). We identified candidate TFBSs and reconstructed regulons for 60 TFs including 30 novel regulators (Supplementary Table S2). The majority of genes from sugar catabolic pathways were identified as candidate members of respective sugar-specific TF regulons in UCC genomes.

In *Rhodobacteriaceae* bin08, we identified a new gene locus encoding an ABC-family transporter, an aldolase from the tagatose-1,6-bisphosphate aldolase family (COG3684) and two kinases (COG1940 from the ROK family and COG0529 from the adenylylsulfate kinase family). The substrate-binding component of this ABC transport system has 83% similarity to Avi_5339 from *Agrobacterium vitis*, which was previously found to bind mannoheptulose (Steven Almo and John Gerlt, unpublished observation). Mannoheptulose is a heptose, which is structurally similar to D-tagatose. Based on these observations and known substrate specificities for other kinases from the COG1940 and COG0529 families, we propose the following hypothetical pathway for mannoheptulose utilization. The COG1940 kinase first phosphorylates the substrate to produce mannoheptulose-7-phosphate, then the COG0529 kinase further produces mannoheptulose-1,7-biphosphate, which is subject to the COG3684 aldolase reaction producing glycerone phosphate and erythrose-4-phosphate.

Three UCC members are predicted to utilize xylose though the classical isomerase pathway (XylB, XylA), whereas both *Halomonas* species have a different pathway for xylose utilization including xylose dehydrogenase XylD. In *Caulobacter crescentus*, xylose is converted to α-ketoglutarate by xylose dehydrogenase, xylonolactonase, and xylonate dehydratase (Stephens et al., 2007). However, we have not identified candidate genes

TABLE 2 | Carbon sources that support growth of four heterotrophic UCC members.

Carbon sources tested by growth curves and/or Api50 assay[1]				
Hexoses	**Pentoses**	**Disaccharides**	**Sugar alcohols**	**Sugar acids**
***Halomonas* sp. HL-48**				
D-Fructose	D-Xylose	Maltose	**D-Arabitol**	**D-Gluconate**
D-Galactose	L-Arabinose	Sucrose	**D-Mannitol**	
D-Glucose		Trehalose	**D-Sorbitol**	
			Glycerol	
***Halomonas* sp. HL-93**				
D-Fructose	D-Xylose	**Maltose**	D-Arabitol	D-Galacturonate
D-Fucose	L-ARABINOSE	**Sucrose**	D-Mannitol	D-Glucuronate
D-Galactose		**Trehalose**	D-Sorbitol	D-Gluconate
D-Glucose			**D-Erythritol**	
D-Mannose			**Glycerol**	
			Inositol	
***Roseibaca calidilacus* HL-91**				
D-Fructose		Cellobiose	**Glycerol**	
D-Glucose		Maltose		
D-Mannose		**Sucrose**		
GlcNAc		**Trehalose**		
***Marinobacter* sp. HL-58**				
D-Fructose		Maltose	**Glycerol**	
D-Glucose		**Sucrose**		
		Trehalose		

[1]*Carbon sources shown in regular font have positive phenotypes as determined by the Api50 assay. Phenotypes validated by both the Api50 assay and the growth curves measurements are in bold. Phenotypes with measured growth curves but without Api50 tested phenotypes are underlined. GlcNAc, N-acetylglucosamine. The detailed comparison of the predicted and experimentally determined growth phenotypes is provided in Supplementary Table S4.*

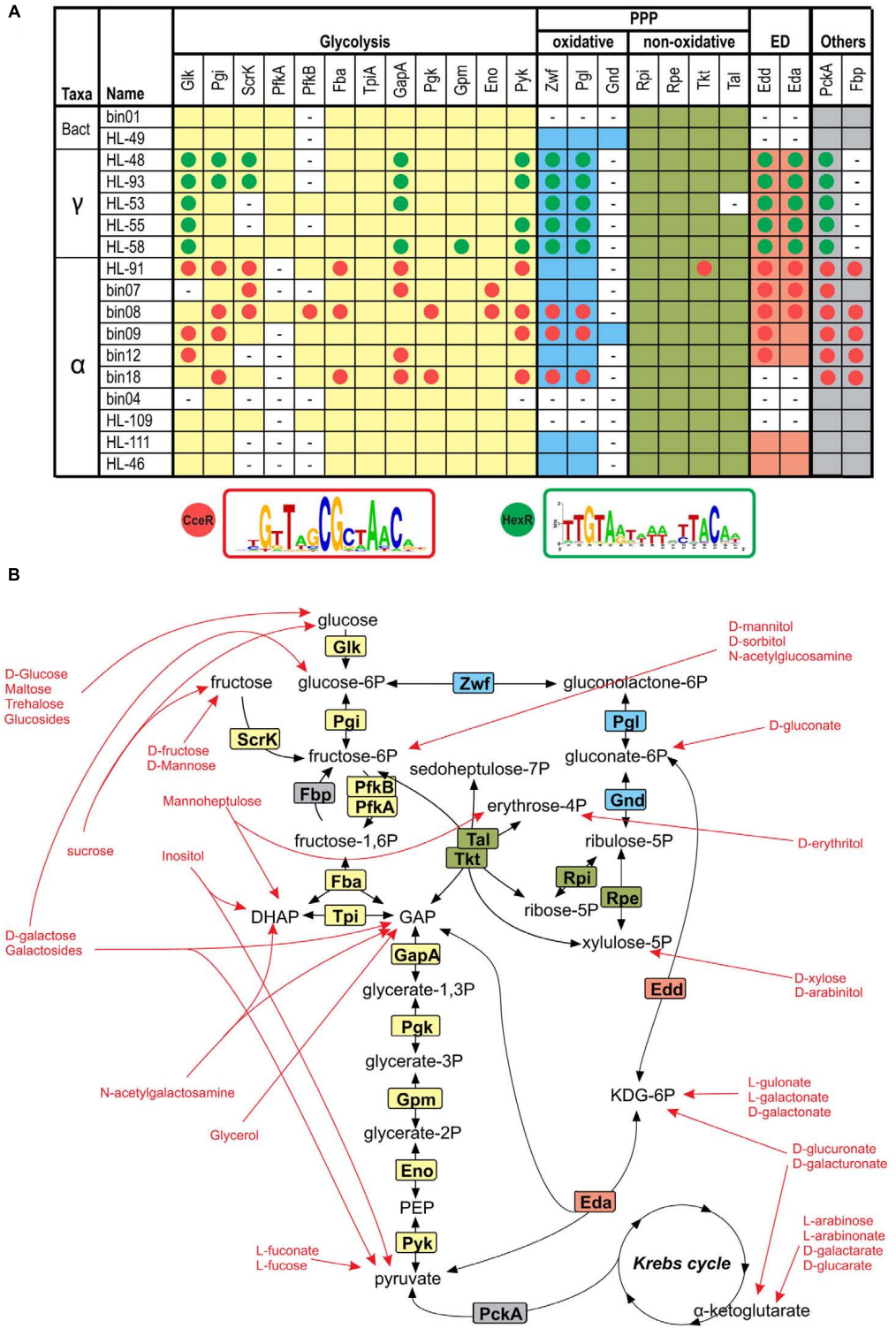


FIGURE 6 | Reconstruction of CCM in heterotrophic UCC organisms. **(A)** Occurrence and regulation of genes involved in glycolysis, oxidative and non-oxidative pentose phosphate (PP) and Entner-Doudoroff (ED) glycolytic pathways in the analyzed genomes. Aliases for analyzed UCC genomes are described in **Table 1**. The presence of genes encoding respective enzymes is shown by background colors matching similar pathways. Candidate members of HexR and CceR regulons are shown by green and red circles, respectively. Sequence logos for the predicted DNA motifs of HexR and CceR regulators are given below the table. The details on the predicted regulator-binding sites and gene locus tags for the identified genes are included in Supplementary Table S2. **(B)** Overview of the CCM and metabolite entrances from peripheral carbohydrate utilization pathways. Enzymes are colored according to the occurrence table in **(A)**. Red arrows point toward final products of each carbohydrate utilization pathway reconstructed in this study.

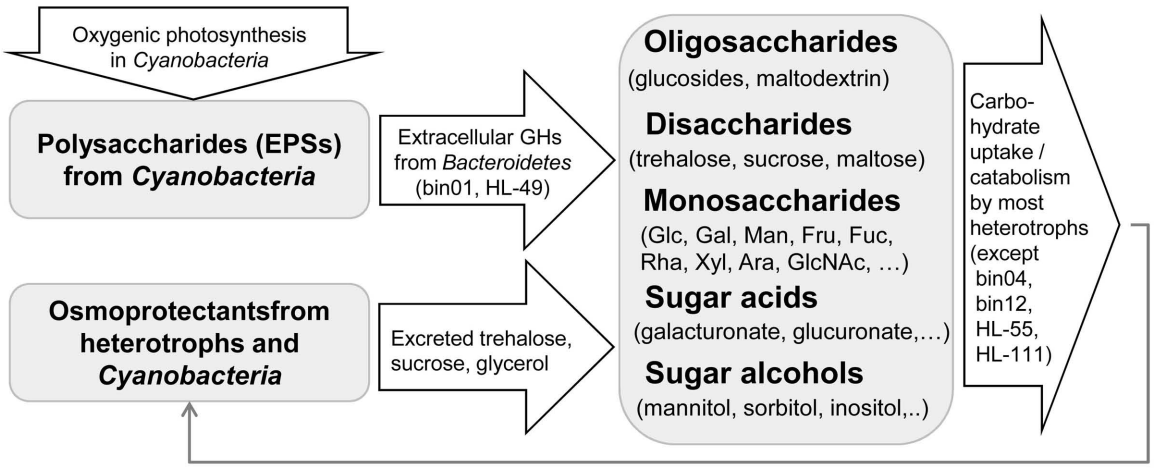


FIGURE 7 | Proposed scheme for carbon flow in the UCC.

for xylonolactonase and xylonate dehydratase in *Halomonas* spp. In contrast, the xylose utilization operon in *Halomonas* spp. encodes two hypothetical enzymes, a sugar phosphate isomerase/epimerase (COG1082) and a Gfo/Idh/MocA-family oxidoreductase (COG0673), however, their exact biochemical functions require further experimental characterization.

Another yet unknown pathway involved in rhamnose utilization was identified in *Rhodobacteriaceae* bin08. Its putative rhamnose catabolic locus contains genes encoding the RhaFGHJ transporter and the RhaM mutarotase, however, bin08 lacks other candidate genes required for utilization of rhamnose (Rodionova et al., 2013b). However, function of other six genes in this locus encoding putative enzymes is unclear. Comparative genomics shows that orthologous loci are present in other *Rhodobacteriaceae* genomes and in some cases they include genes encoding rhamnose dehydrogenase and rhamnonate dehydratase, however bin08 lacks orthologs of these genes, suggesting the existence of a novel yet uncharacterized rhamnose catabolic pathway in the *Rhodobacteriaceae* species.

Growth Phenotype Testing

We tested four UCC isolates including Halomonas sp. HL-48, HL-93, R. calidilacus HL-91, and Marinobacter sp. HL-58 for growth phenotypes on a panel of various hexoses, pentoses, disaccharides, sugar acids, and sugar alcohols (Supplementary Table S4). For the majority of tested carbohydrates we used the Api^{TM}50 CH strip assay. Additionally, we confirmed the selected phenotypes by growing the UCC isolates in defined media using carbohydrates as a single carbon source (see example growth curves provided in Supplementary Figure S1). In accordance with the predicted absence of complete carbohydrate utilization pathways, M. excellens HL-55 did not grow on sugars but it grows on glutamate and lactate (data not shown). We were unable to grow other UCC isolates on defined medium, thus below we report results of the growth tests only for above four UCC organisms.

All four tested organisms demonstrated the ability to grow on two hexoses (glucose, fructose) and three disaccharides (trehalose, maltose, sucrose), as well as on glycerol (**Table 2**). Both Halomonas spp. are able to grow on arabinose, xylose, galactose, arabinitol, mannitol, sorbitol and gluconate, whereas HL-91 grow on N-acetylglucosamine, mannose, and cellobiose (α β-glucoside). In addition, HL-93 has growth phenotypes on fucose, mannose, erythritol, inositol, glucoronate, and galacturonate. With a single exception of the fructose and mannose utilization in HL-91 when we were unable to predict specific catabolic pathways, the measured phenotypes are consistent with the reconstructed catabolic pathways.

Central Carbohydrate Metabolism

Peripheral catabolic pathways produce intermediates that are further catabolized through the CCM pathways including the glycolysis, the oxidative and non-oxidative branches of the pentose phosphate (PP) pathway and the Entner-Doudoroff (ED) pathway (**Figure 6**). To understand downstream parts of the reconstructed catabolic pathways, we searched the genomes of 17 heterotrophic UCC members for known CCM genes (Supplementary Table S3). The complete glycolysis pathway was identified in 13 species, whereas the ED pathway (Zwf, Pgl, Edd, Eda) is present in 12 organisms including three α-proteobacteria that have missing 6-phosphofructokinase Pfk. The bin04 genome lacks the Pfk, Glk, Pyk, Zwf, Edd, and Eda enzymes, suggesting this organism cannot utilize carbohydrates. In agreement with these findings, our genomic analysis did not identify any carbohydrate utilization pathway in bin04. The non-oxidative PP pathway, which is essential for nucleic acid synthesis, was found in all studied genomes. The oxidative PP pathway, which is characterized by the presence of Gnd (in addition to Zwf and Pgl), was identified only in HL-49 and bin09. Thus, bin09 has the most diverse set of CCM pathways involved in sugar utilization.

Bacterial CCM genes are often controlled by global transcriptional regulators, such as HexR and FruR in

γ-proteobacteria, and CceR and GluR in α-proteobacteria (Ravcheev et al., 2014; Imam et al., 2015) Orthologs of HexR and CceR were identified in UCC proteobacteria, and their regulons were reconstructed using the comparative genomics approach (Supplementary Table S3 and **Figure 6**). The RpiR-family regulator HexR that responds to 2-keto-3-deoxy-gluconate-6P (Leyn et al., 2011) was identified in all five γ-proteobacteria, where it mostly regulates the glycolysis and ED pathway genes, as well as the gluconeogenesis gene *pckA*. The LacI-family regulator, CceR, that senses gluconate-6P (Imam et al., 2015) was identified in six α-proteobacteria from the *Rhodobacteriaceae* family, where it controls a broad range of genes involved in the glycolysis, gluconeogenesis, ED and PPP pathways, as well as the ATP synthase genes. Thus, the transcriptional control of the CCM and peripheral sugar catabolic pathways in heterotrophic UCC organisms is mediated by distinct global and local TFs that co-regulate non-overlapping sets of genes.

CONCLUSION

By applying the subsystem-based comparative genomics approach, we reconstructed carbohydrate utilization pathways and predicted catabolic potential for heterotrophic members of UCC consortia. Overall, the reconstructed sugar utilization subsystems include almost 800 genes unevenly distributed across 17 analyzed genomes. Functional roles of 171 genes were first proposed in this study. 13 UCC members were predicted to utilize at least some carbohydrates as a source of carbon and energy using the dedicated catabolic pathways (**Figure 1**). The *Halomonas* strains HL-48 and HL-93 have capabilities to utilize over 20 substrates including pentoses, hexoses, disaccharides, sugar acids, and alcohols. *Rhodobacteriaceae* bin08 is able to grow on 18 substrates including mannoheptulose, a heptose for which we proposed a novel catabolic pathway/transporter/regulon. Each catabolic pathway includes a specialized, often multicomponent, transport system and a set of intracellular enzymes catalyzing biochemical transformations of a particular sugar into one of the common CCM intermediates (**Figure 6**). For the majority of reconstructed pathways we also mapped their cognate TFs that are involved in transcriptional regulation (induction) of catabolic genes, and reconstructed the TF regulons, often allowing the identification of missing transporters and enzymes. Further assessment of two *Halomonas* strains and two other UCC organisms for the growth on a large number of carbon sources allowed us to confirm the majority of the *in silico* predicted catabolic phenotypes. The results of the Api50 testes and extended growth profiling revealed a remarkable consistency between the predicted and observed phenotypes.

Exopolysaccharides produced by autotrophic *Cyanobacteria* serve as the main carbon sources for heterotrophic UCC members (**Figure 7**). Secreted GHs (such as glucosidases) identified in the two *Bacteroidetes* members could benefit the other carbohydrate-utilizing heterotrophs by producing transportable mono- and oligosaccharides. Utilization pathways for disaccharides (maltose/trehalose/sucrose), β-glucosides and glucose are the most abundant among UCC members (present in 8–10 genomes). All other peripheral pathways are present in <30% of the UCC organisms, and among them, 15 pathways are present only in 1–2 genomes. The obtained carbohydrate utilization profiles of UCC heterotrophs are in agreement with the carbohydrate composition of cyanobacterial EPSs (Pereira et al., 2009). Cumulatively, they have pathways to utilize most of the known monosaccharide components of EPSs including glucose, galactose, mannose, fructose, xylose, arabinose, fucose, rhamnose, glucuronate, and galacturonate. Glucosides and disaccharides could be also generated from EPSs. Additionally, trehalose and sucrose are known osmolytes produced by many bacteria to protect them against high salinity levels. There are a plenty of osmoprotectants released into the UCC growth media including glycerol, gluconate, trehalose, and sucrose (Cole et al., 2014). Indeed, the glycerol utilization pathway was identified in five UCC members, while gluconate is utilized by two *Halomonas* isolates. Thus we propose that the UCC community has two levels of carbon donors: (i) *Cyanobacteria* that provide both EPS and osmoprotectants, and (ii) heterotrophic bacteria that could use the cyanobacteria-generated substrates to synthesize their own osmoprotectants and in turn share them with the community. Four UCC members do not rely on carbohydrates for growing. These organisms could use other by-products (such as lactate) secreted by *Cyanobacteria* and other heterotrophs and thus serve as yard cleaners for the community.

Our ability to predict a phenotype of organism from its genome is one of the key goals in microbiology. A systematic application of this omics approach for metabolic reconstruction in a growing number of microbial genomes would allow us to establish the capability of highly accurate automated annotation and assertion of carbohydrate catabolic phenotypes in microbial communities.

AUTHOR CONTRIBUTIONS

SL performed the majority of bioinformatics analysis and wrote the paper; YM performed experimental validation of UCC isolate phenotypes; MR annotated the analyzed genomes; DR designed the study, analyzed the obtained data and wrote the paper.

FUNDING

This research was supported by the Russian Science Foundation (grant #14-14-00289). Additional funding for experimental validation of growth phenotypes was provided by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area.

ACKNOWLEDGMENT

The authors thank Andrei L. Osterman for useful discussions on biochemistry and presentation of the reconstructed metabolic pathways.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01304/full#supplementary-material

REFERENCES

Beliaev, A. S., Romine, M. F., Serres, M., Bernstein, H. C., Linggi, B. E., Markillie, L. M., et al. (2014). Inference of interactions in cyanobacterial-heterotrophic co-cultures via transcriptome sequencing. *ISME J.* 8, 2243–2255. doi: 10.1038/ismej.2014.69

Bendtsen, J. D., Kiemer, L., Fausboll, A., and Brunak, S. (2005a). Non-classical protein secretion in bacteria. *BMC Microbiol.* 5:58. doi: 10.1186/1471-2180-5-58

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005b). Prediction of twin-arginine signal peptides. *BMC Bioinformatics* 6:167. doi: 10.1186/1471-2105-6-167

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., et al. (2016). UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Methods Mol. Biol.* 1374, 23–54. doi: 10.1007/978-1-4939-3167-5_2

Carpenter, E.J., and Foster, R.A. (2002). "Marine cyanobacterial symbioses," in *Cyanobacteria in Symbiosis*, eds A. N. Rai, B. Bergman, and U. Rasmussen (Dordrecht: Springer), 11–17.

Chen, I. M., Markowitz, V. M., Palaniappan, K., Szeto, E., Chu, K., Huang, J., et al. (2016). Supporting community annotation and user collaboration in the integrated microbial genomes (IMG) system. *BMC Genomics* 17:307. doi: 10.1186/s12864-016-2629-y

Chiovitti, A., Bacic, A., Burke, J., and Wetherbee, R. (2003). Heterogeneous xylose-rich glycans are associated with extracellular glycoproteins from the biofouling diatom *Craspedostauros australis* (Bacillariophyceae). *Eur. J. Phycol.* 38, 351–360.

Cole, J. J. (1982). Interactions between bacteria and algae in aquatic ecosystems. *Ann. Rev. Ecol. Syst.* 13, 291–314.

Cole, J. K., Hutchison, J. R., Renslow, R. S., Kim, Y. M., Chrisler, W. B., Engelmann, H. E., et al. (2014). Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Front. Microbiol.* 5:109. doi: 10.3389/fmicb.2014.00109

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

Decho, A. W. (1990). Microbial exopolymer secretions in ocean environments: their role(s) in food webs and marine processes. *Oceanogr. Mar. Biol. Ann. Rev.* 28, 73–153.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Flemming, H. C., and Wingender, J. (2010). The biofilm matrix. *Nat. Rev. Microbiol.* 8, 623–633. doi: 10.1038/nrmicro2415

Gu, Y., Ding, Y., Ren, C., Sun, Z., Rodionov, D. A., Zhang, W., et al. (2010). Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC Genomics* 11:255. doi: 10.1186/1471-2164-11-255

Haft, D. H. (2015). Using comparative genomics to drive new discoveries in microbiology. *Curr. Opin. Microbiol.* 23, 189–196. doi: 10.1016/j.mib.2014.11.017

Hagemann, M. (2011). Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol. Rev.* 35, 87–123. doi: 10.1111/j.1574-6976.2010.00234.x

Holden, H. M., Rayment, I., and Thoden, J. B. (2003). Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J. Biol. Chem.* 278, 43885–43888. doi: 10.1074/jbc.R300025200

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Palaniappan, K., et al. (2015). The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). *Stand. Genomic Sci.* 10, 86. doi: 10.1186/s40793-015-0077-y

Imai, K., Asakawa, N., Tsuji, T., Akazawa, F., Ino, A., Sonoyama, M., et al. (2008). SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in gram-negative bacteria. *Bioinformation* 2, 417–421.

Imam, S., Noguera, D. R., and Donohue, T. J. (2015). CceR and AkgR regulate central carbon and energy metabolism in alphaproteobacteria. *MBio* 6:e02461-14. doi: 10.1128/mBio.02461-14

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. doi: 10.1110/ps.0303703

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Kuivanen, J., and Richard, P. (2014). The yjjN of *E. coli* codes for an L-galactonate dehydrogenase and can be used for quantification of L-galactonate and L-gulonate. *Appl. Biochem. Biotechnol.* 173, 1829–1835. doi: 10.1007/s12010-014-0969-0

Leyn, S. A., Gao, F., Yang, C., and Rodionov, D. A. (2012). N-acetylgalactosamine utilization pathway and regulon in proteobacteria: genomic reconstruction and experimental characterization in *Shewanella*. *J. Biol. Chem.* 287, 28047–28056. doi: 10.1074/jbc.M112.382333

Leyn, S. A., Li, X., Zheng, Q., Novichkov, P. S., Reed, S., Romine, M. F., et al. (2011). Control of proteobacterial central carbon metabolism by the HexR transcriptional regulator: a case study in *Shewanella oneidensis*. *J. Biol. Chem.* 286, 35782–35794. doi: 10.1074/jbc.M111.267963

Nelson, W. C., Maezato, Y., Wu, Y. W., Romine, M. F., and Lindemann, S. R. (2015). Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. *Appl. Environ. Microbiol.* 82, 255–267. doi: 10.1128/AEM.02274-15

Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., et al. (2013). RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* 14:745. doi: 10.1186/1471-2164-14-745

Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., et al. (2010). RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* 38, W299–W307. doi: 10.1093/nar/gkq531

Osterman, A., and Overbeek, R. (2003). Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.* 7, 238–251.

Overbeek, R., Bartels, D., Vonstein, V., and Meyer, F. (2007). Annotation of bacterial and archaeal genomes: improving accuracy and consistency. *Chem. Rev.* 107, 3431–3447. doi: 10.1021/cr068308h

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Pereira, S., Zille, A., Micheletti, E., Moradas-Ferreira, P., De Philippis, R., and Tamagnini, P. (2009). Complexity of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. *FEMS Microbiol. Rev.* 33, 917–941. doi: 10.1111/j.1574-6976.2009.00183.x

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Ravcheev, D. A., Godzik, A., Osterman, A. L., and Rodionov, D. A. (2013). Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* 14:873. doi: 10.1186/1471-2164-14-873

Ravcheev, D. A., Khoroshkin, M. S., Laikova, O. N., Tsoy, O. V., Sernova, N. V., Petrova, S. A., et al. (2014). Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front. Microbiol.* 5:294. doi: 10.3389/fmicb.2014.00294

Rodionov, D. A. (2007). Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem. Rev.* 107, 3467–3497. doi: 10.1021/cr068309+

Rodionov, D. A., Rodionova, I. A., Li, X., Ravcheev, D. A., Tarasova, Y., Portnoy, V. A., et al. (2013). Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. *Front. Microbiol.* 4:244. doi: 10.3389/fmicb.2013.00244

Rodionov, D. A., Yang, C., Li, X., Rodionova, I. A., Wang, Y., Obraztsova, A. Y., et al. (2010). Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 11:494. doi: 10.1186/1471-2164-11-494

Rodionova, I. A., Leyn, S. A., Burkart, M. D., Boucher, N., Noll, K. M., Osterman, A. L., et al. (2013a). Novel inositol catabolic pathway in *Thermotoga maritima*. *Environ. Microbiol.* 15, 2254–2266. doi: 10.1111/1462-2920.12096

Rodionova, I. A., Li, X., Thiel, V., Stolyar, S., Stanton, K., Fredrickson, J. K., et al. (2013b). Comparative genomics and functional analysis of rhamnose catabolic pathways and regulons in bacteria. *Front. Microbiol.* 4:407. doi: 10.3389/fmicb.2013.00407

Rodionova, I. A., Scott, D. A., Grishin, N. V., Osterman, A. L., and Rodionov, D. A. (2012). Tagaturonate-fructuronate epimerase UxaE, a novel enzyme in the hexuronate catabolic network in *Thermotoga maritima*. *Environ. Microbiol.* 14, 2920–2934. doi: 10.1111/j.1462-2920.2012.02856.x

Romine, M. F. (2011). Genome-wide protein localization prediction strategies for gram negative bacteria. *BMC Genomics* 12(Suppl. 1):S1. doi: 10.1186/1471-2164-12-S1-S1

Romine, M. F., Rodionov, D. A., Maezato, Y., Osterman, A. L., and Nelson, W. C. (2017). Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. *ISME J.* 11, 1434–1446. doi: 10.1038/ismej.2017.2

Saier, M. H. Jr., Reddy, V. S., Tamang, D. G., and Vastermark, A. (2014). The transporter classification database. *Nucleic Acids Res.* 42, D251–D258. doi: 10.1093/nar/gkt1097

Seymour, J. R., Ahmed, T., Durham, W. M., and Stocker, R. (2010). Chemotactic response of marine bacteria to the extracellular products of *Synechococcus* and *Prochlorococcus*. *Aquat. Microb. Ecol.* 59, 161–168.

Stal, L. J., and Moezelaar, R. (1997). Fermentation in cyanobacteria. *FEMS Microbiol. Rev.* 21, 179–211.

Stephens, C., Christen, B., Fuchs, T., Sundaram, V., Watanabe, K., and Jenal, U. (2007). Genetic analysis of a novel pathway for D-xylose metabolism in *Caulobacter crescentus*. *J. Bacteriol.* 189, 2181–2185. doi: 10.1128/JB.01438-06

Tsirigos, K. D., Elofsson, A., and Bagos, P. G. (2016). PRED-TMBB2: improved topology prediction and detection of beta-barrel outer membrane proteins. *Bioinformatics* 32, i665–i671. doi: 10.1093/bioinformatics/btw444

Vetting, M. W., Al-Obaidi, N., Zhao, S., San Francisco, B., Kim, J., Wichelecki, D. J., et al. (2015). Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* 54, 909–931. doi: 10.1021/bi501388y

Warda, A. K., Siezen, R. J., Boekhorst, J., Wells-Bennik, M. H., de Jong, A., Kuipers, O. P., et al. (2016). Linking *Bacillus cereus* genotypes and carbohydrate utilization capacity. *PLoS ONE* 11:e0156796. doi: 10.1371/journal.pone.0156796

Watanabe, S., Kodaki, T., and Makino, K. (2006a). Cloning, expression, and characterization of bacterial L-arabinose 1-dehydrogenase involved in an alternative pathway of L-arabinose metabolism. *J. Biol. Chem.* 281, 2612–2623. doi: 10.1074/jbc.M506477200

Watanabe, S., Kodaki, T., and Makino, K. (2006b). A novel alpha-ketoglutaric semialdehyde dehydrogenase: evolutionary insight into an alternative pathway of bacterial L-arabinose metabolism. *J. Biol. Chem.* 281, 28876–28888. doi: 10.1074/jbc.M602585200

Wichelecki, D. J., Vendiola, J. A., Jones, A. M., Al-Obaidi, N., Almo, S. C., and Gerlt, J. A. (2014). Investigating the physiological roles of low-efficiency D-mannonate and D-gluconate dehydratases in the enolase superfamily: pathways for the catabolism of L-gulonate and L-idonate. *Biochemistry* 53, 5692–5699. doi: 10.1021/bi500837w

Wong, T. Y., and Yao, X. T. (1994). The DeLey-doudoroff pathway of galactose metabolism in *Azotobacter vinelandii*. *Appl. Environ. Microbiol.* 60, 2065–2068.

Yang, C., Rodionov, D. A., Li, X., Laikova, O. N., Gelfand, M. S., Zagnitko, O. P., et al. (2006). Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J. Biol. Chem.* 281, 29872–29885. doi: 10.1074/jbc.M605052200

Yew, W. S., Fedorov, A. A., Fedorov, E. V., Rakus, J. F., Pierce, R. W., Almo, S. C., et al. (2006). Evolution of enzymatic activities in the enolase superfamily: L-fuconate dehydratase from *Xanthomonas campestris*. *Biochemistry* 45, 14582–14597. doi: 10.1021/bi061687o

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

Zhang, L., Leyn, S. A., Gu, Y., Jiang, W., Rodionov, D. A., and Yang, C. (2012). Ribulokinase and transcriptional regulation of arabinose metabolism in *Clostridium acetobutylicum*. *J. Bacteriol.* 194, 1055–1064. doi: 10.1128/JB.06241-11

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Leyn, Maezato, Romine and Rodionov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 July 2017
doi: 10.3389/fmicb.2017.01268

Genetic Diversity of Nitrogen-Fixing and Plant Growth Promoting *Pseudomonas* Species Isolated from Sugarcane Rhizosphere

Hai-Bi Li[1†], Rajesh K. Singh[1†], Pratiksha Singh[1], Qi-Qi Song[1], Yong-Xiu Xing[1], Li-Tao Yang[1] and Yang-Rui Li[1,2*]*

[1] *Agricultural College, State Key Laboratory of Subtropical Bioresources Conservation and Utilization, Guangxi University, Nanning, China,* [2] *Key Laboratory of Sugarcane Biotechnology and Genetic Improvement Guangxi, Ministry of Agriculture, Sugarcane Research Center, Chinese Academy of Agricultural Sciences, Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China*

OPEN ACCESS

Edited by:
Florence Abram, NUI Galway, Ireland

Reviewed by:
Romy Chakraborty, Lawrence Berkeley National Lab, United States
Jay Prakash Verma, Banaras Hindu University, India

***Correspondence:**
Li-Tao Yang
liyr@gxu.edu.cn
Yang-Rui Li
liyr@gxaas.net

[†]These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *23 January 2017*
Accepted: *23 June 2017*
Published: *14 July 2017*

Citation:
Li H-B, Singh RK, Singh P, Song Q-Q, Xing Y-X, Yang L-T and Li Y-R (2017) Genetic Diversity of Nitrogen-Fixing and Plant Growth Promoting Pseudomonas Species Isolated from Sugarcane Rhizosphere. Front. Microbiol. 8:1268. doi: 10.3389/fmicb.2017.01268

The study was designed to isolate and characterize *Pseudomonas* spp. from sugarcane rhizosphere, and to evaluate their plant- growth- promoting (PGP) traits and nitrogenase activity. A biological nitrogen-fixing microbe has great potential to replace chemical fertilizers and be used as a targeted biofertilizer in a plant. A total of 100 isolates from sugarcane rhizosphere, belonging to different species, were isolated; from these, 30 isolates were selected on the basis of preliminary screening, for *in vitro* antagonistic activities against sugarcane pathogens and for various PGP traits, as well as nitrogenase activity. The production of IAA varied from 312.07 to 13.12 μg mL^{-1} in tryptophan supplemented medium, with higher production in AN15 and lower in CN20 strain. The estimation of ACC deaminase activity, strains CY4 and BA2 produced maximum and minimum activity of 77.0 and 15.13 μmoL mg^{-1} h^{-1}. For nitrogenase activity among the studied strains, CoA6 fixed higher and AY1 fixed lower in amounts (108.30 and 6.16 μmoL C_2H_2 h^{-1} mL^{-1}). All the strains were identified on the basis of 16S rRNA gene sequencing, and the phylogenetic diversity of the strains was analyzed. The results identified all strains as being similar to *Pseudomonas* spp. Polymerase chain reaction (PCR) amplification of *nifH* and antibiotic genes was suggestive that the amplified strains had the capability to fix nitrogen and possessed biocontrol activities. Genotypic comparisons of the strains were determined by BOX, ERIC, and REP PCR profile analysis. Out of all the screened isolates, CY4 (*Pseudomonas koreensis*) and CN11 (*Pseudomonas entomophila*) showed the most prominent PGP traits, as well as nitrogenase activity. Therefore, only these two strains were selected for further studies; Biolog profiling; colonization through green fluorescent protein (GFP)-tagged bacteria; and *nifH* gene expression using quantitative real-time polymerase chain reaction (qRT-PCR) analysis. The Biolog phenotypic profiling, which comprised utilization of C and N sources, and tolerance to osmolytes and pH, revealed the metabolic versatility of the selected strains. The colonization ability of the selected strains was evaluated by genetically tagging them with a constitutively expressing GFP-pPROBE-pTet^{r}-OT plasmid. qRT-PCR results showed that both strains had the ability to express the *nifH* gene at 90 and 120 days,

as compared to a control, in both sugarcane varieties GT11 and GXB9. Therefore, our isolated strains, *P. koreensis* and *P. entomophila* may be used as inoculums or in biofertilizer production for enhancing growth and nutrients, as well as for improving nitrogen levels, in sugarcane and other crops. The present study, to the best of our knowledge, is the first report on the diversity of *Pseudomonas* spp. associated with sugarcane in Guangxi, China.

Keywords: antibiotic gene, genetic diversity, GFP, *Pseudomonas*, *nifH*, sugarcane, Biolog

INTRODUCTION

Sugarcane (*Saccharum officinarum* L.) is one of the most important industrial agricultural crops, being cultivated in over 110 tropical and subtropical countries and, providing a source of sugar, renewable energy, and biomaterials (Fischer et al., 2012). The main sugarcane producing country in the world market is Brazil, and the next major producers are India, China and Thailand (FAO, 2016). More than fifty diseases are caused by plant pathogens in sugarcane (Croft and Magarey, 2000; Rao et al., 2002) with 10–15% of sugar being lost due to such diseases. Among them, red rot, smut, wilt, and pineapple diseases caused by fungi, and ratoon stunting disease caused by bacteria, are found to cause considerable yield loss (Viswanathan and Rao, 2011). Sugarcane is a long duration economical crop, so it requires large amounts of plant nutrients i.e., N, P, K, as well as of other micro nutrients. An abundant supply of nitrogen is required for the early stages of plant growth. However, in many countries, farmers apply even higher doses of fertilizers, chemicals, and pesticides to sugarcane to promote early growth and development and to increase yields. Although, N fertilizer use is comparatively low in Brazil ($\sim$50 kg N ha^{-1}), other countries average $\sim$120–300 kg N ha^{-1} with extreme rates in excess of 700 kg N ha^{-1} (Robinson et al., 2011). However, a higher dose of fertilizer not only raises the production cost, but also causes serious environmental pollution (Herridge et al., 2008; Li and Yang, 2015). It may have negative and unpredictable effects on the environment, and contribute to the pollution of soil, water, and natural areas.

There is vast microbial flora available globally, and microbes are found in all types of soils, such as sands, deserts, and soils of volcanic origin, and in bogs and moors, snow covered soils, sediments, and semi-aquatic ecosystems, and on rocks (Manoharachary and Mukerji, 2006). There is a clear incentive to exploit this microbial diversity and to isolate and develop functional microbes that can be used, in effect, as targeted fertilizers as an alternative to traditional fertilizer applications. Here, we focused on nitrogen-fixing bacterial genera that are often found in large populations in rhizospheric soils and that exhibit general disease-suppression and PGP traits. In principle, biological nitrogen fixation (BNF) promises an alternative approach to plant N fertilizer requirements (Xing et al., 2006, 2015). Some Brazilian sugarcane varieties are capable of obtaining substantial nitrogen from the soil through BNF (Lima et al., 1987; Urquiaga et al., 1992, 2012). Studies using long-term N balances, ^{15}N natural abundance, and ^{15}N isotope dilution methods have shown that some sugarcane cultivars can obtain a significant amount of their nitrogen requirements in this way (Urquiaga et al., 1992, 2012), but the bacteria responsible remains unknown (Boddey et al., 1995; James and Olivares, 1998; James, 2000).

A diverse array of bacteria, including species of Azoarcus, Azospirillum, Arthrobacter, Azotobacter, Bacillus, Burkholderia, Erwinia, Enterobacter, Gluconacetobacter, Herbaspirillum seropedicae, Klebsiella, Kosakonia, Paenibacillus, Pantoea, Pseudomonas, Stenotrophomonas, Serratia, and Xanthomonas are among the main plant growth promoting rhizobacteria used to promote the growth of several crops, including sugarcane (Somers et al., 2004; Bhattacharyya and Jha, 2012; Carvalho et al., 2014; Rafikova et al., 2016; Xing et al., 2016; Solanki et al., 2017). Strains of some bacterial genera, e.g., Azotobacter, Bacillus, Enterobacter, Pseudomonas, Serratia, and Azospirillum are already being used as biofertilizers for enhancing the growth and yield of crops, as well as for maintaining soil fertility (De Souza et al., 2015). Nitrogen-fixing microorganisms play an important role both in the soil and in plants. Many plant-associated rhizobacteria are recognized for their PGP ability, for their capacity to increase disease resistance, and for their of phytohormones under various stress conditions. However, the use of inoculated microbial activity requires observation of the efficiency and colonization rate to track and identify the inoculated strain within the host plant. We used a popular marker gene that encodes green fluorescent protein (GFP), which was easily detected in cell samples by using confocal microscopy (Unge et al., 1999). Confocal laser scanning electron microscopy (CLSEM), in combination with GFP is a powerful tool for studying plant-microbe interactions (Chi et al., 2004; Liu et al., 2006). N_2-fixing microorganisms contain dinitrogenase, one of the subunits of which is encoded by nifH, the detection of nifH mRNA indicates the presence of N_2-fixing bacteria, as well as indicating N_2-fixation in plants (Young, 1992). Quantitative real-time polymerase chain reaction (qRT-PCR) has been found to be a powerful approach for quantification of an active N_2-fixing population within a multifarious community (Wallenstein, 2004). The undeviating contribution of plant-inhabiting diazotrophs (*Azospirillum* spp., *Rhizobium* spp., etc.) to nitrogen-fixation and nitrogen-uptake in cucumber has been estimated by quantifying the nifH gene copy number using qRT-PCR (Juraeva et al., 2006).

Researchers have explored and focused on identifying N_2-fixing *Pseudomonas* associated with plants to increase crop production, reduce harmful chemicals and protect the soil

and environment. *Pseudomonas* spp. belong to the family *Pseudomonadaceae,* which contains a large number of species and is divided into subdivisions (Mehnaz, 2011). *Pseudomonas fluorescens* and *Pseudomonas putida* are very well-known and well-studied species of this genus that have been used as inoculums to promote plant growth. Some earlier reports on the isolation of *Pseudomonas* from sugarcane are: *Pseudomonas* spp. (Li and Macrae, 1991; Antwerpen et al., 2002; Magnani et al., 2010), *P. aeruginosa* (Viswanathan et al., 2003), *P. aurantia*ca (Mehnaz et al., 2009b), *P. fluorescens* (Viswanathana and Samiyappan, 2002; Mendes et al., 2007; Mehnaz et al., 2009a), *P. putida* (Viswanathana and Samiyappan, 2002; Mehnaz et al., 2009a), and *P. reactans* (Mehnaz et al., 2010).

In this study, *Pseudomonas* strains were isolated from the rhizosphere of sugarcane plants grown in the field in Guangxi, China. We specifically focused on the diversity of *Pseudomonas* spp. and the major objectives were: (1) to investigate the antagonistic ability of *Pseudomonas* spp. isolated from sugarcane against pathogens; (2) to evaluate their PGP traits as well as nitrogenase activities in order to use them further as bio-fertilizers; (3) to use polymerase chain reaction (PCR) and qRT-PCR based techniques to detect the *nifH* gene, characterize antibiotic genes, and assess their genetic diversity through BOX, ERIC and REP-PCR; (4) to analyse 16S rRNA gene sequences for effective identification of *Pseudomonas* spp.; (5) to test the utilization of numerous sources of carbon, and nitrogen, as well as tolerance to osmolytes and different pH conditions and (6) to investigate the interaction mechanisms between the sugarcane plant and selected potential strains through a GFP technique. As for the available literature, to the best of our knowledge, this is the first report on nitrogen-fixing *Pseudomonas koreensis* and *Pseudomonas entomophila* isolated from sugarcane in China.

MATERIALS AND METHODS

Locations and Collection of Soil Sampling Site and Properties

The study area is Nanning City, Guangxi Autonomous Region in South China. It has a warm with an average temperature of 21.7°C and humid subtropical climate. Summers are hot and the average temperature is 25 (lowest), 33°C (highest) in July, and winters are the coldest, 10°C in January. The average annual rainfall is between 1,000–2,800 mm and precipitation is 1,372 mm. It is situated between 22°49′1.21″ N latitude and 108°21′59.55″ E longitude, and elevation is 79.51 m.

Soil samples were randomly collected from sugarcane fields that have highly fertile soils. Five healthy plants were sampled using a sterile auger from different locations in a sterile specimen container and immediately transported to the laboratory. In all cases, soil samples were taken from 2 to 20 cm layers in April 2015. The soil particles attached to roots were carefully collected after uprooting plants and mixed well. Root debris was removed by sieving through 2 mm mesh. Samples were stored at 4°C for further studies and processed within 24 h of collection. The pulverized soil samples were used for analysis of physico-chemical properties.

Media and Growth Conditions of Bacterial Strains

We selected four different enrichment media (Ashbey's medium, Yeast Mannitol Agar, LGI, and Dworkin and Foster salts minimal medium) for the isolation of bacteria from soil samples; all the media contained some component that permitted the growth of specific types of nitrogen-fixing bacteria. A universal nutrient agar (NA) medium was also used for the isolation of all types of bacterial strains (Table S1). Ten grams of soil from each sample were separately suspended in 90 mL of saline water (0.85% of NaCl) in a flask and placed on an orbital shaker (at 100 rpm) at 30 ± 2°C for 1 h.

In vitro Test of Plant-Growth-Promoting Attributes

The growth promotion traits of all bacterial isolates were evaluated by performing standard protocol for the estimation of indole acetic acid (IAA), P-solubilization, siderophore, hydrogen cyanide (HCN) and ammonia production according to Glickmann and Dessaux (1995), Brick et al. (1991), Schwyn and Neilands (1987), Lorck (1948) and Dey et al. (2004), respectively. For P-solubilization, plates containing Pikovskaya's media amended with tri-calcium phosphate were observed for clearing or solubilisation zones around the colonies. For siderophore production the Chrome Azurol S (CAS) medium was prepared by mixing 25 mL of solution A (composition in gL^{-1}: 60.5 mg CAS was dissolved in 50 mL of distilled water and 10 mL iron (III) solution; 1 mM $FeCl_3.6H_2O$, 10 mm HCl). This solution was slowly added to 72.9 mg hexadecyl-trimethyl ammonium bromide (HDTMA) dissolved in 40 mL of water. The resultant dark medium was autoclaved and 75 mL of solution B (Nutrient agar) after autoclaving separately mixed at about 40–50°C and chromazurol sulphonate agar plate has been prepared. The bacterial strains were spot inoculated on chromazurol sulphonate agar plate medium and incubated at 28 ± 2°C for 2–6 days. After incubation of plates siderophore production was assayed by the change in the color of the medium from blue to orange haloes zone formation.

Antifungal Activity

All isolates were evaluated for their *in vitro* antifungal activity by dual culture assays on NA + potato dextrose agar (1:1) plate against the plant pathogens, *Ustilago scitaminea* and *Ceratocystis paradoxa* according to Singh et al. (2013, 2014). The strains exhibiting more than 50% inhibition in mycelial growth were considered as promising antagonists.

Nitrogen Fixation by Acetylene Reduction Assay (ARA)

Nitrogen-fixing ability of all isolate was tested by using ARA previously described by Hardy et al. (1968). All bacterial isolate was inoculated in a 25 mL flask containing 10 mL semi solid JNFb medium (Table S1) and bacteria were grown at 30 ± 2°C for 3 days. Five Percent air from the tubes was replaced by acetylene through a syringe, incubated for 12 h, 0.5 mL gas was withdrawn from the tube, and ethylene formation was analyzed through a gas

chromatograph (GC-17A, Shimadzu, Kyoto, Japan) with a flame ionization detector and a column filled with DB-1701 (Agilent, Santa Clara, USA).

1-Aminocyclopropane-1-Carboxylate (ACC) Deaminase Activity

Screening for ACC deaminase activity of all isolates was done based on their ability to use ACC as a sole nitrogen source, a trait that is consequence of the activity of the enzyme ACC deaminase. All the isolates were grown in 10 mL of LB broth medium incubated at 30 ± 2°C at 120 rpm for 24–36 h. The cells were harvested by centrifugation at 10,000 rpm for 5 min and washed twice with sterile 0.1 M Tris-HCl (pH 7.5). And it was spotted on petri plates of the modified nitrogen free Dworkin and Foster (DF) medium (Jacobson et al., 1994). Plates without ACC were used as negative control and those with ACC (3 mM) or $(NH_4)_2SO_4$ (0.2% w/v) plates were used as positive control. The plates were incubated at 30 ± 2°C for 3–5 days. The isolates had ability to grow on ACC plates confirmed that it possessed ACC deaminase activity.

Quantitative estimations of ACC deaminase activities of the bacterial isolates were determined for selected isolates according to the procedures described by Honma and Shimomura (1978) with a standard curve of α-ketobutyrate ranging between 0.01 and 1.0 μ mole. The absorbance was measured at 540 nm (Shimadzu UV-1800, Japan). Enzyme activity was expressed as μmol mg^{-1} protein h^{-1}.

Genomic DNA Extraction

Extraction of total genomic DNA from all the selected bacterial strains was performed using a DNA isolation kit (CWBIO, Beijing, China). After extraction, the quantity, integrity, and quality of the DNA obtained were checked by 0.8% (wt/vol) agarose gel electrophoresis, followed by staining in ethidium bromide, and visualization under UV light. The extracted DNA was further quantified using a nano-photometer (Pearl, Implen-3780, USA).

PCR Amplification of 16S rRNA Gene

To amplify the 16S rRNA gene, polymerase chain reaction (PCR) was performed from the genomic DNA of strains using a universal primer pair for pA-F and pH-R (**Table 1**). The PCR program for 16S rRNA gene included initial denaturation at 95°C for 5 min, 30 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1 min, extension at 72°C for 1 min and final extension cycle at 72°C for 5 min. Amplified fragments were checked and purified by using PCR purification kit BioFlux (Hangzhou, China) and then sequenced at Sangon Biotech (Shanghai, China).

Phylogenetic Analysis

To perform molecular phylogenetic analysis and evolutionary relationship analysis, the 16S rRNA gene sequences of the isolated *Pseudomonas* strains were compared with reference strain sequences deposited in the National Center for Biotechnology Information (NCBI) GenBank public database. The sequences were aligned by ClustalW (Saitou and Nei, 1987) and the phylogenetic tree was reconstructed using MEGA software version 7.0 (Kumar et al., 2016) and unweighted pair group method with arithmetic mean (UPGMA) (Sneath and Sokal, 1973) in a Kimura two-parameter model (Tamura et al., 2004). To obtain the confidence values, the gaps were treated by pairwise deletions and bootstrap analysis was carried out by the method of Felsenstein (1985) using 1,000 pseudoreplications.

Amplification, Cloning, and Sequencing of the *nifH* Gene

Isolated DNA template from all 30 selected strains were used to amplify a conserved region of the *nifH* gene fragment by PCR, according to the method of Poly et al. (2001) using the primers PolF and PolR (**Table 1**).

Detection of Antibiotic Genes

Genomic DNA of *Pseudomonas* strains was used for the PCR amplification of genes involved in the biosynthesis of three different types of antibiotics genes:- phenazine-1-carboxylic acid (PhCA) (Raaijmakers et al., 1997), pyrrolnitrin (PRN) (Souza and Raaijmakers, 2003), and hydrogen cyanide (HCN) (Ramette et al., 2003) respectively (**Table 1**). The antibiotics PhCA and PRN are broad-spectrum antibiotics produced by several strains of *Pseudomonas*, which play an important role in the suppression of multiple plant pathogenic fungi, whereas HCN is a broad-spectrum antimicrobial compound that play a key role in the biological control of root diseases by many plant-associated *Pseudomonas* isolates.

Genetic Diversity Studies

The genetic PCR fingerprinting was carried out using repetitive consensus with BOX-PCR (based on primers targeting the highly conserved repetitive DNA sequences of the BOXA subunit of the BOX element), ERIC-PCR (based on primers targeting the highly conserved enterobacterial repetitive intergenic consensus) and REP-PCR (based on primers targeting the repetitive extragenic palindromic sequence). The genomic fingerprints were obtained as described by Rademaker and de Bruijn (1997) to determine phylogenetic relatedness and sequences of different primers (**Table 1**). All PCR reactions were carried out in Peltier Thermal Cycler BIORAD. PCR amplifications were performed in a 25 μL reaction volume, the reaction mixture and conditions are given in Table S2. Amplification was analyzed by 1.6% agarose gels electrophoresis containing 0.5 μg mL^{-1} ethidium bromide. A low range ladder (TaKaRa, Dalian, China) was used as molecular size marker. The gels were visualized and gel images were documented in Bio-Rad gel documentation system. The profiles generated by genetic diversity analysis were compared by calculating Jaccard's similarity coefficient for each pairwise comparison and dendrogram was constructed from the similarity matrix by the unweighted pairgroup method with arithmetic average (UPGMA) using NTSYS pc, version 2.02h. All the experiment was carried out in three replicates.

BIOLOG(R) Phenotypic Assays

Assays of utilization of potential carbon (C), nitrogen (N), and tolerance to different osmotic and pH conditions were tested using BIOLOG Phenotype Micro-Array™ plates GENIII, PM3B,

TABLE 1 | PCR primers used for identification, functional genes amplifications and *nifH* gene expression.

Gene	Primer	Sequence (5′→ 3′)	Product size	References
16S	pA-F	AGAGTTTGATCCTGGCTCAG	1,300–1,500 bp	Edwards et al., 1989
	pH-R	AAGGAGGTGATCCAGCCGCA		
NifH	Pol-F	TGCGAYCC-SAARGCBGACTC	360 bp	Poly et al., 2001
	Pol-R	ATSGCCATCATYTCRCCGGA		
PhCA	PhCA-F	TTGCCAAGCCTCGCTCCAAC	1,150 bp	Raaijmakers et al., 1997
	PhCA-R	CCGCGTTGTTCCTCGTTCAT		
PRN	Prn-F	GGGGCGGGCCGTGGTGATGGA	786 bp	Souza and Raaijmakers, 2003
	Prn-R	YCCCGCSGCCTGYCTGGTCTG		
HCN	HCN-F	ACTGCCAGGGGCGGATGTGC	587 bp	Ramette et al., 2003
	HCN-R	ACGATGTGCTCGGCGTAC		
BOX	BOX A1R	CTACGGCAAGGCGACGCTGACG	50–5,000 bp	Rademaker and de Bruijn, 1997
ERIC	ERIC-1	ATGTAAGCTCCTGGGGATTCAC	50–5,000 bp	
	ERIC-2	AAGTAAGTGACTGGGGTGAGCG		
REP	REP-F	IIIICGICGICATCIGGC	50–5,000 bp	
	REP-R	ICGICTTATCIGGCCTAC		
Q-RT-PCR PRIMER				
NifH	Pol-F	TGCGAYCC-SAARGCBGACTC		Poly et al., 2001
	Pol-R	ATSGCCATCATYTCRCCGGA		
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH-1	CTCTGCCCCAAGCAAAGATG		Niu et al., 2013
	GAPDH-2	TGTTGTGCAGCTAGCATTG		

PM9, and PM10 (Biolog Inc., Hayward, CA). The number of all the possible conditions was assayed in the four different types of microplates. GENIII plates were used to study C sources metabolism, and PM3B plates to assess N metabolism sources, respectively. In addition, PM9 and PM10 plates were used to test the growth under various stress conditions and different pH (Bochner, 2009; Mazur et al., 2013). Two isolates (CY4 and CN11) were grown at 30 ± 2°C on LB agar medium and then suspended in an inoculation fluid (IF) after washing to get the transmittance of 90–98% according to procedure. A 100 mL cell suspension was then transferred into the 96 wells of all Micro-Plate, and then incubated at 30 ± 2°C for 48 h to allow the phenotypic fingerprint to form. During incubation there is an increased respiration in the wells and cells can utilize different sources and grow. Increased the respiration causes reduction of the tetrazolium dye and forming a purple color. After incubation the readings were obtained using automated BIOLOG(R) Micro-Station Reader according to the instructions of the manufacturer.

Green Fluorescent Protein Technique

Plasmid Transformation

The *Pseudomonas* strains (CY4 and CN11) resistant to ampicillin (40 μg mL^{-1}) were chosen as recipients for genetic tagging with GFP-pPROBE-pTet^r-OT. Both the strains were sensitive to kanamycin (100 μg mL^{-1}). Plasmid pPROBE-pTet^r-OT containing the GFP and kanamycin genes expressed under the control of a Tet^r promoter was introduced by biparental mating using donor strain *E. coli* TG1. Plasmid pPROBE-pTet^r-OT is a derivative of plasmid pBBR1, which is a small (2.6 kb), broad-host range plasmid and stably maintained in a number of gram (+) and (−) bacteria. The recipient strains and donor strains were mixed at a ratio of 1:2. An aliquot (100 μL) of this mixture was spread onto LB agar. After overnight incubation of the plates at 30°C, the bacteria were washed off the plates and suitable dilutions of the cultures were plated onto selective media containing ampicillin (40 μg mL^{-1}) and kanamycin (100 μg mL^{-1}). Ex-conjugants showing green fluorescence under UV illumination were selected for further study.

Inoculation of Micro-propagated Sugarcane Plantlets

Sugarcane micro-propagated plantlets were inoculated with *Pseudomonas* strains and GFP-tagged *Pseudomonas* strain/pPROBEpTet^r-TT as described by Oliveira et al. (2002). Five individual rooted plantlets were transferred into a glass bottle of 300 mL capacity containing 50 mL of liquid one-tenth MS medium (sucrose and basal salt mixture) (Reis et al., 1999). Two days after plant transfer, the media free of clear microbial contamination were inoculated with bacterial suspension, providing an initial bacterial cell numbers of approximately 2.0×10^5 mL^{-1} medium. Plantlets without inoculation were prepared as control.

TABLE 2 | *In vitro* biochemical characterization of antagonistic *Pseudomonas* species on the plant growth promotion traits isolated from sugarcane.

Culture code	Phosphate	Siderophore	Ammonia	HCN	ACC	Antifungal activity	
						Ustilago scitaminea	*Ceratocystis paradoxa*
CoY1	++	+++	−	−	+	++	−
CoA5	++	−	++	−	−	+	++
CoA6	++	+++	++	−	−	+	++
AY1	++	++	−	−	−	++	−
AY2	+++	+++	−	−	+	−	−
AA8	++	−	−	+++	−	++	+
AN15	−	+++	++	++	−	−	+++
BY2	++	+++	−	++	+	++	−
BY3	+++	−	−	−	−	++	−
BY10	+++	−	−	−	+	−	++
BA2	++	−	−	−	+	++	−
BA4	+++	−	−	−	−	++	−
BA7	+++	−	+++	−	−	−	−
BA12	+++	+++	−	++	+	++	−
BA13	+++	−	+++	−	−	−	+
BA17	++	+++	−	++	+	−	−
BN6	++	+++	+++	+++	−	−	−
BN7	++	+++	+++	+++	−	−	−
CY4	++	++	+++	++	+	+	−
CY7	++	++	−	−	+	+	++
CA5	++	+++	+++	+++	+	−	−
CA7	+++	+++	−	+++	+	−	+++
CN1	+++	++	−	−	+	−	−
CN2	−	++	+++	−	+	−	−
CN3	+++	−	−	−	+	+	−
CN9	+++	+++	−	++	−	−	+
CN11	+++	+++	+++	+++	+	−	+
CN12	+++	−	−	−	+	−	+++
CN15	−	+++	−	+++	+	−	−
CN20	−	++	++	−	+	+	−

+, showing low activity; ++, moderate activity; +++, strong activity; −, showing no activity.
Phosphate and Siderophore activity of the bacterial isolates 3 mm or greater clear zone of inhibitions on suitable medium after 3–5 days of incubation at 30 ± 2°C.
Antifungal activity by dual culture plate measured as zone of inhibition after 3–5 days of incubation at 26 ± 2°C.

Plantlets were grown in a growth chamber at 30°C with a 14 h photoperiod at a 60 μ moL m^{-2} s^{-1} photon flux density.

Laser Scanning Confocal Microscopy (Olympus SXZ16)

Three days after inoculation, sugarcane tissue culture plantlets (inoculated and un-inoculated) were taken out from the tubes and sugarcane plantlets were washed with autoclaved distilled water. After cut into small pieces, the root and stem tissues were mounted on bridge slide with 10% (v/v) glycerol. Whole root parts and optical sections of the root and stem pieces were observed with a Leica DMI 6000 microscope attached to a Leica TCS SP5 laser scanning confocal microscope (Leica Microsystems, Mannheim, Germany). GFP fluorescence was detected with an emission band from 500 nm to 530 nm while root auto fluorescence was detected by adjusting the band width from 600 nm to 800 nm depending on the intensity of auto fluorescence (Lin et al., 2012).

RNA Extraction and cDNA Synthesis

Total RNA was extracted using Trizol reagent (Tiangen, Beijing, China), according to the manufacturer's instructions. DNA contamination of the RNA was removed by DNase I (Promega, USA). The extracted RNA was further quantified using a Nano photometer (Pearl, Implen-3780, USA). Synthesis of first-strand cDNA from the RNA was carried out using the Prime-ScriptTM RT Reagent Kit (TaKaRa, Dalian, China).

Quantitative Real-Time Polymerase Chain Reaction

Expression patterns of the target *nifH* gene during plant-microbe interaction were studied in a greenhouse experiment for the selected strains (CY4 and CN11) in two sugarcane varieties

TABLE 3 | Screening of different *Pseudomonas* species for IAA, ARA, and ACC deaminase activity.

Culture Code	IAA ($\mu g\ mL^{-1}$)		ARA ($\mu moL\ C_2H_2h^{-1}\ mL^{-1}$)	ACC ($\mu moL\ mg^{-1}\ h^{-1}$)
	A-Tryptophane	P-Tryptophane		
CoY1	21.84^{a}	118.12^{e}	13.61^{ijk}	30.32^{ghi}
CoA5	12.92^{k}	130.86^{c}	18.41^{g}	–
CoA6	18.92^{bc}	153.31^{b}	108.30^{a}	–
AY1	15.27^{fgh}	76.19^{f}	6.16^{n}	–
AY2	14.23^{g-k}	128.35^{d}	17.46^{gh}	44.23^{d}
AA8	14.23^{g-k}	20.63^{h}	22.32^{ef}	–
AN15	13.90^{g-k}	312.07^{a}	25.91^{bcd}	–
BY2	17.44^{cde}	17.43^{ij}	22.34^{ef}	32.92^{fgh}
BY3	16.47^{ef}	16.32^{ijk}	25.71^{b-e}	–
BY10	13.84^{g-k}	14.23^{kl}	29.26^{b}	19.04^{k}
BA2	16.38^{ef}	16.37^{ijk}	19.30^{fg}	15.13^{k}
BA4	17.17^{de}	17.40^{ij}	27.08^{bcd}	–
BA7	13.92^{g-k}	13.74^{kl}	25.59^{cde}	–
BA12	15.38^{fg}	15.37^{jkl}	26.04^{bcd}	36.49^{ef}
BA13	14.72^{g-j}	14.92^{jkl}	18.12^{g}	–
BA17	15.14^{fgh}	15.08^{jkl}	23.83^{de}	60.78^{b}
BN6	15.21^{fgh}	15.21^{jkl}	17.34^{gh}	–
BN7	23.13^{a}	23.12^{g}	26.08^{bcd}	–
CY4	20.25^{b}	310.63^{a}	26.33^{bcd}	77.00^{a}
CY7	14.90^{f-i}	14.90^{jkl}	28.30^{bc}	24.33^{j}
CA5	13.84^{g-k}	13.84^{kl}	8.49^{lmn}	25.97^{ij}
CA7	14.51^{g-k}	14.36^{kl}	10.82^{kl}	52.92^{c}
CN1	13.62^{h-k}	13.59^{l}	6.87^{mn}	32.72^{fgh}
CN2	13.32^{ijk}	13.26^{l}	13.62^{ijk}	29.68^{hi}
CN3	13.99^{g-k}	14.06^{kl}	18.41^{g}	34.99^{fg}
CN9	23.24^{a}	23.31^{g}	14.24^{hij}	–
CN11	23.00^{a}	23.26^{g}	16.89^{ghi}	33.17^{fgh}
CN12	18.16^{cd}	18.29^{i}	11.11^{jkl}	29.64^{hi}
CN15	13.77^{g-k}	13.77^{kl}	9.71^{lm}	40.23^{de}
CN20	13.12^{jk}	13.12^{l}	18.58^{g}	15.62^{k}
SEM	0.499	0.785	1.082	1.479
CD ($P = 0.05$)	1.413	2.224	3.063	4.252
CV (%)	5.3	2.5	8.6	7.3

Means followed by same letter within a row are not significantly different ($P \leq 0.05$) according to Duncan's Multiple Range Test (DMRT). SEM standard error of the difference between means, CD, critical difference; CV, coefficient of variation.

(GT11 and GXB9) at 90 and 120 days, as well as for a control. Leaf samples of both sugarcane varieties were used as the experimental material. The relative expression of the target genes was calculated as the expression level of the inoculated sample minus the level of the control at each corresponding time point. Each qRT-PCR experiment was conducted in triplicate. qRT-PCR was carried out with SYBR Premix Ex Tap™ II (TaKaRa, Japan) in Real Time PCR Detection System (Bio-Rad, USA). Reactions were carried out in a final volume of 20 μL, which contained 10 μL SYBR Premix Ex Tap™ II, 2 μL template (10 × diluted cDNA), 0.8 μL of each 10 μM primer and 6.4 μL ddH_2O. PCR with distilled water as the template was performed as a control. The primer sequences for *nifH* are presented in **Table 1**. The qRT-PCR program was as follows: 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 20 s (Niu et al., 2013). Melting curve analysis was conducted at the end of amplification to confirm the specificity of the reaction. The $2^{-\Delta\Delta Ct}$ method was used to quantify the relative gene expression (Livak and Schmittgen, 2001).

Statistical Analysis

Experimental data was analyzed using standard analysis of variance (ANOVA) followed by Duncan's multiple range test (DMRT). Standard errors were calculated for all mean values. Differences were considered significant at the $p \leq 0.05$ level. All biochemical experiments were performed in triplicate, and the results were expressed as mean values. OrigiPro 9.1 (2013) software was used for the principle component analysis (PCA).

RESULTS

Analysis of Chemical and Trace elements Composition in the Soil

The soil texture was analyzed, and characterized as a medium loam. The pH varied from a low of 5.99 to a high of 6.70, and the electrical conductivity of the soil samples varied from 111.0 to 72.8 μS cm^{-1}. The chemical analysis of the soil revealed that the total amounts of nitrogen (N), phosphorus (P) and potassium (K) were 0.30, 0.43, and 13.81 g kg^{-1}, respectively, at the time of sampling, (April 2015); calcium and magnesium levels were 784.6 and 147.6 mg kg^{-1}, respectively. The amounts of trace elements (mg kg^{-1}) were Fe, 106.14; Mn, 85.22; Zn, 7.49; B, 0.42; SO_4^{2-}, 136.67; Cl^-, 30.64.

Isolation and PGP Potential

A total of 350 bacterial strains were isolated from the sugarcane rhizospheric soil samples. Of these, 100 strains were selected on basis of morphology. These selected strains were further tested by *in vitro* screening for antifungal activity against sugarcane pathogens (*U. scitaminea* and *C. paradoxa*), and for selected PGP traits, as well as for nitrogenase activity. PGP activities were evident from the ability of the selected isolates to produce plant hormones. Of these, only 30 strains were selected that showed various different PGP traits, such as P-solubilization, or siderophore, ammonia, HCN, ACC, or indole-3-acetic acid IAA production, or acetylene reduction activity for molecular identification. In the case of phosphate solubilization, out of the 30 isolates, 26 (87%) isolates showed halo zone formation on Pikovskaya's agar plates, confirming their ability that solubilized the tricalcium phosphate in the medium. An *in vitro* siderophore production assay revealed that 20 (66.67%) of the strains were able to produce siderophores, 40% of the strains were able to produce ammonia, and 43.33% were able to produce HCN. In the case of qualitative screening of ACC-utilizing bacterial strains, only 18 (60%) exhibited ACC deaminase activity, the strains consumed 3 mM ACC in DF-ACC medium after 48 h incubation at 30 ± 2°C and the color in the DF-ACC medium containing the bacterial strains appeared weaker compared with the non-inoculated medium. In the quantitative estimation of ACC deaminase activity, all strains ranged from

TABLE 4 | Identification of putative plant growth-promoting bacterial strains isolated from sugarcane based on 16S rDNA sequence and NCBI GenBank databases.

Isolates	16S rRNA gene	% Similarity	Accession number match	No. of nucleotides	Accession numbers
CoY1	*Pseudomonas monteilii*	97	KX785170	1,501	KY460972
CoA5	*Pseudomonas aeruginosa*	99	KF929419	1,438	KY460973
CoA6	*Pseudomonas putida*	99	KC952984	1,440	KY460974
AY1	*Pseudomonas putida*	99	KU187966	1,414	KY460975
AY2	*Pseudomonas monteilii*	99	LC015566	1,383	KY460976
AA8	*Pseudomonas koreensis*	99	KC790278	1,424	KY460977
AN15	*Pseudomonas koreensis*	99	KC790275	1,446	KY460978
BY2	*Pseudomonas putida*	98	KT759148	1,040	KY460979
BY3	*Pseudomonas* sp.	99	KX891558	1,162	KY460980
BY10	*Pseudomonas plecoglossicida*	99	LT671913	1,396	KY460981
BA2	*Pseudomonas plecoglossicida*	97	JQ976892	1,436	KY460982
BA4	*Pseudomonas* sp.	99	GU372931	1,433	KY460983
BA7	*Pseudomonas* sp.	99	LC093430	1,486	KY460984
BA12	*Pseudomonas taiwanensis*	99	KU597525	1,417	KY460985
BA13	*Pseudomonas putida*	99	KT273281	1,458	KY460986
BA17	*Pseudomonas taiwanensis*	99	KU597525	1,433	KY460987
BN6	*Pseudomonas entomophila*	99	KU601313	1,386	KY460988
BN7	*Pseudomonas entomophila*	99	KU601313	1,409	KY460989
CY4	*Pseudomonas koreensis*	99	KF484686	1,444	KY460990
CY7	*Pseudomonas* sp.	99	GU372931	1,429	KY460991
CA5	*Pseudomonas* sp.	98	KP126777	1,445	KY460992
CA7	*Pseudomonas mosselii*	98	KU550148	1,353	KY460993
CN1	*Pseudomonas mosselii*	99	LC015563	1,419	KY460994
CN2	*Pseudomonas plecoglossicida*	99	KY006168	1,404	KY460995
CN3	*Pseudomonas* sp.	99	KX168054	1,460	KY460996
CN9	*Pseudomonas mosselii*	98	LN995691	1,490	KY460997
CN11	*Pseudomonas entomophila*	98	KU601313	1,398	KY460998
CN12	*Pseudomonas monteilii*	98	KX785170	1,514	KY460999
CN15	*Pseudomonas putida*	98	KJ850211	1,410	KY460100
CN20	*Pseudomonas* sp.	97	KT034417	1,448	KY460101

77.0 to 15.13 μmoL mg^{-1} h^{-1}, with a maximum activity in CY4 and a minimum in the BA2 strain (**Table 3**). Twenty-one isolates showed antagonistic activity against sugarcane pathogens. Fourteen isolates were antagonistic to *U. scitaminea* (46.66%), and eleven were antagonistic to *C. paradoxa* (36.67%) (**Table 2**).

Biosynthesis of IAA showed differences among the strains, which are summarized in **Table 3**. The quantitative production of IAA varied from 312.07 to 13.12 μg mL^{-1} in tryptophan supplemented medium, and a higher level of production was recorded in strain AN15, and a lower one in CN20. In the case of medium without tryptophan, the maximum IAA production was observed from the strain BN7 (23.24 μg mL^{-1}) and the minimum from CoA5 (12.92 μg mL^{-1}). Nitrogen fixation efficiency of all strains was estimated by the acetylene reduction assay (ARA) under laboratory conditions. The data revealed considerable variability in the nitrogenase activity among the studied strains, which ranged from 108.30 to 6.16 μmoL C_2H_2 h^{-1} mL^{-1}. On average, under our experimental conditions, the strain CoA6 fixed higher amounts and AY1 fixed lower amounts of nitrogen as compared to the other strains (**Table 3**).

Molecular Identification and Phylogenetic Analysis

Identification of all isolates was performed based on partial 16S rRNA gene sequencing. The strain sequences obtained were compared using the BlastN tool, with NCBI GenBank database nucleotide sequences and similarity values ≥97% being obtained. The results are suggestive that all the isolates belonged to different species of *Pseudomonas,* i.e., *P. monteilii* (3), *P. aeruginosa* (1), *P. putida* (5), *P. koreensis* (3), *P.* spp. (7), *P. plecoglossicida* (3), *P. taiwanensis* (2), *P. entomophilla* (3), and *P. mosselii* (3), and the nucleotide sequence data have been submitted to the NCBI GenBank database (**Table 4**). These DNA sequences were aligned and used to reconstruct a phylogenetic tree, possessing five clusters with 1,000 bootstrap samplings with representative strains of related taxa, as shown in **Figure 1**.

Detection of *nifH* and Antibiotic Genes

PCR products of the correct size were amplified from the total genomic DNA extracted from all strains. Of these, 10 strains were positive for *nifH* gene amplification, producing an amplified

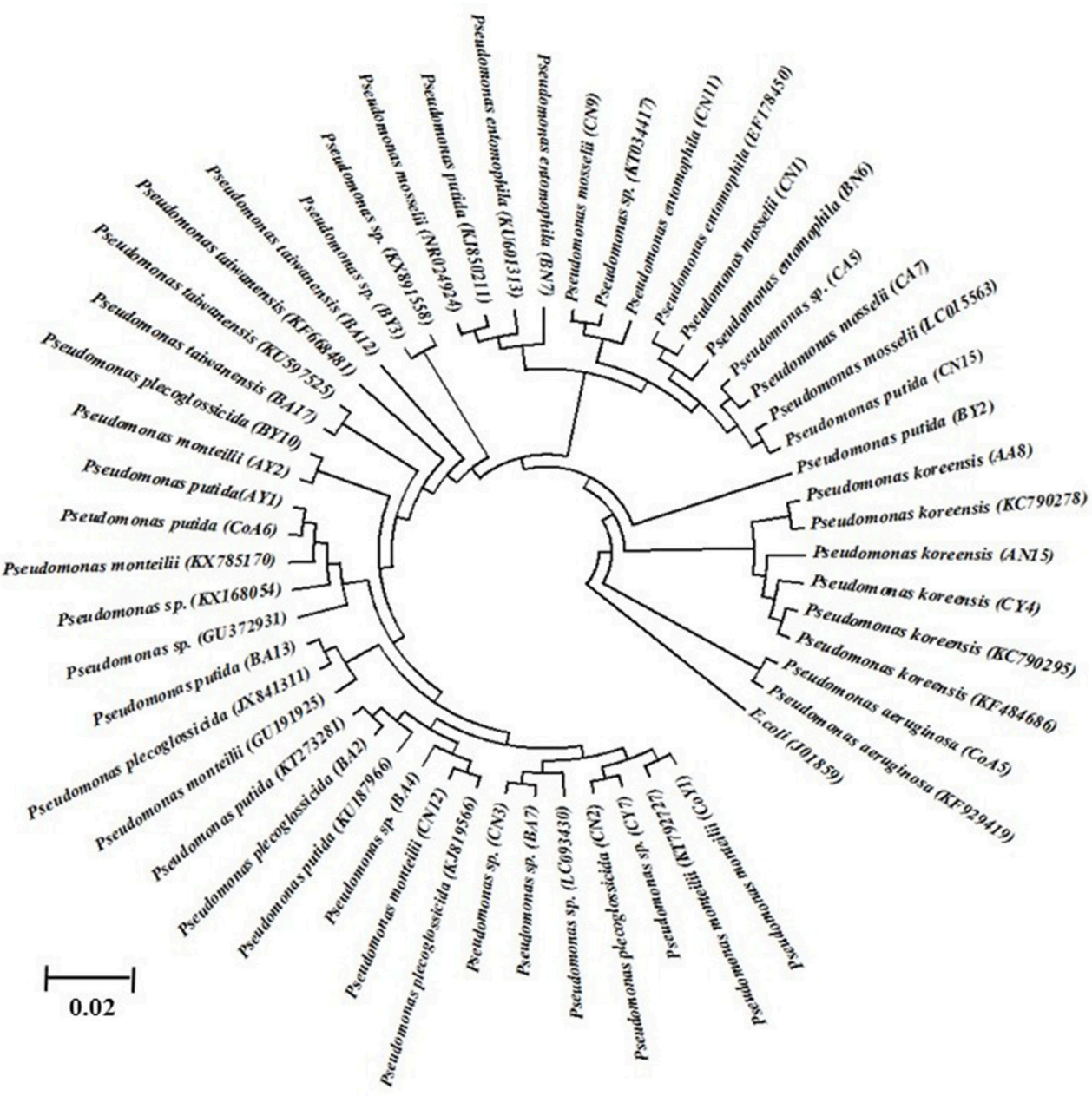


FIGURE 1 | The 16S rDNA phylogenetic tree of *Pseudomonas* isolates from sugarcane. The evolutionary history was inferred using the UPGMA method. Bootstrap values of 1,000 replications are indicated as percent confidence values for particular branching. All positions containing gaps and missing data were eliminated. Sequences indicated in code were determined in this study. The *E. coli* was used as an out group.

fragment of about 360 bp (**Figure 2**); these strains were CoA5, BA12, CY4, CN9, CN15, AY1, CN3, BA4, BA17, and AN15. The All positive strains were selected and used to establish *nifH* clone libraries. A total of ten clones selected from each strain were analyzed by sequencing. All clones were sequenced, analyzed, and identified by BlastN. All sequences were related to the partial *nifH* gene. The clones obtained had similarity levels that varied from 90 to 100% and the NCBI GenBank accession numbers are KY508382–KY508391.

Amplifications of three genes of antibiotic biosynthesis; (involved in biocontrol activity in the biosynthesis of PhCA, PRN, and HCN) were used in this study. Only four strains showed the PhCA related amplified band size of 1.15 kb, and five showed PRN related gene amplification at around 786 bp. In the case of the HCN related gene, nine strains showed positive amplification at 587 bp. All the amplified products were purified and sequenced. The antibiotic related gene sequences were analyzed and identified using the BlastN program, but only

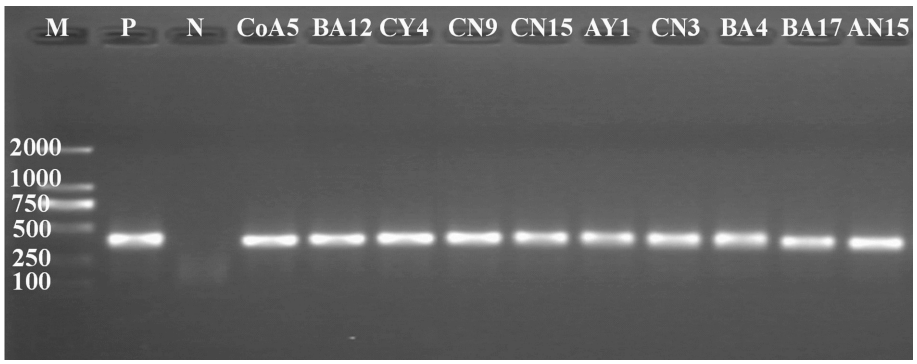


FIGURE 2 | PCR-amplification of *nifH* gene from genomic DNA of *Pseudomanas* species. M, molecular size marker from 100 to 2,000 bp. P is a positive control (*Klebsiella verticola*- DX120E) and N is a negative control (Water). M, molecular size marker (100 bp–2 kb), (Takara).

the HCN related sequences submitted to the NCBI GenBank database, under accession numbers KY508373–KY508381. On the basis of the above results, we selected two strains (CY4 and CN11) for the study of substrate richness using different Biolog plates (carbon, nitrogen, and osmolytes), and for study of the plant-microbe interaction mechanism by means of GFP.

Genotypic Diversity

In the present study, the genetic diversity fingerprints of all the selected strains isolated from sugarcane rhizosphere were investigated through BOX, ERIC, and REP-PCR fingerprints. A number of polymorphic bands ranging between 50 bp and about 5 kb were observed, and the DNA fingerprints were clearly differentiated from each other. Nearly all the selected strains showed high-quality DNA fingerprint profiles generated with each primer set (**Figure 3**). The fingerprint patterns of the thirty isolates generated by BOX, ERIC, and REP-PCR were complex, producing a large number of polymorphic bands of variable intensity. Differences among strains were assessed visually, on the basis of the banding patterns of PCR products. BOX PCR generated a total of 268 bands ranging from 150 bp to 4 kb for all the 30 selected *Pseudomonas* strains studied. The maximum numbers of bands (12) were observed for CN1 and CN20 while the lowest (5) was found in the case of BN7 (**Figure 3A**). A total of 229 bands were identified by ERIC-PCR for all the *Pseudomonas* strains, and the band sizes ranged from 50 bp to 4 kb. CoY1, CoA5, BY10, BA17, CA5, and CN15 showed the highest number of bands (11), while AN15 showed the lowest number of bands (3) (**Figure 3B**). For REP-PCR, 210 bands were identified, of approximately 50 bp to 5 kb, and faint bands were also frequently observed. The highest number of bands (12) was observed for AY1, while the lowest (4) was found in the case of BN7 and CN11 (**Figure 3C**). In the case of all these PCRs, BOX-PCR fingerprints revealed a high genotypic diversity for all the *Pseudomonas* strains.

To determine the relatedness of the isolates, a dendrogram was reconstructed based on BOX, ERIC, and REP-PCR fingerprint bands (**Figure 4**), and the data were analyzed by using Jaccard similarity coefficients and the neighbor-joining clustering method based on pair-wise similarity coefficients with UPGMA. The dendrogram generated through BOX PCR fingerprints revealed two major clusters, one containing five strains and the other containing 25 strains (**Figure 4A**). The clustering was clear, and represented 28 distinct clusters of *Pseudomonas* spp. In the ERIC-PCR, all 30 *Pseudomonas* spp. showed two major clusters containing 20 and 10 strains (**Figure 4B**), all grouped into different clusters. As with BOX and ERIC-PCR, the dendrogram generated based on the REP-PCR fingerprints also showed two major clusters, containing one and 29 strains (**Figure 4C**) respectively, and all 30 strains were grouped into 28 different clusters of *Pseudomonas* spp.

Substrate Utilization Patterns

Selected strains were tested for metabolic potential with several compounds as sole sources for carbon (C), using GNIII and nitrogen (N) PM3B micro-plates. Strain tolerance of osmotic stress was examined with PM9 micro-plates, and metabolic activity over a wide range of pH, 3.5–10, was determined by using PM10 micro-plates. The Biolog system was used to detect the biochemical, physiological, and chemical sensitivity of strains on the basis of substrate richness (Table S3). The results indicated that strain CY4, followed by CN11, showed the highest utilization of carbon sources, i.e., sugars (28.17%), carboxylic acids (18.30%), hexose acids (12.68%), and amino acids (9.86%), and the highest chemical sensitivity (86.95%) (**Figure 5**).

A qualitative analysis of metabolic differentiation between the selected strains was performed through principal component analysis (PCA), using PM3B, PM9, and PM10 (**Figure 6** and Table S4). The main principle for the grouping of every substrate group to the individual component was the relationship between different metabolites and strain utilization. For these studies, separate evaluations of nitrogen, osmolytes, and pH were performed. A scatter plots diagram of PCA showed the nitrogen of different components accounted for 86.48% of the metabolic variation in PC1 for the selected strains (**Figure 6A**). The osmolytes accounted for 62.57% of the variance and the pH for 82.89% of the variance observed in PC1 (**Figures 6B,C**). Based on the data obtained through the Biolog analysis, the diversity index was measured and was evident for the selected strains, which showed high metabolic diversity (**Table 5**).

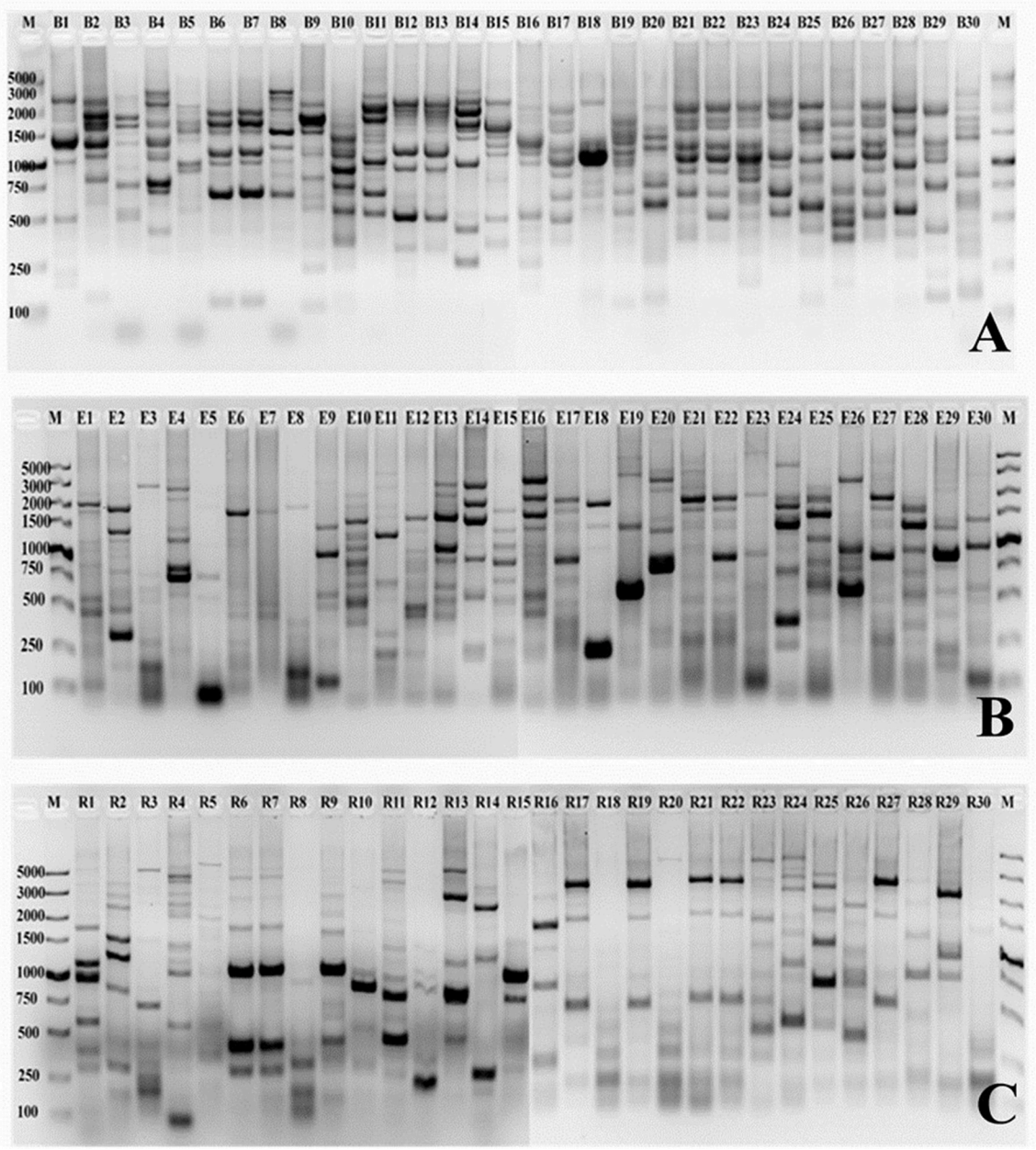


FIGURE 3 | PCR finger printing patterns from genomic DNA of *Pseudomonas* strains isolated from sugarcane. The BOX, ERIC and REP patterns are shown in **(A–C)**, respectively. B1-B30 (BOX), E1-E30 (ERIC), and R1-R30 (REP) is a strain codes used in this study. A 12–15 μL of each product is loaded onto a 1.5% agarose gel. M, molecular size marker (100 bp–5 kb), low range DNA ruler (Takara). Strain codes: 1. CoY1, 2. CoA5, 3. CoA6, 4. AY1, 5. AY2, 6. AA8, 7. AN15, 8. BY2, 9. BY3, 10. BY10, 11. BA2, 12. BA4, 13. BA7, 14. BA12, 15. BA13, 16. BA17, 17. BN6, 18. BN7, 19. CY4, 20. CY7, 21. CA5, 22. CA7, 23. CN1, 24. CN2, 25. CN3, 26. CN9, 27. CN11, 28. CN12, 29. CN15, and 30. CN20.

Colonization of Sugarcane Plants by GFP-Tagged Bacteria

Visualization of root colonization was carried out by CLSEM. This technique facilitated study of the interaction of the selected potential strains in sugarcane. Bacterial colonization in the internal tissues of plants was observed in almost all parts of the sugarcane plant. After 72 h of incubation with the inoculated strains (CY4 and CN11), it was observed that the bacterial cell density had increased, and green fluorescent cells were distributed throughout the plant organs, including roots, stems, and leaves (**Figure 7**). No fluorescent bacteria were detected by CLSEM in the uninoculated plants served as a control (**Figures 7A–C**). In roots, the GFP-tagged bacteria were observed by CLSEM to colonize mostly mature root hair zones. Many

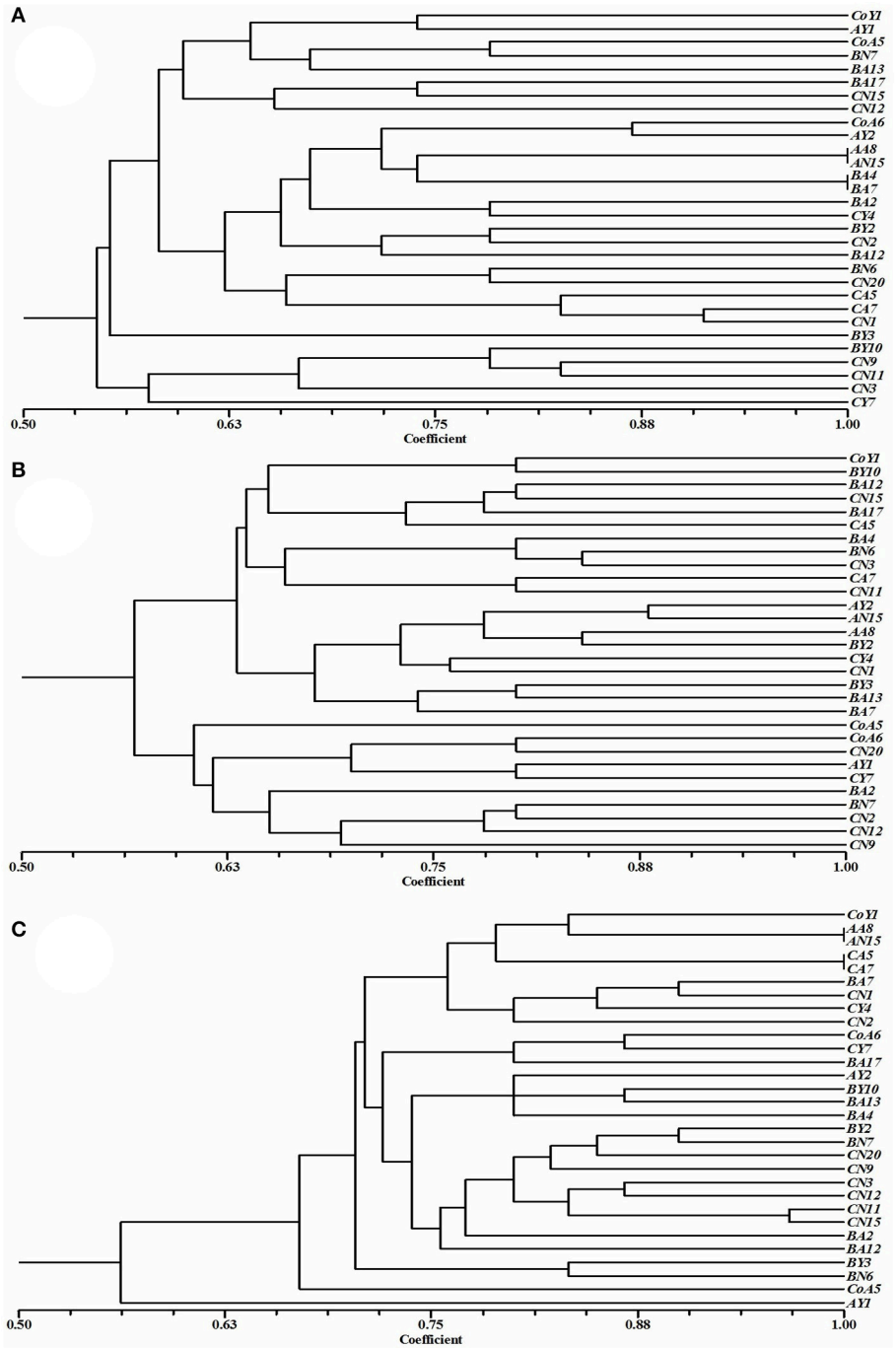


FIGURE 4 | Cluster analyses of **(A)** BOX, **(B)** ERIC, and **(C)** REP-PCR fingerprints showing the genotypic diversity of *Pseudomonas* species isolated from sugarcane. Dendrogram was obtained from the similarity coefficient calculations and clustering was done using unweighted pair-grouping method based on arithmetic averages (UPGMA) using NTSYS software and the Jaccard coefficient.

of the cells were over shadowed by the green fluorescence light emitted from the cell walls of the epidermal cells, endodermal, xylem vessels, and junction sites between the primary and lateral root zones, which were excited by the blue light of the fluorescence microscope (**Figures 7E,G**). In the case of leaves, colonization of GFP-tagged bacteria was clearly observed through the green auto-fluorescence emitted as small dots in all plant parts (**Figures 7D,F**), in the case of roots, the cells emitted fluorescence in the maturation zone and within the body of the root, detected by CLSEM in transvers hand-cut optical sections.

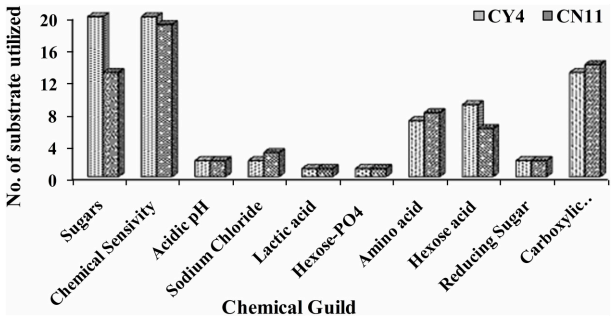


FIGURE 5 | Metabolic differences among CY4 and CN11 in presence of sugars, chemical sensivity, acidic pH, sodium chloride, lactic acid, hexose PO_4, amino acid, hexose acid, reducing sugar, and carboxylic acid using BIOLOG Phenotype Micro-Array™ plates GNIII.

nifH Gene Expression Determined by qRT-PCR

The *nifH* gene expression pattern for two sugarcane varieties (GT11 and GXB9) inoculated with CY4 and CN11 strains at 90 and 120 days was studied (**Figure 8**). The results showed that, using total RNA extracted from sugarcane leaf samples, *nifH* gene expression was positively detected by means of qRT-PCR. The highest expression of the *nifH* gene was recorded at 90 days, in GXB9 inoculated with strain CN11.

DISCUSSION

In this study, soil samples from sugarcane were used for the isolation of PGP nitrogen-fixing bacteria. Guangxi is the major sugarcane producing province in South China, and more than 60% of the total sugar production is from this area (Li and Yang, 2015). In China, the nitrogen application rate is very high, that is about 500–700 kg ha^{-1} annually for commercial sugarcane production and this is several times higher than in Brazil and other countries (Li et al., 2015). Higher application of nitrogen fertilizers not only raises the production cost, but also has an adverse effect on the environment and on soil health. Root-associated microbes have a positive effect on soil nutrient availability for plants (Glick et al., 1994). PGPR increases root surface area, increasing nutrient uptake and improving plant production (Mantelin and Touraine, 2004). Therefore, PGPR is an alternative method, the use of chemicals, for sugarcane nutrition because this crop requires large quantities of nitrogen fertilizer for growth and development. A number of nitrogen-fixing microbes have been reported in sugarcane (Xing et al., 2006, 2015; Mehnaz et al., 2009a; Lin et al., 2012; Solanki et al., 2017). *Pseudomonas* spp. are important PGPRs used as biofertilizer, and are able to enhance crop yield by direct and indirect mechanisms (Walsh et al., 2001), in addition to their characteristics of antibiotic production, phosphate solubilization, siderophore, IAA, HCN, and ammonia production (Ahemad and Khan, 2012a,b), nitrogen fixation, ACC deaminase activity, plant hormone production, and biological control. Of all the strains isolated in the present study, only 30 were selected on the basis of various different PGP traits, and nitrogenase activity.

Phosphorus is one of the most important plant nutrients and it greatly affects the growth of plants (Wang et al., 2009). In soil, P is highly insoluble and is therefore unavailable to plants. Phosphate-solubilizing microorganisms increase the performance of plants by providing them with soluble phosphorus. Phosphate solubilization has already been reported for different strains of *Pseudomonas frederiks-bergensis* (Zeng et al., 2016). In this study, we found that 26 (87%) isolates were positive for phosphate solubilization to convert insoluble tricalcium phosphate to the soluble form. Another important PGPR trait, which may indirectly influences plant growth, is the production of siderophores. Microorganisms play an important role in several essential biological processes and developed specific mechanisms for the assimilation of iron by production of low molecular weight iron-chelating compounds siderophores, which transport this element into their cells (Schwyn and Neilands, 1987; Arora et al., 2013). In the present investigation, 20 (66.67%) strains displayed siderophores production ability; siderophore production by *Pseudomonas* spp. is well known (Liu et al., 2011; Luo, 2014). HCN production by bacterial strains is important for diseases suppression and protection of plants from fungal diseases. HCN production by *P. fluorescens* strain CHA0 was related to biocontrol ability and root colonization; for example, suppression of tobacco black root rot caused by *Thielaviopsis basicola* (Voisard et al., 1989; Laville et al., 1992). The results of qualitative test of HCN production showed that 43.33% of the bacteria were capable of producing HCN. Volatile compounds, such as ammonia produced by a number of rhizobacteria, have been reported to play an important role in biocontrol (Brimecombe et al., 2001). Our results showed that 40.0% of the *Pseudomonas* strains were able to produce ammonia and might therefore play an important role in biocontrol activity. A previous isolate, *P. fluorescens* BAM-4, from semi-arid soil, is a potential biocontrol agent against *M. phaseolina* I mung bean and *P. aeruginosa* RM-3 showed lysis of several pathogenic fungi (Minaxi and Saxena, 2010a,b). To test strains for antifungal activity, we screened against two sugarcane pathogens (*U. scitaminea* and *C. paradoxa*); against these antagonistic activity was observed in 46.66 and 33.33% of strains, respectively (**Table 2**).

The role of ACC deaminase has been documented as one of the major mechanisms of PGP bacteria in promoting root and plant growth (Glick et al., 2007). Inoculation with ACC-deaminase containing bacteria promotes root growth of developing seedlings of various crops (Zahir et al., 2003). Inoculation with rhizobacteria having ACC deaminase activity resulted in the development of a better root system, which subsequently had a positive effect on shoot growth (Glick et al., 1998; Belimov et al., 2002). The screening of bacterial isolates obtained from sugarcane with DF medium and DF-ACC medium showed that 60% of the bacteria contained ACC deaminase, when grown on medium containing ACC as the sole nitrogen source. Most ACC-utilizing bacterial isolates are known to belong to genus *Burkholderia* and genus *Pseudomonas* (Blaha et al., 2006; Onofre-Lemus et al., 2009). All the selected bacterial

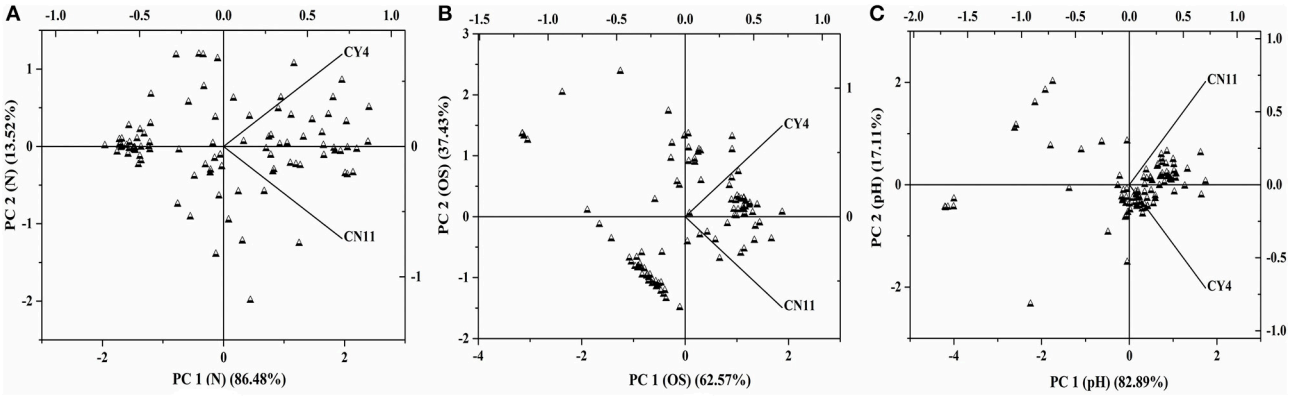


FIGURE 6 | The scatterplots of selected two strains *Pseudomonas koreensis* (CY4) and *Pseudomonas entomophila* (CN11) were analyzed through principle component analysis (PCA) under the different treatments by using the BIOLOG(R) micro-plates **(A)** nitrogen, **(B)** osmolytes, and **(C)** pH.

strains showed ACC deaminase activity ranging from 77.0 to 15.13 μmoL mg^{-1} h^{-1}, and higher activity was observed in strain CY4. In this study, we found that all bacterial strains were able to produce IAA in the range of 312.07–13.12 μg mL^{-1}. The potential of the bacterial strains to produce IAA is indicative of their capability to be used as growth hormones or growth regulators. The results for nitrogen fixation as determined by ARA indicated that a large population of sugarcane-associated nitrogen-fixing bacterial strains is present in the soil and may be beneficial in improving the nitrogen level of sugarcane. The selected isolates were evaluated for their nitrogen-fixation efficiency, which was highly variable, ranging from 108.30 to 6.16 μmoL C_2H_2 h^{-1} mL^{-1}. The incidence of nitrogen fixation in *Pseudomonas* spp. has been long debated, but recently several such *Pseudomonas* strains have been identified (Mirza et al., 2006). Several studies have also shown that nitrogenase producing bacteria can be isolated from sugarcane (Xing et al., 2006; Mehnaz et al., 2010; Lin et al., 2012; Solanki et al., 2017).

In this paper, we have also described a molecular approach for analyzing nitrogen-fixing genes in pure cultures isolated from sugarcane. Several primer have been used to amplify *nifH* genes, but after comparison of the sequences obtained from the *nifH* clones, we found the primer sets were is suitable for this study. All strains showed nitrogen-fixing activity by ARA in N-free medium, but the *nifH* gene was amplified only in 10 strains. *nifH* is one of the earliest characterized and best known functional genes (Rosado et al., 1998), and its amplification using degenerate primers is a useful tool for confirming nitrogen-fixation potential (Zehr and Capone, 1996). On the other hand, if amplification does not occur using the primers, this is not proof that strains are not proficient in nitrogen fixation, because the gene may show diverse nucleotide sequences between species and even within the same species (Zehr et al., 2003). An important objective of this study was to evaluate all the strains for antibiotic biosynthetic gene targets, because the production of various antimicrobial compounds is one of the biocontrol that helps to degrade pathogen cell walls and it is also an important factor for disease suppression (Sasirekha et al., 2013).

TABLE 5 | Substrate richness diversity indices calculated for nitrogen, osmolytes and pH through Biolog for selected strains *Pseudomonas koreensis* (CY4) and *Pseudomonas entomophila* (CN11).

Index measure	Diversity indices through Biolog Micro-Plates					
	Strain CY4			Strain CN11		
	Nitrogen	Osmolytes	pH	Nitrogen	Osmolytes	pH
	PM3B	PM9	PM10	PM3B	PM9	PM10
Dominance D	0.01507	0.01634	0.01178	0.01488	0.01137	0.01131
Simpson 1-D	0.9849	0.9837	0.9882	0.9851	0.9886	0.9887
Shannon H	4.327	4.193	4.46	4.33	4.499	4.501
Evenness e^H/S	0.7891	0.6899	0.9006	0.7912	0.9368	0.9385
Brillouin	1.357	2.177	2.729	1.311	2.072	2.522
Menhinick	10.71	8.977	7.258	11.29	7.911	7.339
Margalef	25.58	21.56	19.31	26.51	20.86	19.64
Equitability J	0.9481	0.9187	0.9771	0.9487	0.9857	0.9861
Fisher alpha	0	282	87.25	0	119.6	90.32
Berger Parker	0.01244	0.01749	0.01143	0.01384	0.01358	0.01169

In the present study, we have demonstrated that repetitive extragenic sequences such as BOX, ERIC and REP are present in the genome of *Pseudomonas* bacteria. There little information published regarding the use of PCR to study the genetic diversity of *Pseudomonas* isolated from sugarcane. A method that was described for bacterial fingerprinting by examining strain specific banding patterns obtained from PCR amplification of repetitive DNA elements presented entire bacterial genomes (Versalovic et al., 1991). This technique is useful in the classification and differentiation of strains in many Gram-positive and Gram-negative bacteria, and also proved to be a powerful tool for initially screening within the strains. REP patterns are generally less complex than BOX and ERIC patterns, but all could give good discrimination at the strain level for the strains isolated from sugarcane (**Figure 3**). Our results also support a previous study in *P. aeruginosa* (Dawson et al., 2002), and the technique proved to be

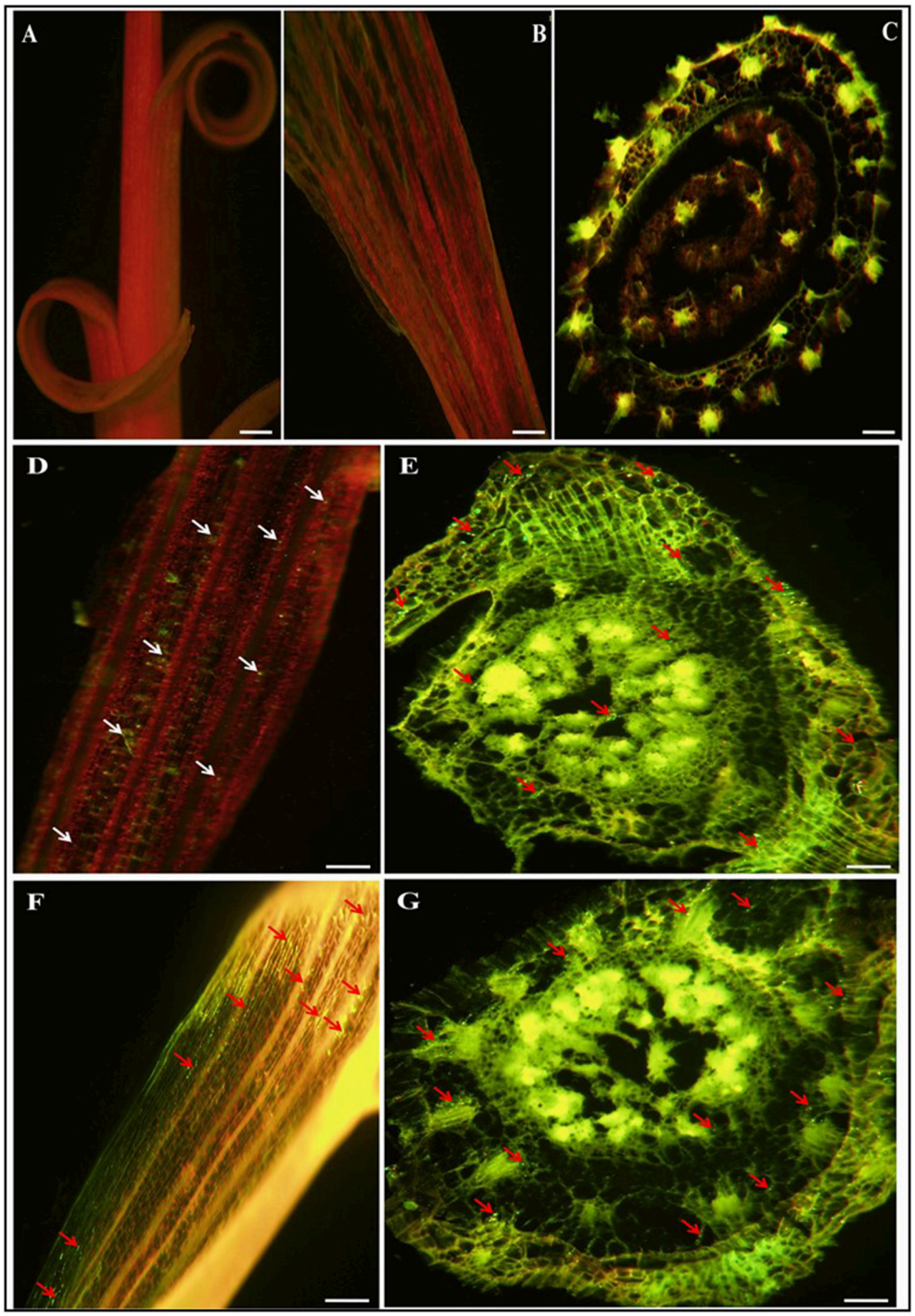


FIGURE 7 | Confocal laser scanning micrographs images showing gfp-tagged strains (CY4 and CN11) colonized in and on roots and leaves of sugarcane micropropagated plantlets GT11 (variety). **(A–C)** is control sugarcane plantlets parts i.e., stem, leaf and root, without inoculated strains. Confocal microscopic images **(D–G)** present inoculated bacterial GFP fluorescence (500–530 nm) in green dots and auto-fluorescence in everywhere in leaf and root. Arrow heads point indicates bacterial cells present in a single or grouped of bacteria. **(D,E)** represents CY4 and **(F,G)** is CN11 strain. Bars present 50 μm.

effective for evaluating the diversity of nitrogen-fixing bacterial strains.

Of all the strains, only two CY4 and CN11, were selected for further studies, such as Biolog profiling, GFP localization, and *nifH* gene expression, through qRT-PCR in sugarcane. Metabolic profiling of the selected isolates was performed using Biolog microplates, comprising analyses of the utilization of nutritional compounds (carbon and nitrogen), as well

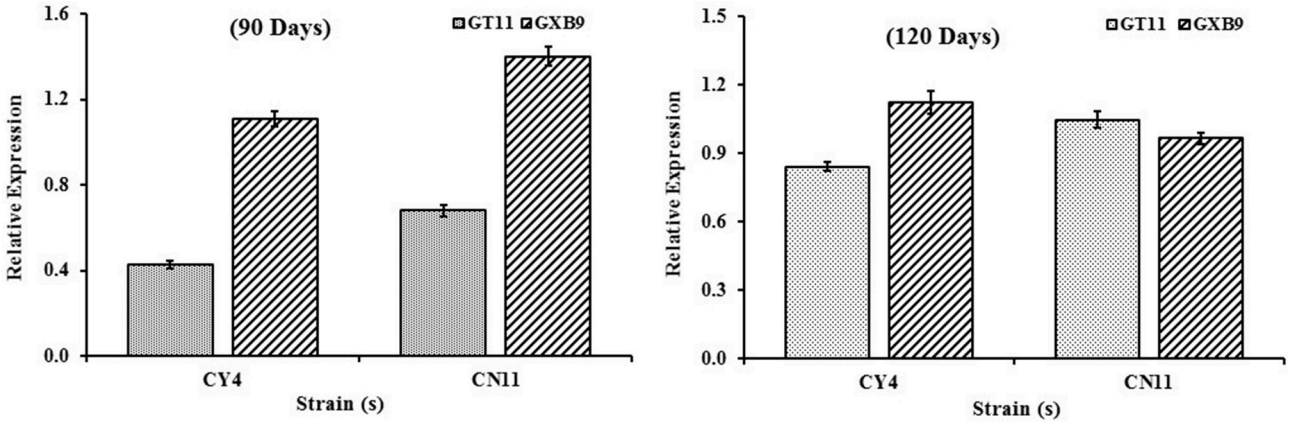


FIGURE 8 | qRT-PCR analysis of the *nifH* expression patterns in sugarcane varieties (GT11 and GXB9) and potent strains (*Pseudomonas koreensis*-CY4 and *Pseudomonas entomophila*-CN11) interaction. Data were normalized to the GAPDH expression level. All data points (with the deduction of their controls) are the means ± SE ($n = 3$).

as tolerance to osmolytes and different pH conditions. The metabolic assets of an organism could contribute toward a particular adaptation and therefore might provide valuable information about bacteria supportive for root colonization (Mazur et al., 2013). The patterns of phenotypes obtained were suggestive that the selected strains were capable of utilizing a variety of metabolic substrates (**Figure 5**). Previously, it was suggested that more metabolically useful strains were more successful competitors in host plant nodulation (Wielbo et al., 2007). Remarkably, in our studies, the more metabolically diverse strain was CY4, rather than CN11. Through metabolic profiling, tolerance to osmolytes and different pH conditions was studied, using diversity indices such as Simpson 1-D, Shannon H, Evenness e^H/S, Brillouin and Equitability J, and these methods revealed the substrate richness of the selected strains. Phenotypic profiling is important for understanding genotype differences, stress responses, media composition, and changes in environmental conditions for microorganisms (Chojniak et al., 2015).

We also investigated the effect of the selected *Pseudomonas* strains CY4 and CN11 on sugarcane by genetically tagging them with GFP-pPROBE-pTet^r-OT and observing the level of colonization in plantlets. Both *Pseudomonas* strains colonized the whole plantlets as for roots, stems and leaves when inoculated separately. Uninoculated sugarcane plantlets served as the control and the plantlets did not show any appearance of fluorescence after 3 and 5 days of inoculation. This result showed that there was no fluorescence from whole plantlet tissues, but after comparison of strains CY4 and CN11, appeared that strain CY4 was a better colonizer for sugarcane plantlets. CLSEM and GFP have previously been used to demonstrate the colonization pattern of *Bacillus megaterium* in rice (Liu et al., 2006), *Klebsiella pneumoniae* in maize (Chelius and Triplett, 2000), *Rhizobium* sp. and *Burkholderia* sp. in rice (Singh et al., 2009), and *Microbacterium* sp. in sugarcane (Lin et al., 2012). This study shows that the GFP technique can be used effectively to evaluate competitive colonization capability for more than one plant-associated bacterium (Singh et al., 2009), or to select a potent strain isolated from different crops.

In, previous studies, it was considered that there were no nitrogen-fixing strains in the genus *Pseudomonas* (Setten et al., 2013). In fact, the inability of *Pseudomonas* spp. to fix nitrogen had been proposed as an important taxonomic character (Young, 1992; Anzai et al., 2000; Solanki et al., 2017). However, recent studies have confirmed that some strains belonging to the genus *Pseudomonas sensu stricto*, such as *P. stutzeri* A1501, *P. stutzeri* DSM4166, *P. azotifigens* 6HT33b^T and *Pseudomonas* sp. K1, have the capability to fix nitrogen (Mehnaz, 2011; Setten et al., 2013; Solanki et al., 2017). In our study, the selected *Pseudomonas* strains CY4 and CN11 also showed *nifH* gene expression in sugarcane, detected through qRT-PCR. Therefore, the presence of the *nifH* gene is indicative of the existence of diazotrophs, and the expression of the *nifH* gene is suggestive of the occurrence of BNF (Akter et al., 2014). The qRT-PCR technique is advantageous because of its high sensitivity and specificity, and detection of mRNA has been reported in ecological samples (Noda et al., 1999; Brown et al., 2003).

CONCLUSION

To the best of our knowledge, this study is the first systematic report that has provided an awareness of the bacterial genus *Pseudomonas* associated with sugarcane rhizosphere. We have isolated bacteria that show useful activities in phosphate solubilization, siderophore production, ACC deaminase activity, and IAA-production, as well as N_2-fixing activity and disease management. These features are measured as important PGP traits and have been found to be effective in improving the growth and nitrogen content of sugarcane plants. The strains

CY4 (*P. koreensis*) and CN11 (*P. entomophila*) turned out to be very efficient in terms of enhancing the growth and development of plants, and disease control, as well as having nitrogenase activity. These organisms have greater potential to be used as biofertilizer due to their properties of nitrogen fixation, phytohormone production, and biocontrol capability. Assessment of these strains in field trials is now required, to determine their efficiency in plant growth promotion under field conditions. The inoculation of PGPR isolates may be an imminent development biofertilizer applications, for sustainable crop production, in reducing environmental pollution, and in biological agri-business.

AUTHOR CONTRIBUTIONS

HL, RS, LY, and YL conceived and proposed the idea and drafted the manuscript. HL, RS, PS, QS, and YX carried out the experiments and conducted data analysis. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

This present study was supported by the National Natural Science Foundation of China (31471449, 31171504, 31101122), the National High Technology Research and Development Program ("63" Program) of China (2013AA102604), Guangxi Special Funds for Bagui Scholars and Distinguished Experts (2013), Guangxi Natural Science Foundation (2011GXNSFF018002, 2012GXNSFDA053011, 2013NXNSFAA019073) and Guangxi Academy of Agriculture Sciences Fund (GNK2014YD01, GNKB2014021). Fund for Guangxi Innovation Teams of Modern Agriculture Technology (gjnytxgxcxtd-03-01) and Fund of Guangxi Academy of Agricultural Sciences (2015YT02).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01268/full#supplementary-material

REFERENCES

Ahemad, M., and Khan, M. S. (2012a). Effect of fungicides on plant growth promoting activities of phosphate solubilizing Pseudomonas putida isolated from mustard (*Brassica compestris*) rhizosphere. *Chemosphere.* 86, 945–950. doi: 10.1016/j.chemosphere.2011.11.013

Ahemad, M., and Khan, M. S. (2012b). Evaluation of plant growth promoting activities of rhizobacterium Pseudomonas putida under herbicide-stress. *Ann. Microbiol.* 62, 1531–1540. doi: 10.1007/s13213-011-0407-2

Akter, Z., Pageni, B. B., Lupwayi, N. Z., and Balasubramanian, P. M. (2014). Biological nitrogen fixation and nifH gene expression in dry beans (*Phaseolus vulgaris* L.). *Can. J. Plant Sci.* 94, 203–212. doi: 10.4141/cjps2013-200

Antwerpen, T. V., Rutherford, R. S., and Vogel, J. L. (2002). Assessment of sugarcane endophytic bacteria and rhizospheric Burkholderia species as antifungal agents. *Proc. S. Afr. Sug. Technol. Assoc.* 76, 301–304.

Anzai, Y., Kim, H., Park, J. Y., Wakabayashi, H., and Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int. J. Syst. Evol. Microbiol.* 4, 1563–1589. doi: 10.1099/00207713-50-4-1563

Arora, N. K., Tewari, S., and Singh, R. (2013). "Multifaceted plant-associated microbes and their mechanisms diminish the concept of direct and indirect PGPRs," in *Plant Microbe Symbiosis; Fundamentals and Advances,* ed N. K. Arora (Lucknow: Springer), 411–449.

Belimov, A. A., Safronova, V. I., and Mimura, T. (2002). Response of spring rape (*Brassica napus* L. var. Oleifera) to inoculation with plant growth promoting rhizobacteria containing 1-aminocyclopropane-1-carboxylate deaminase depends on nutrient status of the plant. *Can. J. Microbiol.* 48, 189–199. doi: 10.1139/w02-007

Bhattacharyya, P. N., and Jha, D. K. (2012). Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. *World. J. Microbiol. Biotechnol.* 28, 1327–1350. doi: 10.1007/s11274-011-0979-9

Blaha, D., Prigent-Combaret, C., Mirza, M. S., and Moenne-Loccoz, Y. (2006). Phylogeny of the 1-amino cyclopropane-1-carboxylic acid deaminase encoding gene acdS in phytobeneficial and pathogenic Proteobacteria and relation with strain biogeography. *FEMS Microbiol. Ecol.* 56, 455–470. doi: 10.1111/j.1574-6941.2006.00082.x

Bochner, B. R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiol Rev.* 33, 191–205. doi: 10.1111/j.1574-6976.2008.00149.x

Boddey, R. M., de Oliveira, O. C., Urquiaga, S., Reis, V. M., de Olivares, F. L., Baldani, V. L., et al. (1995). Biological nitrogen fixation associated with sugar cane and rice: contributions and prospects for improvement. *Plant Soil.* 174, 195–209. doi: 10.1007/BF00032247

Brick, J. M., Bostock, R. M., and Silverstone, S. E. (1991). Rapid in situ assay for indole acetic acid production by bacteria immobilized on nitrocellulose membrane. *Appl. Environ. Microbiol.* 57, 535–538.

Brimecombe, M. J., De Liej, F. A., and Lynch, J. M. (2001). "The effect of root exudates on rhizosphere microbial populations," in *The Rhizosphere,* eds R. Pinton, Z. Varanini, and P. Nannipieri (New York, NY: Marcel Dekker), 95–140.

Brown, M. M., Friez, M. J., and Lovell, C. R. (2003). Expression of nifH genes by diazotrophic bacteria in the rhizosphere of short form *Spartina alterniflora*. *FEMS Microbiol. Ecol.* 43, 411–417. doi: 10.1111/j.1574-6941.2003.tb01081.x

Carvalho, T. L. G., Balsemao-Pires, E., Saraiva, R. M., Ferreira, S. G., and Hemerly, A. S. (2014). Nitrogen signalling in plant interactions with associative and endophytic diazotrophic bacteria. *J. Exp. Bot.* 65, 5631–5642. doi: 10.1093/jxb/eru319

Chelius, M. K., and Triplett, E. W. (2000). Immuno localization of dinitrogenase reductase produced by *Klebsiella pneumoniae* in association with *Zea mays* L. *Appl. Environ. Microbiol.* 66, 783–787. doi: 10.1128/AEM.66.2.783-787.2000

Chi, F., Shen, S. H., Chen, S. F., and Jing, Y. X. (2004). Migration of *Azospirillum brasilense* Yu62 from root to stem and leaves inside rice and tobacco plants. *Acta. Bot. Sin.* 46, 1065–1070.

Chojniak, J., Wasilkowski, D., Płaza, G., Mrozik, A., and Brigmon, R. (2015). Application of Biolog microarrays techniques for characterization of functional diversity of microbial community in phenolic-contaminated water. *Int. J. Environ. Res.* 9, 785–794.

Croft, B. J., and Magarey, R. C. (2000). "Pachymetra root rot," in *A Guide to Sugarcane Diseases,* eds P. Rott, R. A. Bailey, J. C. Comstock, B. J. Croft, and A. S. Saumtally (Montpellier: CIRAD/ISSCT, CIRAD Publication Services), 126–130.

Dawson, S. L., Fry, J. C., and Dancer, B. N. (2002). A comparative evaluation of five typing techniques for determining the diversity of fluorescent pseudomonads. *J. Microbiol. Methods* 50, 9–22. doi: 10.1016/S0167-7012(02)00003-9

De Souza, R., Ambrosini, A., and Passaglia, L. M. P. (2015). Plant growth-promoting bacteria as inoculants in agricultural soils. *Genet. Mol. Biol.* 38, 401–419. doi: 10.1590/S1415-475738420150053

Dey, R., Pal, K. K., Bhatt, D. M., and Chauhan, S. M. (2004). Growth promotion and yield enhancement of peanut (*Arachis hypogaea* L.) by application of plant growth promoting rhizobacteria. *Microbiol. Res.* 159, 371–394. doi: 10.1016/j.micres.2004.08.004

Edwards, U., Rogall, T., Blöcker, H., Emde, M., and Böttger, E. C. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* 17, 7843–7853.

FAO (2016). *Top Sugarcane Production: Food and Agriculture Organization of the United Nations*. Rome: FAO.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Fischer, D., Pfitzner, B., Schmid, M., Simoes-Araujo, J. L., Reis, V. M., Pereira, W., et al. (2012). Molecular characterisation of the diazotrophic bacterial community in uninoculated and inoculated field-grown sugarcane (*Saccharum* sp.). *Plant Soil* 356, 83–99. doi: 10.1007/s11104-011-0812-0

Glick, B. R., Jacobson, C. B., and Pasternak, J. J. (1994). Partial purification and characterization of ACC deaminase from the plant growth-promoting rhizobacterium *Pseudomonas putida* GR12-2. *Can. J. Microbiol.* 40, 1019–1025. doi: 10.1139/m94-146

Glick, B. R., Penrose, D. M., and Li, J. (1998). A model for the lowering of plant ethylene concentration by plant growth-promoting bacteria. *J. Theor. Biol.* 190, 63–68. doi: 10.1006/jtbi.1997.0532

Glick, B. R., Todorovic, B., Czarny, J., Cheng, Z., Duan, J., and McConkey, B. (2007). Promotion of plant growth by bacterial ACC deaminase. *Crit. Rev. Plant. Sci.* 26, 227–242. doi: 10.1080/07352680701572966

Glickmann, E., and Dessaux, Y. (1995). A critical examination of the specificity of the Salkowski reagent for indolic compounds produced by phytopathogenic bacteria. *Appl. Environ. Microbiol.* 61, 793–796.

Hardy, R. W. F., Holsten, R. D., Jackson, E. K., and Burns, R. C. (1968). The acetylene ethylene assay for N2 fixation: laboratory and field evaluation. *Plant Physiol.* 43, 1185–1207. doi: 10.1104/pp.43.8.1185

Herridge, D. F., Peoples, M. B., and Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 3, 1–18. doi: 10.1007/s11104-008-9668-3

Honma, M., and Shimomura, T. (1978). Metabolism of 1-aminocyclopropane-1-carboxylic acid. *Agric. Biol. Chem.* 42, 1825–1831.

Jacobson, C. B., Pasternak, J. J., and Glick, B. R. (1994). Partial purification and characterization of 1-aminocyclopropane-1-carboxylate deaminase from the plant growth promoting rhizobacterium *Pseudomonas putida* GR12-2. *Can. J. Microbiol.* 40, 1019–1025. doi: 10.1139/m94-162

James, E. K. (2000). Nitrogen fixation in endophytic and associative symbiosis. *F. Crop. Res.* 65, 197–209. doi: 10.1016/S0378-4290(99)00087-8

James, E. K., and Olivares, F. L. (1998). Infection and colonization of sugarcane and other graminaceous plants by endophytic diazotrophs. *Crit. Rev. Plant Sci.* 50, 77–119. doi: 10.1016/S0735-2689(98)00357-8

Juraeva, D., George, E., Davranov, K., and Ruppel, S. (2006). Detection and quantification of the nifH gene in shoot and root of cucumber plants. *Can. J. Microbiol.* 52, 731–739. doi: 10.1139/w06-025

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Laville, J., Voisard, C., Keel, C., Maurhofer, M., Defago, G., and Haas, D. (1992). Global control in *Pseudomonas fluorescens* mediating antibiotic synthesis and suppression of black root rot of tobacco. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1562–1566. doi: 10.1073/pnas.89.5.1562

Li, R. P., and Macrae, I. C. (1991). Specific association of diazotrophic Acetobacers with sugarcane. *Soil Biol. Biochem.* 23, 999–1002. doi: 10.1016/0038-0717(91)90181-I

Li, Y. R., Zhou, X. Z., and Yang, L. T. (2015). Biological nitrogen fixation in sugarcane and nitrogen transfer from sugarcane to cassava in an intercropping system. *Inter. J. Sci. Nature.* 6, 214–218.

Li, Y. R., and Yang, L. T. (2015). Sugarcane agriculture and sugar industry in China. *Sugar Tech.* 17, 1–8. doi: 10.1007/s12355-014-0342-1

Lima, E., Boddey, R. M., and Dobereiner, J. (1987). Quantification of biological nitrogen fixation associated with sugar cane using a 15N-aided nitrogen balance. *Soil Biol. Biochem.* 19, 165–170. doi: 10.1016/0038-0717(87)90077-0

Lin, L., Guo, W., Xing, Y., Zhang, X., Li, Z, Hu, C., et al. (2012). The actinobacterium *Microbacterium* sp. 16SH accepts pBBR1-based pPROBE vectors, forms biofilms, invades roots, and fixes N2 associate d with micropropagated sugarcane plants. *Appl. Microbiol. Biotechnol.* 93, 1185–1195. doi: 10.1007/s00253-011-3618-3

Liu, X., Zhao, H., and Chen, S. (2006). Colonization of maize and rice plants by strain Bacillus megaterium C4. *Curr. Microbiol.* 52, 186–190. doi: 10.1007/s00284-005-0162-3

Liu, Y. P., Teng, S. S., and Zhao, L. (2011). Identification of a siderophore-producing bacterium Pseudomonas putida A3 and its growth- promoting effects on cucumber seedlings. *J. Plant Nutrit.* Fertil. *Sci.* 17, 1507–1514.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lorck, H. (1948). Production of hydrocyanic acid by bacteria. *Physiol. Plant* 1, 142–146. doi: 10.1111/j.1399-3054.1948.tb07118.x

Luo, L. J. (2014). *Isolation and Identification of Endophytic Bacteria in Guangxi Sugarcane Wild Germplasms and Endogenous Rhizobia in Sugarcane Intercropping Beans.* Master Thesis, Guangxi University.

Magnani, G. S., Didonet, C. M., Cruz, L. M., Picheth, C. F., Pedrosa, F. O., and Souza, E. M. (2010). Diversity of endophytic bacteria in Brazilian sugarcane. *Genet. Mol. Res.* 9, 250–258. doi: 10.4238/vol9-1gmr703

Manoharachary, C., and Mukerji, K. G. (2006). "Microbial activity in the Rhizosphere," in *Soil Biology*, eds K. G. Mukerji, C. Manoharachary, and J. Singh (Berlin; Heidelberg: Springer-Verlag), 7.

Mantelin, S., and Touraine, B. (2004). Plant growth-promoting bacteria and nitrate availability impacts on root development and nitrate uptake. *J. Exp. Bot.* 55, 27–34. doi: 10.1093/jxb/erh010

Mazur, A., Stasiak, G., Wielbo, J., Koper, P., Kubik-Komar, A., and Skorupsk, A. (2013). Phenotype profiling of *Rhizobium leguminosarum* bv. trifolii clover nodule isolates reveal their both versatile and specialized metabolic capabilities. *Arch. Microbiol.* 195, 255–267. doi: 10.1007/s00203-013-0874-x

Mehnaz, S. (2011). "Plant growth-promoting bacteria associated with sugarcane" in *Bacteria in Agrobiology*, ed D. K. Maheshwari (Berlin; Heidelberg: Crop Ecosystems, Springer-Verlag), 165–187.

Mehnaz, S., Baig, D. N., Jamil, F., Weselowski, B., and Lazarovits, G. (2009a). Characterization of a phenazine and hexanoyl homoserine lactone producing Pseudomonas aurantiaca strain PB-St2, isolated from sugarcane stem. *J. Microbiol. Biotechol.* 19, 1688–1694. doi: 10.4014/jmb.0904.04022

Mehnaz, S., Baig, D. N., and Lazarovits, G. (2010). Genetic and phenotypic diversity of plant growth promoting rhizobacteria isolated from sugarcane plants growing in Pakistan. *J. Microbiol. Biotechnol.* 20, 1614–1623. doi: 10.4014/jmb.1005.05014

Mehnaz, S., Weselowski, B., Mufti, F. A., Zahid, S., Lazarovits, G., and Iqbal, J. (2009b). Isolation, characterization and effect of fluorescent pseudomonads on micropropagated sugarcane. *Can. J. Microbiol.* 55, 1007–1011. doi: 10.1139/W09-050

Mendes, R., Pizzirani-Kleiner, A. A., Araujo, W. L., and Raaijmakers, J. M. (2007). Diversity of cultivated endophytic bacteria from sugarcane: genetic and biochemical characterization of *Burkholderia cepacia* complex isolates. *Appl. Environ. Microbiol.* 73, 7259–7267. doi: 10.1128/AEM.01222-07

Minaxi and Saxena, J. (2010a). Characterization of Pseudomonas aeruginosa RM-3 as a potential biocontrol agent. *Myco Path.* 170, 181–193. doi: 10.1007/s11046-010-9307-4

Minaxi and Saxena, J. (2010b). Disease suppression and crop improvement in moong beans (Vignaradiata) through Pseudomonas and Burkholderia strains isolated from semi-arid region of Rajasthan, India. *J. Biocontrol.* 55, 799–810. doi:10.1007/s10526-010-9292-z

Mirza, M. S., Mehnaz, S., Normand, P., Prigent-Combaret, C., Moenne-Loccoz, Y., Bally, R., et al. (2006). Molecular characterization and PCR-detection of a nitrogen-fixing Pseudomonas strain promoting rice growth. *Biol. Fertil. Soil.* 43, 136–170. doi: 10.1007/s00374-006-0074-9

Niu, J. Q., Wang, A. Q., Huang, J. L., Li, Y. R., and Yang, L. T. (2013). Cloning and expression analysis of a soluble acid invertase gene (SoSAI1) of sugarcane. *Sci. Agric. Sin.* 46, 5248–5260. doi: 10.3864/j.issn.0578-1752.2013.24.019

Noda, S., Ohkuma, M., Usami, R., Horikoshi, K., and Kudo, T. (1999). Culture-independent characterization of a gene responsiblefor nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. *Appl. Environ. Microbiol.* 65, 4935–4942.

Oliveira, A. L. M., Urquiaga, S., Dobereiner, J., and Baldani, J. I. (2002). The effect of inoculating endophytic N2 -fixing bacteria on micropropagated sugarcane plants. *Plant Soil* 242, 205–215. doi: 10.1023/A:1016249704336

Onofre-Lemus, J., Hernandez-Lucas, I., Girard, L., and Caballero-Mellado, J. (2009). ACC (1-amino cyclopropane-1-carboxylate) deaminase activity, a widespread trait in Burkholderia species, and its growth-promoting effect on tomato plants. *Appl. Environ. Microbiol.* 75, 6581–6590. doi: 10.1128/AEM.01240-09

Poly, F., Monrozier, L. J., and Bally, R. (2001). Improvement in the RFLP procedure for studying the diversity of nifH genes in communities of nitrogen fixers in soil. *Res. Microbiol.* 152, 95–103. doi: 10.1016/S0923-2508(00)01172-4

Raaijmakers, J. M., Weller, D. M., and Thomashow, L. S. (1997). Frequency of antibiotic-producing *Pseudomonas* spp. in natural environments. *Appl. Environ. Microbiol.* 63, 881–887.

Rademaker, J. L. W., and de Bruijn, F. J. (1997). "Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis," in *DNA Markers: Protocols, Applications and Overviews,* eds G. Caetano-Anolles and P. M. Gresshoff (New York, NY: John Wiley), 151–171.

Rafikova, G. F., Korshunova, T.,Yu., Minnebaev, L. F., Chetverikov, S. P., and Loginov, O. N. (2016). A new bacterial strain, *Pseudomonas koreensis* IB-4, as a promising agent for plant pathogen biological control. *Microbiology* 85, 333–341. doi: 10.1134/S0026261716030115

Ramette, A., Frapolli, M., Defago, G., and Moenne-Loccoz, Y. (2003). Phylogeny of HCN synthase-encoding hcnBC genes in biocontrol fluorescent pseudomonads and its relationship with host plant species and HCN synthesis ability. *Mol. Plant Microb. Interact.* 16, 525–535. doi: 10.1094/MPMI.2003.16.6.525

Rao, G. P., Viswanathan, R., and Singh, S. B. (2002). "Current situation of sugarcane diseases in India" in *Sugarcane Crop Management,* eds S. B. Singh, G. P. Rao, and S. Easwaramoorthy, (Houstan, TX: SCI Tech Publishing LLC), 734.

Reis, V. M., Olivares, F. L., de Oliveira, A. L. M., dos Reis, J. F. B., Baldani, J. I., and Dobereiner, J. (1999). Technical approaches to inoculate micro-propagated sugarcane plants with *Acetobacter diazotrophicus*. *Plant Soil.* 206, 205–211. doi: 10.1023/A:1004436611397

Robinson, N., Brackin, R., Vinall, K., Soper, F., Holst, J., Gamage, H., et al. (2011). Nitrate paradigm does not hold up for sugarcane. *PLoS ONE* 6:e19045. doi: 10.1371/journal.pone.0019045

Rosado, A. S., Duarte, G. F., Seldin, L., and Elsas, J. D. V. (1998). Genetic diversity of nifH gene sequences in Paenibacillus azotofixans strains and soil samples analyzed by denaturing gradient gel electrophoresis of PCR amplified gene fragments. *Appl. Environ. Microbiol.* 64, 2770–2779.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Sasirekha, B., Srividya, S., and Shankar, B. S. (2013). Molecular detection of antibiotic related genes from *Pseudomonas aeruginosa* FP6, an antagonist towards *Rhizoctonia solani* and *Colletotrichum gloeosporioides. Turk. J. Biol.* 37, 289–295. doi: 10.3906/biy-1207-56

Schwyn, B., and Neilands, J. B. (1987). Universal chemical assay for the detection and determination of siderophores. *Anal. Biochem.* 160, 47–56. doi: 10.1016/0003-2697(87)90612-9

Setten, L., Soto, G., Mozzicafreddo, M., Fox, A. R., Lisi, C., Cuccioloni, M., et al. (2013). Engineering *Pseudomonas protegens* Pf-5 for nitrogen fixation and its application to improve plant growth under nitrogen-deficient conditions. *PLoS ONE* 8:e63666. doi: 10.1371/journal.pone.0063666

Singh, M. K., Kushwaha, C., and Singh, R. K. (2009). Studies on endophytic colonization ability of two upland rice endophytes, *Rhizobium* sp. and *Burkholderia* sp., using green fluorescent protein reporter. *Curr. Microbiol.* 59, 240–243. doi: 10.1007/s00284-009-9419-6

Singh, R. K., Kumar, D. P., Solanki, M. K., Singh, P., Srivastva, A., Kumar, S., et al. (2013). Optimization of media components for chitinase production by chickpea rhizosphere associated *Lysinibacillus fusiformis* B-CM18. J. *Basic Microb.* 53, 451–460. doi: 10.1002/jobm.201100590

Singh, R. K., Kumar, D. P., Solanki, M. K., Singh, P., Srivastava, S., Srivastva, A. K., et al. (2014). Multifarious plant growth promoting characteristics of chickpea rhizosphere associated *Bacilli* help to suppress soil-borne pathogens. *Plant Growth Regul.* 73, 91–101. doi: 10.1007/s10725-013-9870-z

Sneath, P. H. A., and Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco, CA: Freeman.

Solanki, M. K., Wang, Z., Wang, F.-Y., Li, C.-N., Lan, T.-J., Singh, R. K., et al. (2017). Intercropping in sugarcane cultivation influenced the soil properties and enhanced the diversity of vital diazotrophic bacteria. *Sugar Tech.* 19, 136–147. doi: 10.1007/s12355-016-0445-y

Somers, E., Vanderleyden, J., and Srinivasan, M. (2004). Rhizosphere bacterial signalling: a love parade beneath our feet. *Crit. Rev. Microbiol.* 30, 205–240. doi: 10.1080/10408410490468786

Souza, J. T., and Raaijmakers, J. M. (2003). Polymorphisms within the prnD and pltC genes from pyrrolnitrin and pyoluteorin-producing *Pseudomonas* and *Burkholderia* spp. *FEMS Microbiol. Ecol.* 43, 21–34. doi: 10.1111/j.1574-6941.2003.tb01042.x

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030–11035. doi: 10.1073/pnas.0404206101

Unge, A., Tombolini, R., Molbak, L., and Jansson, J. K. (1999). Simultaneously monitoring of cell number and metabolic activity of specific bacterial populations with a dual gfp-lux AB marker system. *Appl. Enviorn. Microbiol.* 65, 813–821.

Urquiaga, S., Cruz, K. H. S., and Boddey, R. M. (1992). Contribution of nitrogen fixation to sugarcane: nitrogen-15 and nitrogen-balance estimates. *Soil Sci. Soc. Am. J.* 56, 105–114. doi: 10.2136/sssaj1992.03615995005600010017x

Urquiaga, S., Xavier, R. P., de Morais, R. F., Batista, R. B., Schultz, N., Leite, J. M., et al. (2012). Evidence from field nitrogen balance and 15N natural abundance data for the contribution of biological N2 fixation to Brazilian sugarcane varieties. *Plant Soil* 356, 5–21. doi: 10.1007/s11104-011-1016-3

Versalovic, J., Koeuth, T., and Lupski, J. R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.* 19, 6823–6831. doi: 10.1093/nar/19.24.6823

Viswanathan, R., Rajitha, R., Sundar, A. R., and Ramamoorthy, V. (2003). Isolation and identification of endophytic bacterial strains isolated from sugarcane stalks and their *in vitro* antagonism against red rot pathogen. *Sugar Tech.* 5, 25–29. doi: 10.1007/BF02943760

Viswanathan, R., and Rao, G. P. (2011). Disease scenario and management of major sugarcane diseases in India. *Sugar Tech.* 13, 336–353. doi: 10.1007/s12355-011-0102-4

Viswanathana, R., and Samiyappan, R. (2002). Induced systemic resistance by fluorescent pseudomonads against red rot disease of sugarcane caused by *Colletotrichum falcatum*. *Crop Prot.* 21, 1–10. doi: 10.1016/S0261-2194(01)00050-3

Voisard, C., Keel, C., Haas, D., and Defago, G. (1989). Cyanide production by *Pseudomonas fuorescens* helps suppress black root rot of tobacco under gnotobiotic conditions. *EMBO J.* 8, 351–358.

Wallenstein, M. D. (2004). *Effect of Increased Nitrogen Deposition on Forest Soil Nitrogen Cycling and Community Structure*. Ph.D. Dissertation, Duke University, Durham, NC.

Walsh, U. F., Morrissey, J. P., and O'Gara, F. (2001). Pseudomonas for biocontrol of phytopathogens: from functional genomics to commercial exploitation. *Curr. Opin. Biotechnol.* 12, 289–295. doi: 10.1016/S0958-1669(00)00212-3

Wang, X., Wang, Y., Tian, J., Lim, B. L., Yan, X., and Liao, H. (2009). Over expressing at PAP15 enhances phosphorus efficiency in soybean. *Plant Physiol.* 151, 233–240. doi: 10.1104/pp.109.138891

Wielbo, J., Marek-Kozaczuk, M., Kubik-Komar, A., and Skorupska, A. (2007). Increased metabolic potential of *Rhizobium* spp. is associated with bacterial competitiveness. *Can. J. Microbiol.* 53, 957–967. doi: 10.1139/W07-053

Xing, Y.-X., Wei, C.-Y., Mo, Y., Yang, L.-T., Huang, S.-L., and Li, Y.-R. (2016). Nitrogen-fixing and plant growth-promoting ability of two endophytic bacterial strains isolated from sugarcane stalks. *Sugar Tech.* 18, 373–379. doi: 10.1007/s12355-015-0397-7

Xing, Y. X., Yang, L. T., Huang, S. L., and Li, Y. R. (2006). Identification of a new nitrogen fixing endo-bacterium strain isolated from sugarcane stalk. *Sugar Tech.* 8, 49–53. doi: 10.1007/BF02943741

Xing, Y., Yang, L., Huang, S., and Y., Li (2015). "A new nitrogen fixing endo-bacterium strain isolated from sugarcane stem," in *Proceedings of International Symposium on Technologies to improve Sugar Productivity in Developing Countries*, ed P. R. Guilin (Beijing: China Agricultural Science), 487–490.

Young, J. P. W. (1992). "Phylogenetic classification of nitrogen fixing organisms" in *Biological Nitrogen Fixation,* Vol. 1544, eds G. Stacey, R. H. Burris, and H. J. Evans (New York, NY: Chapman & Hall), 43–86.

Zahir, Z. A., Muhammad, A., and Frankenberger, W. T. J. (2003). Plant growth promoting rhizobacteria: applications and perspectives in agriculture. *Adv. Agron.* 81, 97–168. doi: 10.1016/S0065-2113(03)81003-9

Zehr, J. P., and Capone, D. G. (1996). Problems and promises of assaying the genetic potential for nitrogen fixation in the marine environment. *Microb. Ecol.* 32, 263–281. doi: 10.1007/BF00183062

Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ. Microbiol.* 5, 539–554. doi: 10.1046/j.1462-2920.2003.00451.x

Zeng, Q., Wu, X., and Wen, X. (2016). Identification and characterization of the rhizosphere phosphate-solubilizing bacterium *Pseudomonas frederiksbergensis* JW-SD2, and its plant growth-promoting effects on poplar seedlings. *Ann. Microbiol.* 66, 1343–1354. doi: 10.1007/s13213-016-1220-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Li, Singh, Singh, Song, Xing, Yang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 August 2017
doi: 10.3389/fmicb.2017.01516

Long-term Fertilization Structures Bacterial and Archaeal Communities along Soil Depth Gradient in a Paddy Soil

*Yunfu Gu[1], Yingyan Wang[1], Sheng'e Lu[1], Quanju Xiang[1], Xiumei Yu[1], Ke Zhao[1], Likou Zou[1], Qiang Chen[1], Shihua Tu[2] and Xiaoping Zhang[1]**

[1] Department of Microbiology, College of Resource Science and Technology, Sichuan Agricultural University, Chengdu, China, [2] Soil and Fertilizer Institute, Sichuan Academy of Agricultural Sciences, Chengdu, China

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Steffen Kolb, Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF), Germany
Jacynthe Masse, Institut de Recherche en Biologie Végétale, Canada

***Correspondence:**
Xiaoping Zhang
zhangxiaopingphd@126.com

Specialty section:
This article was submitted to Terrestrial Microbiology, a section of the journal Frontiers in Microbiology

Received: *01 June 2017*
Accepted: *27 July 2017*
Published: *15 August 2017*

Citation:
Gu Y, Wang Y, Lu S, Xiang Q, Yu X, Zhao K, Zou L, Chen Q, Tu S and Zhang X (2017) Long-term Fertilization Structures Bacterial and Archaeal Communities along Soil Depth Gradient in a Paddy Soil. Front. Microbiol. 8:1516. doi: 10.3389/fmicb.2017.01516

Soil microbes provide important ecosystem services. Though the effects of changes in nutrient availability due to fertilization on the soil microbial communities in the topsoil (tilled layer, 0–20 cm) have been extensively explored, the effects on communities and their associations with soil nutrients in the subsoil (below 20 cm) which is rarely impacted by tillage are still unclear. 16S rRNA gene amplicon sequencing was used to investigate bacterial and archaeal communities in a Pup-Calric-Entisol soil treated for 32 years with chemical fertilizer (CF) and CF combined with farmyard manure (CFM), and to reveal links between soil properties and specific bacterial and archaeal taxa in both the top- and subsoil. The results showed that both CF and CFM treatments increased soil organic carbon (SOC), soil moisture (MO) and total nitrogen (TN) while decreased the nitrate–N content through the profile. Fertilizer applications also increased Olsen phosphorus (OP) content in most soil layers. Microbial communities in the topsoil were significantly different from those in subsoil. Compared to the CF treatment, taxa such as *Nitrososphaera*, *Nitrospira*, and several members of *Acidobacteria* in topsoil and Subdivision 3 *genera incertae sedis*, *Leptolinea*, and *Bellilinea* in subsoil were substantially more abundant in CFM. A co-occurrence based network analysis demonstrated that SOC and OP were the most important soil parameters that positively correlated with specific bacterial and archaeal taxa in topsoil and subsoil, respectively. *Hydrogenophaga* was identified as the keystone genus in the topsoil, while genera *Phenylobacterium* and *Steroidobacter* were identified as the keystone taxa in subsoil. The taxa identified above are involved in the decomposition of complex organic compounds and soil carbon, nitrogen, and phosphorus transformations. This study revealed that the spatial variability of soil properties due to long-term fertilization strongly shapes the bacterial and archaeal community composition and their interactions at both high and low taxonomic levels across the whole soil profile.

Keywords: soil profile, miseq sequencing, soil bacteria, soil archaea, specific taxa, network analysis, 16S rRNA gene amplicon

INTRODUCTION

Soil microbes play key roles in the functions of ecosystems by cycling nutrients, degrading organic material and pollutants as well as in maintaining the quality of groundwater (Madsen, 1995; Fierer et al., 2003a; King, 2014). In the topsoil (0–20 cm, tilled layer) both the microbial biomass and diversity are the greatest, yet in the subsoil (below 20 cm) microbes are also diverse and abundant due to the large volume of soil on a depth-weighted basis (Will et al., 2010; Li et al., 2014). Soil microbial communities varied significantly along with the soil depth, and the microbial diversity of microorganisms typically decreases with depth (Li et al., 2014; He et al., 2017). Fertilization, an essential agricultural practice used primarily to increase nutrient availability to crop plants, causes concomitant changes in the soil properties and microbial communities (Marschner et al., 2003). Studies have reported that over fertilization with nitrogen can result in negative consequences to the environment such as soil acidification, decreased soil microbial activity, enhanced nitrification and the leaching of nitrate nitrogen (Nowinski et al., 2008; Guo et al., 2010; Ramirez et al., 2010). In contrast, the combined use of inorganic and organic fertilizer likely increase the microbial diversity, biomass and activity, and helps to maintain even increase the soil nutrients (Birkhofer et al., 2008; Kumar et al., 2017). Nutrients added to the soil by fertilizer may directly alter the abundance and composition of microbial community in topsoil; in subsoil soluble nutrients leaked from topsoil change these microbial parameters indirectly (Stowe et al., 2010; Eilers et al., 2012). However, little is known about the characterization and spatial variability of microbial communities with depth in fertilized paddy soils (Bao et al., 2015; Yu et al., 2015).

Certain bacterial and archaeal taxa at high taxonomic levels (e.g., phylum or class) have shown ecological coherence since they respond predictably to environmental variables (Philippot et al., 2010; Cederlund et al., 2014). The abundance of *Proteobacteria* and *Actinobacteria* in the topsoil increased under long-term combined use of chemical and organic fertilizer while the abundance of *Acidobacteria* decreased (Li et al., 2014). However, little is known about the response of specific soil microbial taxa at low taxonomic levels (e.g., genus or species) to long-term fertilization. Long-term application of organic fertilization seems to select for certain microbial taxa at low taxonomic levels that feed primarily on organic substrates and proliferate greatly, resulting in the changes in microbial community composition and soil nutrient status (Cederlund et al., 2014; Li et al., 2014, 2017). As a result, specific microbial taxa of which the abundances are substantially increased by long-term fertilization should show some degree of connections with soil nutrients. These predictable responses of specific microbial taxa make them possibly useful indicators of soil nutrient status and sustainability (Smit et al., 2001; Hartman et al., 2008). Moreover, microbes including these specific microbial taxa formed complex interaction webs in soil ecosystem, and understanding the interactions between them is critical to explore the complexity of functional processes (Faust and Raes, 2012). Co-occurrence network analysis, based on the analysis of correlations between the abundances of microbial taxa, has provided comprehensive perspective into the microbial association patterns and the ecological functions guiding community assembly, such as commensalism, competition, and predation. Thus, network analysis has been used to describe microbial co-occurrence relations in diverse environments including marine sediments, soil, and waste water. However, knowledge on the effects of long-term fertilization on the networks of taxon co-occurrence of microbial communities across the whole soil profiles is still scarce.

To address these questions, we analyzed microbial communities in a long-term fertilization field experiment site in Sichuan, China. Our previous studies on this calcareous purplish paddy soil site showed that the NPK and NPK combined with farmyard manure (NPKM) treatments significantly increased the rice and wheat yields, decreased the soil pH and modified the microbial community in the topsoil (Gu et al., 2008). In this study, two typical fertilizer treatments [chemical fertilizer (CF; NPK), and chemical fertilizer combined with farmyard manure (CFM; NPKM)] were compared with a non-fertilized treatment (CK), and deep 16S rRNA gene amplicon sequencing was used to examine the changes in bacterial and archaeal communities with depth. Differential abundance analysis and co-occurrence based network analysis were performed to unravel the potential effects of specific bacterial and archaeal taxa and their associations with soil nutrients. Specifically, we examined (1) the responses of microbial communities to long-term fertilizer applications across the soil profile, (2) which specific taxa in different soil depths are substantially stimulated by long-term fertilization, and (3) the associations of specific bacterial and archaeal taxa responding to long-term fertilization in the topsoil and subsoil. We hypothesized that (i) the topsoil, which is directly affected by long-term fertilization, would host a distinct composition and diversity of soil microbial communities compared to those in subsoil; and (ii) specific taxa involved in nutrient cycling would be substantially stimulated by long-term NPKM fertilization throughout the depth of soil profile due to the incorporation of farmyard manure. Moreover, we hypothesized that (iii) specific taxa of which the abundances are substantially changed by long-term fertilization would form distinct association patterns in the topsoil and subsoil.

MATERIALS AND METHODS

Field Site and Experimental Design

The experimental site was a 'N, P, K long-term fertilization field experiment' site on a calcareous purplish paddy field established in 1982 in Chuanshan (30°10′50″N, 105°03′26″E), Sichuan, China. The site which has an annual average temperature of 17.4°C and mean annual precipitation of 930 mm was described in a previous study (Gu et al., 2008). The experiment included three treatments, CF (NPK), chemical fertilizer CFM (NPKM) and unfertilized control (CK), in three replicate 13.2 m^2 plots. In the fertilizer treatments, the application rates of inorganic fertilizers were similar to local traditional application rates: N, 120 kg ha^{-1}; P_2O_5, 60 kg ha^{-1}; K_2O, 60 kg ha^{-1}. CFM treatment

received 3×10^4 kg ha^{-1} pig manure. The fertilizer treatments remained the same each year. All P as superphosphate and K as K_2SO_4 were applied as basal fertilizers, while N as urea was split into 70% as basal application and 30% as topdressing for rice at tillering. Soil original physico-chemical properties prior to the experiment are in Gu et al. (2008). The mean grain yields in the CK, CF, and CFM treatments were 2967.7 kg ha^{-1}, 7116.8 kg ha^{-1}, and 7496.5 kg ha^{-1}, which were significantly different ($P < 0.05$) (Gu et al., 2008).

Soil Sampling and Selected Soil Properties Measurement

In 2012, after the summer rice was transplanted and all the plots were flooded for about 55 days, five sampling points were chosen randomly in each plot and combined as one composite sample. Based on the soil profile, soil samples were collected at the following depth intervals (cm): 0–20, 20–40, 40–60, and 60–90 cm. Finally, a total of 36 soil samples were taken from the experiment site. After the removal of visible roots and fresh litter material, the composite samples were homogenized and then separated into two parts: approximately 50 g was packed into a sterile bag, immersed in liquid nitrogen instantly and stored at −70°C for DNA analysis, while the other part (approximately 800 g) was air-dried at ambient room temperature (∼25°C) and sieved through a 6-mm sieve for soil physicochemical parameter determination.

Soil pH was measured with a compound electrode (E-201-C, Shanghai Shengguang Instrument Co. Ltd. Shanghai, China) using a soil-to-water ratio of 1:1. Soil moisture (gravimetric water content) was measured using a wet/dry soil conversion with a soil subsample dried at 105°C for 12 h. Soil organic carbon (SOC) and total N (TN) were determined by dichromate oxidization and Kjeldahl digestion, respectively. Ammonium–N (NH_4^+–N) and nitrate–N (NO_3^-–N) were extracted by shaking 5 g of field-moist soil with 50 mL of 0.01 M $CaCl_2$ for 30 min. Filtered extracts were kept at −20°C until NH_4^+–N and NO_3^-–N concentration were measured in an AA3 flow injection analyzer (FIA SFA CFA, Germany). Soil Olsen phosphorus (OP) and available potassium (AK) were extracted with sodium bicarbonate and ammonium acetate (Lu, 1999), respectively.

DNA Extraction

Soil crude DNA was extracted from 0.5 g fresh soil with three replications using a FastDNA spin kit for soil (Qbiogene, Carlsbad, CA, United States) by following the manufacturer manual. DNA quality was assessed based on the spectrometry absorbance ratios at wavelengths of 260/280 nm and 260/230 nm by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, United States). The integrity of the DNA extracts was confirmed by electrophoresis. DNA samples were stored at −20°C.

Amplification of 16S rRNA Gene

Sequencing of PCR amplicons of 16S rRNA gene was conducted with the Illumina MiSeq (Illumina, San Diego, CA, United States) targeting the V4 hyper variable regions (515F, 5′-GTGCCAGCMGCC-GCGGTAA-3′ and 806R, 5′-GGACTACHVGGGTWTCTAAT-3′) for 2 × 150-bp paired-end sequencing (Caporaso et al., 2012). PCR reactions (Qiagen, Valencia, CA, United States) were performed in a 50 μl mixture containing 0.25 μl of TaKaRa Ex Taq HS at 5 U/μl, 4 μl of dNTP Mixture (2.5 mM each), 5 μl of 10 × Ex Taq Buffer (Mg^{2+} Plus), 1 μl of each primer at 10 mM and 1 μl of 10-fold diluted template DNA. The thermal program was as follows: initial denaturation at 94°C for 4 min, followed by 30 cycles of 94°C for 20 s, 53°C for 25 s, 68°C for 45 s, and a final extension step of 10 min at 68°C. PCR products (3 μl) were examined by agarose gel electrophoresis, and then 5 μl of triplicate reactions were mixed together and quantified with PicoGreen (Invitrogen Ltd., Paisley, United Kingdom). Approximately 200 ng PCR products from each sample were pooled together and purified with a QIAquick PCR Purification Kit (QIAGEN), and then re-quantified with PicoGreen. Denaturation was performed by mixing 10 μl of combined PCR products (2 nM) and 10 μl 0.1 N NaOH. Denatured DNA was diluted to 6 pM and mixed with an equal volume of 6 pM PhiX library. Finally, the 600 μl mixture was loaded into the reagent cartridge and run on a MiSeq sequencer (Illumina) for 300 cycles.

Illumina Data Analysis

All reads were assembled and quality-filtered using a fast length adjustment of short reads (FLASH) software (Magoc and Salzberg, 2011) and QIIME pipeline (Caporaso et al., 2010), respectively. Only those reads with phred-quality score more than 20 and length over 300 bps were considered for further analysis. Those reads containing ambiguous alphabets, harboring two or more mismatches in primer or unable to be assembled were discarded during quality polishing. Operational taxonomic units (OTUs) were picked using the UPARSE pipeline (Edgar, 2013) at 97% identity. Chimeric sequences were removed using UCHIME (Caporaso et al., 2012). The RDP classifier was used to pick representative sequences for the OTUs and to assign taxonomic data to each representative sequence at the 70% threshold (Caporaso et al., 2010). Sequences which could not be assigned were removed. The sequence data were submitted to NCBI Sequence Read Archive[1] with accession number SRS2127454. After OTU clustering at 97% sequence identity, removal of singletons and resample at 9500 sequences per sample, 34030 OTUs for all the 36 soil samples were obtained and used in downstream analysis. The Good's coverage and rarefaction curves were calculated using QIIME pipeline and 'rarecurve' function in R package vegan, respectively (Caporaso et al., 2010; R Development Core Team, 2010). The Good's coverage was ranged from 74.2 to 87.4% with an average of 80.2% (Supplementary Table S1). Meanwhile, the rarefaction curves did not reach a plateau, indicating that further sequencing could have revealed additional species coverage (Supplementary Figure S1).

Statistical Analysis

Normality and variance homogeneity of community data were analyzed using Shapiro–Wilks and Levene's test, respectively.

[1] https://www.ncbi.nlm.nih.gov/sra/

Principal coordinates analysis (PCoA) was applied to estimate the microbial community structure using Bray–Curtis dissimilarities based on "Hellinger" transformed community data (Legendre and Gallagher, 2001). Analysis of similarity (ANOSIM) was used to test the dissimilarity between microbial community structure in topsoil (0–20 cm) and in subsoil (20–90 cm) (Zhou et al., 2012). After the soil edaphic parameters were standardized with the "decostand" function, distance-based redundancy analysis (db-RDA) was performed to correlate these parameters to microbial community structure in both topsoil and subsoil using "capscale" function in R package vegan (R Development Core Team, 2010). The significance of db-RDA models was tested using "permutest" function in vegan with 999 permutations. The goodness-of-fit (R2) and associated statistical significance (*P*-value) of each edaphic factor in db-RDA model were verified using "envfit" function in vegan.

Differential abundance analysis was performed in R package DESeq2 (Love et al., 2014). *P*-values were adjusted for multiple testing with the procedure described by Benjamini and Hochberg (1995), and a false discovery rate (FDR) of 10% was selected to denote statistical significance (Love et al., 2014; Whitman et al., 2016). Enriched and depleted OTUs were defined using the methods described by Li et al. (2017).

Co-abundance Network Analysis

To minimize pairwise comparison and network complexity, only OTUs with large differential abundance ($\log_2$ fold change $>|\pm 1|$ and FDR-adjusted *P*-value < 0.1) were selected for network analysis (Li et al., 2017). The interaction network was inferred based on Spearman rank correlation matrix constructed with Python package scipy (Oksanen et al., 2013). All *P*-values were adjusted using Benjamini and Hochberg FDR controlling procedure (Benjamini et al., 2006) in R package multtest (Pollard et al., 2013). Connections between identical nodes were removed before network construction using Python package igraph (Csardi and Nepusz, 2006). Based on correlation coefficient thresholds and corresponding cutoffs of FDR-adjusted *P*-values for correlation, two interaction networks were inferred using differentially abundant OTUs from 0 to 20 cm topsoil (38 OTUs) and 20–90 cm subsoil (107 OTUs). The cutoffs of correlation coefficient were automatically determined as ± 0.79 for topsoil and ± 0.69 for subsoil through random theory-based methods (Luo et al., 2006). The topological features of the abundance correlation networks were not compared in this study because of the different correlation coefficient cutoffs used in meta-network construction. The threshold of FDR-adjusted *P*-values was 0.001. Network properties were calculated using Python package igraph (Csardi and Nepusz, 2006). Node level topological features including degree and betweenness centrality were calculated to identify potential keystone species. The network modules were detected using greedy modularity optimization algorithm (Newman, 2006). Modules with four or more nodes were selected for further analysis. The sources of nodes in each module were determined using an in-house Python script based on the results of differential abundance analysis. To identify the primary driving force of each network module, mantel test was performed between nodes in each network module and associated environmental factors. Networks were visualized using Cytoscape v.3.2.1 (Shannon et al., 2003).

RESULTS

Soil Physico-Chemical Parameters

Compared to the unfertilized control (CK), the two fertilizer treatments led to lower pH and higher soil moisture, SOC, TN, OP, and AK content in topsoil (0–20 cm) (**Figure 1**). Compared to the topsoil, the soil pH in both the CF and CFM treatments was higher in the subsoil (20–90 cm) (**Figure 1A**), while soil moisture, SOC, and TN were lower in both CF and CFM treatments at the subsoil (**Figures 1B–D**). In the topsoil, both soil moisture and SOC in CFM treatment was significantly higher than those in CF ($P = 0.05$) (**Figures 1B,C**). Fertilizer applications led to higher OP content in most soil layers, and significantly higher in CF treatment that those in CFM ($P = 0.05$) (**Figure 1E**). Soil AK in the topsoil was also significantly higher in the CFM ($P = 0.05$), while they were low in both CF and CFM treatments in the subsoil (**Figure 1F**). In addition, ammonium–N content was higher than nitrate–N content (**Table 1**). Unlike ammonium–N with non-uniform distribution within the soil profile, nitrate–N content significantly decreased with depth. Fertilizer applications produced significant decreases in nitrate–N content through the profile (**Table 1**) ($P = 0.05$).

Relative Abundance of Dominant Phyla and Orders

Proteobacteria and *Acidobacteria* were the most abundant phyla in the topsoil (0–20 cm), with relative abundances of 24–38% and 18–22%, respectively, and were followed by *Verrucomicrobia* (6.2–9.0%), *Chloroflexi* (2.0–7.2%), and *Bacteroidetes* (2.7–5.8%). Although the relative abundance of *Proteobacteria* showed a non-uniform variation pattern through the four different soil depths, the weighted mean values of the relative abundance were highest in the subsoil (20–90 cm) (**Figure 2**) ($P = 0.05$). The relative abundance of *Chloroflexi* increased significantly with increasing soil depth, while the relative abundances of *Acidobacteria*, *Verrucomicrobia*, and *Bacteroidetes* decreased (**Figure 2A**) ($P = 0.05$).

Compared to the control (CK), the relative abundance of *Proteobacteria* was higher in the CF and lower in the CFM treatment in topsoil (**Figure 2B**), and especially CF and also CFM resulted in higher relative abundances of *Betaproteobacteria* and *Gammaproteobacteria*, and lower relative abundances of *Deltaproteobacteria* in topsoil (Supplementary Figure S2) ($P = 0.01$). In addition, compared to the CK treatment, the relative abundance of *Alphaproteobacteria* was higher in the CF and lower in the CFM (Supplementary Figure S2). Finer taxonomic divisions also revealed the effect of fertilizer application in topsoil at the order level (**Table 2**). Compared to the CK, the relative abundance of *Nitrosomonadales* in *Betaproteobacteria* was 36 and 28 times higher in the CF and CFM treatments, respectively, and the abundance

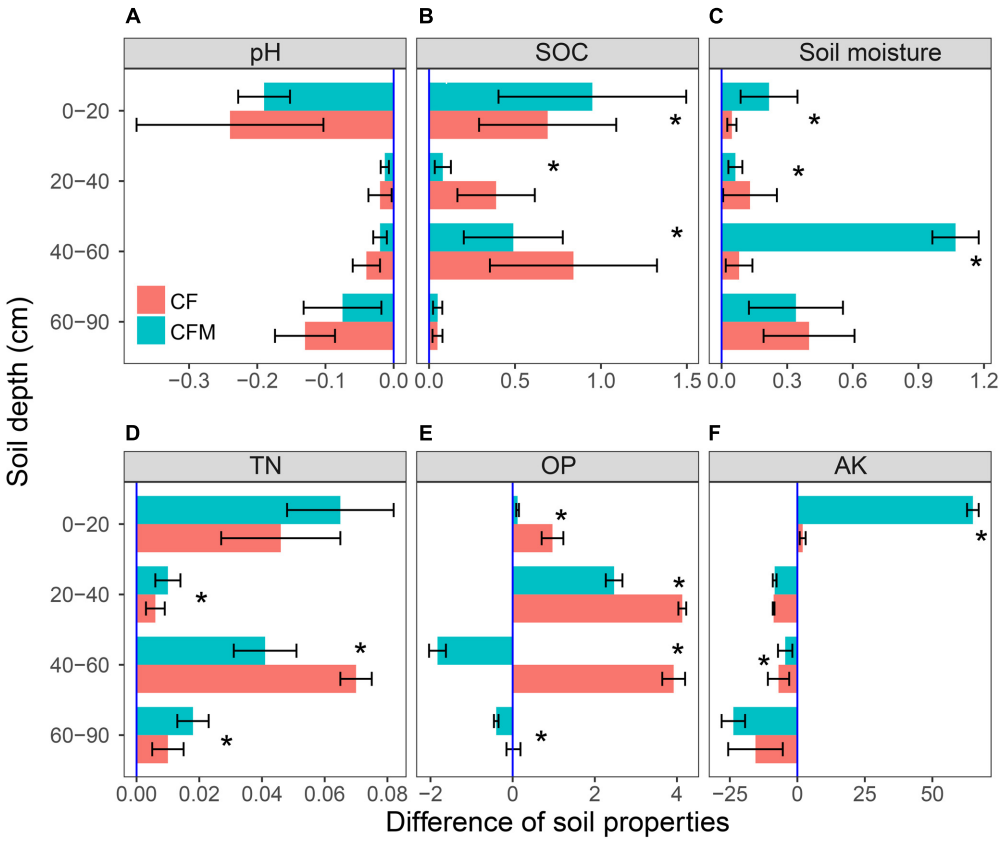


FIGURE 1 | Soil properties at 0–90 cm depths under long-term applications (32 years) of NPK fertilizer (CF) and NPK combined with farmyard manure (CFM) compared with the non-fertilized control (CK). SOC, soil organic carbon; Soil moisture: gravimetric water content; TN, total nitrogen; OP, Olsen phosphorus (available phosphorus); and AK, available potassium. Error bars represent the standard error of the mean (n = 3). * means significantly difference among three different long-term fertilization treatments at the level of P = 0.05.

of *Methylophilales* was 7.3 and 2.9 times higher in the CF and CFM, respectively (P = 0.01). The abundances of *Xanthomonadales*, *Methylococcales,* and *Pseudomonadales* within *Gammaproteobacteria* were also higher in the CF and CFM treatments, whereas those of *Acidobacteria* were lower in the CF and CFM treatment (**Figure 2**). The abundances of minor phyla like *Nitrospirae* (Supplementary Figure S3) were higher in the fertilizer treatments whereas those of *Gemmatimonadetes* were lower (**Figure 2**) (P = 0.05).

Long-term fertilizer application also changed the microbial community in subsoil (20–90 cm). For example, the relative abundance of *Gammaproteobacteria* was lower in the fertilizer treatments, especially in the CFM (Supplementary Figure S2). The relative abundances of *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* were higher in the CF but lower in the CFM treatment compared to the CK (P = 0.05). For the finer taxonomic divisions, the relative abundances of the orders *Xanthomonadales*, *Methylococcales*,

TABLE 1 | Ammonium–N (NH_4^+–N) and nitrate–N (NO_3^-–N) contents at the 0–90 cm depth under long-term fertilizer treatments.

Soil depth (cm)	NH_4^+–N (mg kg^{-1})			NO_3^-–N (mg kg^{-1})		
	CK	CF	CFM	CK	CF	CFM
0–20	3.93 ± 0.11a[a]C[b]	5.14 ± 0.02aB	8.29 ± 0.06aA	6.74 ± 0.15aA	2.43 ± 0.88bA	2.10 ± 0.34bA
20–40	8.71 ± 0.07aA	5.75 ± 2.00bA	7.98 ± 0.03aA	2.72 ± 0.01aB	2.12 ± 0.16aA	2.09 ± 0.12aA
40–60	7.12 ± 0.09aB	5.85 ± 0.07bA	5.39 ± 0.09bC	2.57 ± 0.25aB	1.81 ± 0.30bC	1.71 ± 0.06aA
60–90	7.61 ± 0.06aA	4.32 ± 0.08bC	6.75 ± 0.06aB	1.12 ± 0.15bC	1.19 ± 0.05aB	0.53 ± 0.13cB

CK, no fertilizer; CF, NPK fertilizer; CFM, NPK fertilizer combined with pig manure. Values in the column are mean ± SE (stand error). [a]*Treatment means within a depth followed by the same lower case letter (a, b, and c) are not significantly different (P > 0.05).*
[b]*Depth means within a column followed by the same upper case letter (A, B, and C) are not significantly different (P > 0.05).*

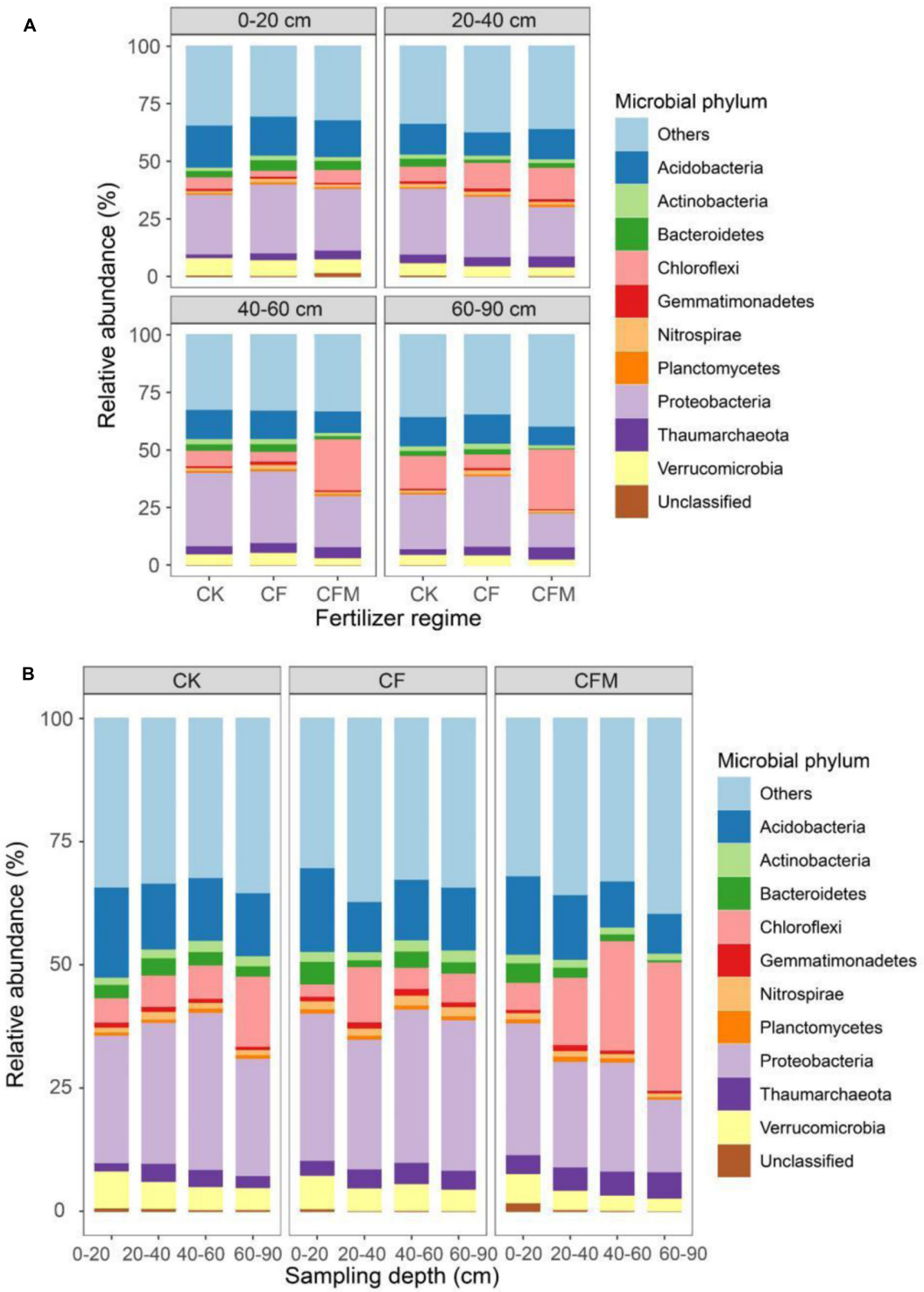


FIGURE 2 | Relative abundances of the top 10 microbial phyla at different depths **(A)** under long-term fertilizer treatments **(B)**. 'Others' refer to those identified phyla beyond the top 10 phyla. CK, no fertilizer; CF, NPK fertilizer; CFM, NPK fertilizer combined with farmyard manure.

and *Pseudomonadales* within *Gammaproteobacteria* were lower in the fertilizer treatments. The relative abundances of orders *Sphingomonadales* and *Rhizobiales* within *Alphaproteobacteria*, and that of order *Myxococcales* within *Gammaproteobacteria* were higher in the CF but lower in the CFM treatment (**Table 2**) ($P = 0.05$). In addition, the abundances of phyla

TABLE 2 | Relative abundances of selected orders within *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria,* and *Gammaproteobacteria* at 0–20 and 20–90 cm depths under long-term fertilizer treatments.

Bacteria		Soil depth (cm)	Treatment		
Class	**Order**		**CK**	**CF**	**CFM**
Relative abundance (%)					
Alphaproteobacteria	*Sphingomonadales*	0–20	1.70 ± 0.17a[a]A[b]	1.94 ± 0.14aA	1.45 ± 0.02aA
		20–90	1.29 ± 0.02aB	1.31 ± 0.28aB	0.72 ± 0.16bB
	Rhizobiales	0–20	0.57 ± 0.02aA	0.66 ± 0.09aA	0.61 ± 0.05aA
		20–90	0.79 ± 0.05aA	0.97 ± 0.14aA	0.68 ± 0.02bA
Betaproteobacteria	*Nitrosomonadales*	0–20	0.00 ± 0.00bB	0.04 ± 0.01aA	0.03 ± 0.01aA
		20–90	0.02 ± 0.01aA	0.02 ± 0.01aB	0.01 ± 0.00aB
	Methylophilales	0–20	0.00 ± 0.00bB	0.03 ± 0.01aA	0.01 ± 0.00aA
		20–90	0.01 ± 0.00aA	0.01 ± 0.00aB	0.01 ± 0.00aA
	Rhodocyclales	0–20	1.52 ± 0.06aA	0.83 ± 0.11bA	0.79 ± 0.05bA
		20–90	0.47 ± 0.06aB	0.41 ± 0.09aB	0.30 ± 0.08aB
Deltaproteobacteria	*Myxococcales*	0–20	3.90 ± 0.46aA	3.86 ± 0.34aA	3.42 ± 0.03aA
		20–90	3.71 ± 0.25aA	4.36 ± 0.27aA	3.09 ± 0.23bA
	Desulfuromonadales	0–20	0.45 ± 0.06aB	0.65 ± 0.07aA	0.54 ± 0.05aA
		20–90	0.72 ± 0.01aA	0.72 ± 0.09aA	0.70 ± 0.10aA
Gammaproteobacteria	*Xanthomonadales*	0–20	1.37 ± 0.09bB	2.10 ± 0.33aA	1.95 ± 0.07aA
		20–90	2.81 ± 1.44aA	1.46 ± 0.40bB	1.52 ± 0.78bA
	Methylococcales	0–20	0.13 ± 0.02bA	0.34 ± 0.04aA	0.24 ± 0.02aA
		20–90	0.14 ± 0.01aA	0.12 ± 0.03aB	0.07 ± 0.03bB
	Pseudomonadales	0–20	0.08 ± 0.01bB	0.16 ± 0.02aB	0.12 ± 0.01aB
		20–90	0.71 ± 0.56aA	0.37 ± 0.17bA	0.48 ± 0.35bA

Values at 20–90 cm depths are weighted means. CK, no fertilizer; CF, NPK fertilizer; CFM, NPK fertilizer combined with manure. Values in the column are mean ± SE (stand error). [a]*Treatment means within a depth followed by the same lower case letter (a, b, and c) are not significantly different (P > 0.05).*
[b]*Depth means within a column followed by the same upper case letter (A, B, and C) are not significantly different (P > 0.05).*

Verrucomicrobia, *Bacteroidetes*, and *Actinobacteria* (**Figure 2**) were lower in the two fertilizer treatments, and all of them were lowest in the CFM ($P < 0.05$). The relative abundances of *Gemmatimonadetes* and *Nitrospirae* were lower in the CFM and higher in the CF treatment (**Figure 2**) ($P = 0.01$).

Community Structure, Variation, and Determinants

The topsoil samples were separated from subsoil samples in the principal coordinates analysis (PCoA) (**Figure 3**). The analysis of similarities (ANOSIM) further confirmed significant differences ($R = 0.57$, $P < 0.001$) between the topsoil and subsoil. Db-RDA that was used to quantify the impacts of edaphic factors on microbial community composition showed that SOC ($R^2 = 0.91$, $P = 0.006$) was the major factor affecting variance in the bacterial community structure in the topsoil (**Figure 4A**). In addition to SOC, variance was also significantly linked to soil MO ($R^2 = 0.90$, $P = 0.006$) and nitrate–N content ($R^2 = 0.90$, $P = 0.009$). In the subsoil (**Figure 4B**), both soil OP and AK were strongly and significantly linked to community variance (OP: $R^2 = 0.45$, $P < 0.001$; AK: $R^2 = 0.45$, $P < 0.001$).

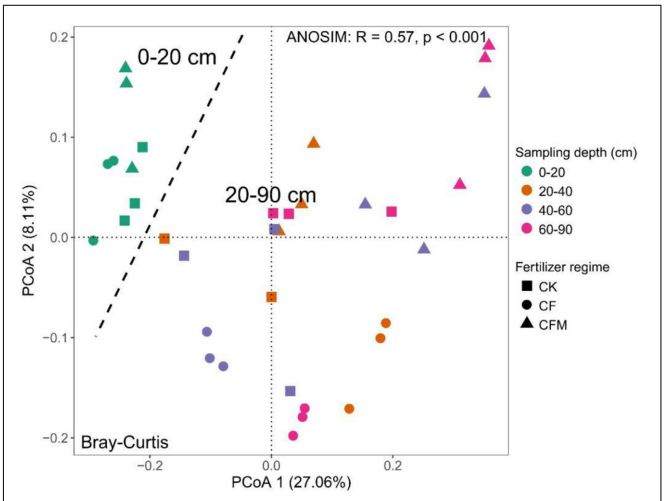


FIGURE 3 | Principal coordinates analysis (PCoA) of the bacterial communities in the soil samples at 0–20 cm, 20–40 cm, 40-60 cm, and 60–90 cm depths under long-term fertilizer treatments. CK, no fertilizer; CF, NPK fertilizer; CFM, NPK fertilizer combined with farmyard manure.

Differential Abundance Analysis

Differential abundance analysis was done to identify OTUs that were affected by fertilization (**Figure 5**). Enriched OTUs (eOTUs) and depleted OTUs (dOTUs) specifically represent OTUs more than two times higher or lower relative abundance ($P < 0.05$) in the long-term fertilization treatments. Altogether there were more OTUs enriched and depleted in CFM compared to CF in

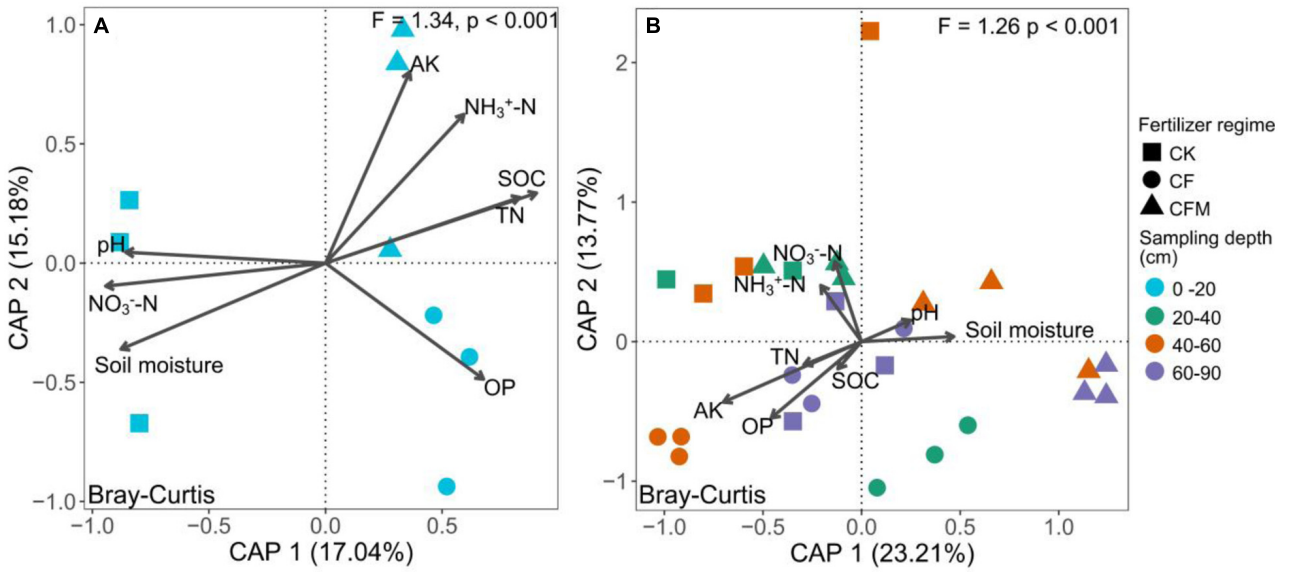


FIGURE 4 | Distance based redundancy analysis (db-RDA, scaling = 2) based on Bray–Curtis dissimilarities of bacterial communities and measured soil properties at 0–20 cm **(A)** and 20–90 cm **(B)** depths under long-term fertilizer treatments. CK, no fertilizer; CF, NPK fertilizer; CFM, NPK fertilizer combined with farmyard manure. See **Figure 1** for the abbreviation of soil properties.

both the topsoil and subsoil. In the topsoil, there were 5 and 16 eOTUs, and 4 and 18 dOTUs in CF and CFM, respectively (**Figures 5A,C** and **Supplementary Data Sheet S1**). In the subsoil, 6 and 42 eOTUs, and 7 and 55 dOTUs were detected in CF and CFM, respectively (**Figures 5B,D** and **Supplementary Data Sheet S1**).

At the phylum level (Supplementary Figure S4A), the eOTUs in CF were mainly identified as *Proteobacteria*, *Bacteroidetes*, and *Gemmatimonadetes*, while those in CFM were *Proteobacteria*, *Acidobacteria*, *Bacteroidetes*, *Nitrospirae*, *Spirochaetes*, and *Thaumarchaeota* in the topsoil. The dOTUs in CF were mainly identified as *Acidobacteria*, *Proteobacteria* and *Verrucomicrobia*, while those in CFM were *Acidobacteria*, *Gemmatimonadetes*, *Proteobacteria*, and *Verrucomicrobia*. In the subsoil, the eOTUs in CF were mainly identified as *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*, while those in CFM were *Chloroflexi*, *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. The dOTUs in CF were *Actinobacteria* and *Bacteroidetes*, while those in CFM were *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, and *Verrucomicrobia* (**Supplementary Data Sheet S1**).

At the genus level (Supplementary Figure S4B), the eOTUs in CF were identified as *Gemmatimonas* and *Bellilinea*, while those in CFM were *Nitrososphaera*, *Nitrospira*, *Dechloromonas*, *Gp4*, *Gp6*, *Gp10*, *Hydrogenophaga*, *Magnetospirillum*, *Bellilinea*, *Sulfuricurvum*, *Thiobacillus,* and *Treponema* in the topsoil. The dOTUs in CF were classified as Subdivision 3 *genera incertae sedis*, while those in CFM were *Gemmatimonas*, *Gp4*, *GP10*, *Gp21*, and Subdivision 3 *genera incertae sedis*. In the subsoil, the eOTUs in CF belonged to *Gp22* and Subdivision 3 *genera incertae sedis*, while those in CFM were Subdivision 3 *genera incertae sedis*, *Gp6*, *Gp9*, *Leptolinea*, and *Bellilinea*. The dOTUs in CF were unidentified, while those in CFM were *Gp3*, *Gp4*, *Gp10*, Subdivision 3 *genera incertae sedis*, *Sideroxydans*, *Steroidobacter*, *Phenylobacterium,* and *Geobacter* (**Supplementary Data Sheet S2**).

Network Associations among Soil Microbial and Soil Properties

To explore the possible ecological interactions within members of microbial communities in top- and subsoil, we inferred two networks based on correlation between differentially abundant OTUs in topsoil and subsoil (**Figure 6**). The network of topsoil captured 37 potential associations among 38 OTUs (**Figure 6A**), while the network of OTUs in subsoil harbored 107 OTUs with 575 potential associations (**Figure 6B**). The basic network level topological features are listed in Supplementary Table S2. In the topsoil network, *Proteobacteria*, *Verrucomicrobia*, *Spirochaetes*, and *Nitrospirae* OTUs showed relatively high degree centrality (Supplementary Figures S5A,C), while several *Proteobacteria* and *Bacteroidetes* OTUs showed higher betweenness centrality compared to other OTUs (Supplementary Figures S5B,C). In the subsoil network, OTUs belonging to *Proteobacteria, Acidobacteria, Verrucomicrobia, Chloroflexi*, and *Bacteroidetes* showed higher degree centrality (Supplementary Figures S6A,C) while several *Proteobacteria*, *Chloroflexi, Bacteroidetes*, and *Verrucomicrobia* OTUs showed relatively higher betweenness centrality (Supplementary Figures S6B,C).

Using greedy modularity optimization algorithm, four and five network modules with more than four nodes were detected in the topsoil and subsoil networks, respectively. The topsoil network modules included *Nitrososphaera*, *Nitrospira*, *Hydrogenophaga*, *Thiobacillus*, *Sulfuricurvum,* and several *Acidobacteria* OTUs

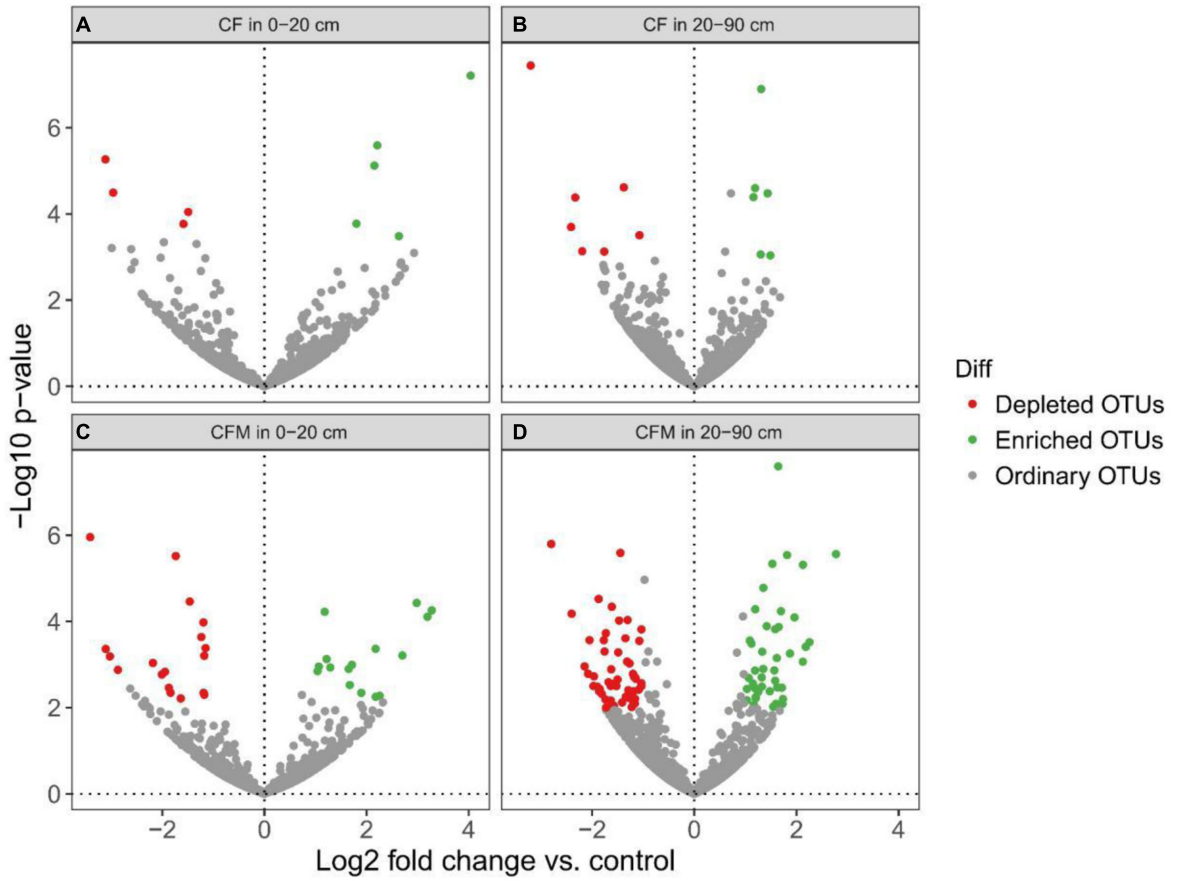


FIGURE 5 | Volcano plots illustrating topsoil and subsoil OTUs that were significantly enriched (green) and depleted (red) by long-term fertilization compared with unfertilized control as determined by differential abundance analysis. Each point represents an individual OTU, and the *Y*-axis indicates the abundance fold change. **(A)** CF vs. control in the topsoil (0–20 cm); **(B)** CF vs. control in the subsoil (20–90 cm); **(C)** CFM vs. control in the topsoil (0–20 cm); **(D)** CFM vs. control in the subsoil (20–90 cm).

(**Supplementary Data Sheet S3**). The subsoil network module M1 contained *Leptolinea* and *Bellilinea* from phylum *Chloroflexi* and *Gp6* from phylum *Acidobacteria*. M2 contained genus *Gp4*, *Gp10* belonged to phylum *Chloroflexi*, *Steroidobacter* from *Proteobacteria*, and Subdivision 3 *genera incertae sedis* from *Verrucomicrobia*. M3 contained genus *Gp22* in *Acidobacteria*, Subdivision 3 *genera incertae sedis* from *Verrucomicrobia*. M4 contained *Sideroxydans* and *Phenylobacterium* from *Proteobacteria*, *Gp3* in *Acidobacteria*, Subdivision 3 *genera incertae sedis* from *Verrucomicrobia*. M5 was composed of only unidentified genus from *Micromonosporaceae* in *Actinobacteria* and *Holophagaceae* in *Acidobacteria* (**Supplementary Data Sheet S4**).

The identification of node sources in each module indicated that almost all nodes in the topsoil network modules were OTUs differentially abundant in CFM (**Supplementary Data Sheet S3**). In the subsoil network, modules M1 and M2 consisted of OTUs differentially abundant in CFM, while M3, M4, and M5 included OTUs differentially abundant in CFM and CF (**Supplementary Data Sheet S4**). Mantel test based on Bray–Curtis dissimilarity indicated that the modules in topsoil network significantly ($p < 0.05$) correlated with multiple soil properties, of which SOC and soil moisture showed highest correlations (**Figure 7A**), suggesting that these two soil properties may be the primary forces in driving the abundance variation of these OTUs in CFM topsoil. In subsoil (**Supplementary Data Sheet S4**), the five modules showed different responses to environmental factors (**Figure 7B**). Modules M1 and M2 derived from CFM treatment showed higher correlations with TN and OP (**Figure 7B**). M3, M4, and M5 showed relatively higher correlations with OP, AK, and TN, respectively (**Figure 7B**).

DISCUSSION

Effects of Depth and Long-Term Fertilization on Soil Physico-Chemical Parameters

We studied how fertilizer treatments affect soil bacterial and archaeal communities and soil properties. Soil physico-chemical parameters were strongly depth-dependent. Similar to earlier

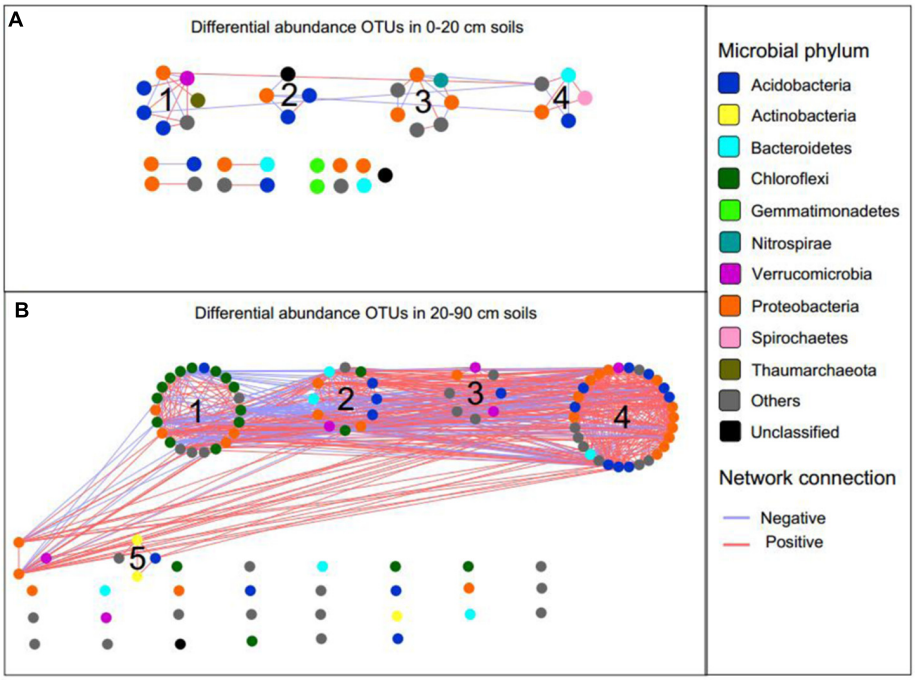


FIGURE 6 | The correlation network of differentially abundant OTUs in topsoil **(A)** and subsoil **(B)**. Blue and red lines represent negative and positive correlation, respectively.

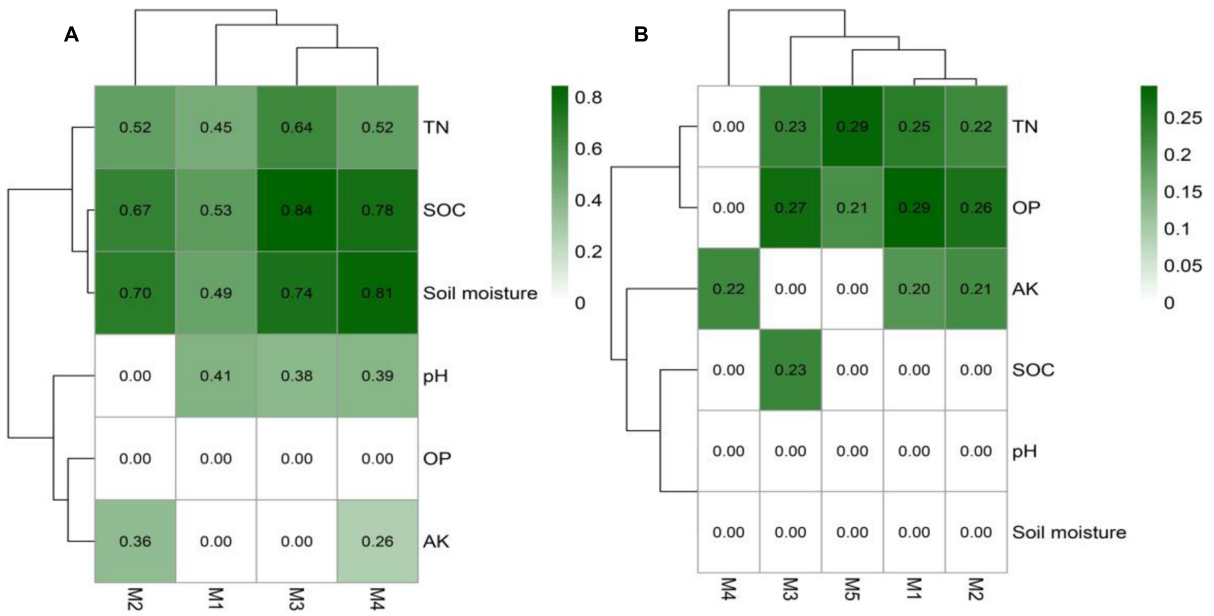


FIGURE 7 | The correlation of modules in the topsoil **(A)** and subsoil **(B)** OTU networks using mantel test based on Bray-Curtis dissimilarity. See **Figure 1** for the abbreviation of soil properties.

studies by Wang and Yang (2003) and Guo et al. (2010), lower pH in the fertilized topsoil might be due to that the fertilization stimulates nitrification and acidification as well as provides more organic acids to the soil. Higher SOC and TN concentrations in the topsoil may be explained by the fact that they are strongly related to root C inputs, and that manure and crop residues often accumulate in the topsoil. As in earlier studies, higher SOC in both CF and CFM treatments in the topsoil could result

in higher soil moisture (Singh and Usha, 2003; Chen et al., 2007). Lower nitrate–N concentration in the subsoil may be due to less aerobic conditions, which result in higher losses of N through denitrification (Roldán et al., 2005), or there are more DNRA (dissimilatory nitrate reduction to ammonium) activity in the subsoil that produces ammonium that can be then assimilated by microorganisms and plants (Burger and Jackson, 2003). Though the application of P containing fertilizer increased the OP concentration, the OP concentration in the subsoil was lower than that in the topsoil which, probably due to the low solubility of phosphorus in slightly alkaline soil (El-Baruni and Olsen, 1979).

Soil properties change with agricultural management practices (Salako et al., 1999). Previous study reported that long-term CF treatments may acidify the soils, similar result was also observed in our study which showed that the soil pH was lowest in the CF soil (Guo et al., 2010). Manure containing fertilizers result in higher SOC content in the soil (Stark et al., 2007; Kuntal et al., 2008). In our study, SOC was higher in both the CF and CFM treatments compared to the control. Moreover, manure improves soil fertility since it contains more different nutrients than CFs (Fan et al., 2005). This was evident in our study where the concentrations of TN and AK were higher in the CFM treatment than in the CF treatment.

Microbial Community Composition in Response to Long-Term Fertilization

Our differential abundance analysis revealed differences in microbial community structure between the CF and CFM treatments in both the topsoil and subsoil. This distinction between topsoil and subsoil should be associated with the direct or indirect effect of fertilization. As we hypothesized, more enriched OTUs belonging to *Nitrospirae*, *Spirochaetes*, *Acidobacteria*, and *Nitrososphaera* in the phylum *Thaumarchaeota* were detected in the soil treated with long-term CFM treatment compared to the CF treatment. *Nitrososphaera* is non-thermophilic Crenarchaeota and capable of oxidizing ammonia (Zhang and He, 2012). *Nitrospirae* played key roles in mediating the nitrite oxidation process in the greenhouse soil, and this result was thought to be attributed to the high substrate affinity of *Nitrospirae* (Freitag et al., 2005; Xia et al., 2010). As *Nitrospirae* phylum is involved in nitrite oxidation (Fierer et al., 2007; Lücker et al., 2010), the enrichment of *Nitrospirae* OTUs in the CFM soil may have had positive effects on nitrification.

The 32 years of CF and CFM fertilizer amendment changed the relative abundances of nitrifying archaea and bacteria in the topsoil and even the subsoil, and there were distinct differences in communities between the CF and CFM treatments. Application of urea fertilizer could lead to the enrichment of NH_4^+–N in the soil due to fast hydrolysis of urea-N, which then leaks downward during irrigation and is gradually re-absorbed in deeper soil layers (Chien et al., 2009). This may favor the growth of ammonia-oxidizing microbes under long-term fertilizer treatments. Our study showed that, comparing to CK, the two fertilizer treatments (CF and CFM) decreased both ammonium–N content and nitrate–N content within the whole profile (**Table 1**), which also indicated the increased activity of ammonia and nitrite oxidizers during fertilizer application. Ammonia and nitrite oxidizers are critical to soil N cycling (Fierer et al., 2007), and the application N containing fertilizer in the current study was likely to result in different composition and structure of these communities.

Changes in Taxa Abundances with Soil Depth

The subsoil microbial communities were distinct in composition and structure from topsoil communities (Fierer et al., 2003a,b). Based on our differential abundance analysis, primarily OTUs from *Proteobacteria*, *Acidobacteria*, and *Gemmatimonadetes* were enriched in the topsoil, while *Chloroflexi* and *Actinobacteria* were enriched in the subsoil. *Betaproteobacteria* is considered as copiotrophic bacteria that flourish in soils with enriched nutrients (Fierer et al., 2007), which probably explains why mostly *Betaproteobacteria* OTUs were enriched in the topsoil. *Acidobacteria* include many oligotrophic members (Nemergut et al., 2010; Pascault et al., 2013), yet *Acidobacteria* like *Gp4* and *Gp6* were abundant in soils with higher contents of soil C whereas *Gp1* and *Gp7* were not (Liu et al., 2014; Li et al., 2017). Similarly, in our study both enriched and depleted *Acidobacteria* were detected in the soil profile, and the relative abundance of *Acidobacteria* was higher in the topsoil than in the subsoil. *Chloroflexi* are facultative anaerobic and have a recognized role as heterotrophic oligotrophs in soils, having the ability to survive on recalcitrant plant polymers (Yabe et al., 2010; Hug et al., 2013; King, 2014). The relative abundance of *Chloroflexi* was relatively lower compared to other studies (Zhao et al., 2014; Chen et al., 2016), but similar to the levels reported by Li et al. (2014). Their enrichment in the subsoil confirmed their adaptation to growth under nutrient limitation (Engel et al., 2010; Hug et al., 2013; Barton et al., 2014).

At the genus level the most enriched OTUs in the topsoil were *Ohtaekwangia*, *Gemmatimonas,* and *Nitrospira*. Members of the *Ohtaekwangia* genus, potential petroleum hydrocarbon degraders, were widely distributed in plant rhizosphere (Zhang et al., 2011; McGenity et al., 2012; Tejeda-Agredano et al., 2013; Gaggìa et al., 2013). *Ohtaekwangia* may contribute to the transformation of soil carbon derived from the plant (Naumoff and Dedysh, 2012), possibly explaining the increase of their relative abundance. *Gemmatimonas* species were able to modulate carbon and nitrogen intake according to their metabolic needs under various conditions and were abundant in soils amended with pyrogenic organic material (Carbonetto et al., 2014; Xu et al., 2014; Whitman et al., 2016), indicating that they are likely to decompose polyaromatic carbon compounds. *Gemmatimonas* species accumulated polyphosphate and were stimulated by P fertilizer (Zhang et al., 2003; Su et al., 2015). Members of *Nitrospira* are mostly uncultured nitrite-oxidizing bacteria (Palomo et al., 2016). In this study, *Nitrospira* OTUs were more relevant to the CFM than to the CF treatment in the topsoil, probably due to the lower C/N ratio of manure, which mobilizes soil N, increases its mineralization and the availability of mineral nitrogen (Leite et al., 2017).

The most enriched OTUs in the subsoil were genus Subdivision 3 *genera incertae sedis, Leptolinea, Bellilinea*, and some subgroups of *Acidobacteria*. Subdivision 3 *genera incertae sedis* is affiliated to the phylum *Verrucomicrobia*, which was negatively correlated with soil fertility and increased in abundance with increasing available N, P, K, and SOC obtained from cotton straw (Huang et al., 2012; Navarrete et al., 2015). Genus *Leptolinea* and *Bellilinea* belong to class *Anaerolineae* in the oligotrophic *Chloroflexi* subphylum I, and needed to associate with other microbes to grow efficiently (Narihiro et al., 2012). Taken together, our results showed that specific microbial taxa involved in decomposition of organic compounds and in C, N, and P transformations were substantially enriched or depleted in the top- and subsoil by long-term fertilization.

OTU Association Networks in Top- and Subsoil

Since C and N are essential nutrients for microbial growth, soil C and N are expected to show strong associations with taxa affected directly or indirectly by long-term fertilization in different soil depths. Our co-occurrence based network analysis revealed that in the topsoil SOC and soil moisture had strong positive correlations with microbial taxa, e.g., genus *Nitrososphaera* from phylum *Thaumarchaeota*, *Nitrospira* of *Nitrospirae*, *Hydrogenophaga*, and *Thiobacillus and Sulfuricurvum* from *Proteobacteria*, and several subgroups of *Acidobacteria*. In the subsoil soil OP and TN showed strong positive correlations with different genera belonging to phyla *Chloroflexi*, *Acidobacteria*, *Proteobacteria*, and *Verrucomicrobia* that participate in soil nutrient transformations as discussed above. Betweenness centrality is linked to the importance of the control potential an OTU exerts over the interactions of other OTUs in that network. OTUs with high betweenness centrality values might possess high impact on other interactions in the community (Greenblum et al., 2012). Keystone nodes in co-occurrence networks tend to have maximum betweenness centrality values (Vick-Majors et al., 2015; Banerjee et al., 2016). Based on betweenness centrality, most keystone bacterial nodes in our study belonged to the most abundant phyla *Proteobacteria* and *Bacteroidetes* in both top- and subsoil. Within the *Proteobacteria*, *Hydrogenophaga* was identified as the keystone species in the topsoil, while genus *Phenylobacterium* and *Steroidobacter* were identified as keystones in the subsoil. These genera have positive effects on soil nutrient cycling and have been reported as beneficial microorganisms in soil. Genus *Hydrogenophaga* belongs to order *Burkholderiales*, members of which have shown ability to decompose degrade high molecular weight organic compounds and utilize sulfanilic acid as the sole carbon and energy source (Ding et al., 2012). Genus *Phenylobacterium* and *Steroidobacter* belong to *Caulobacterales* and *Xanthomonadales*, respectively. *Phenylobacterium* species were reported to be favored by a compound in the organic material applied or metabolites available at a specific stage in the degradation process (Cruz-Barrón et al., 2017). Genus *Steroidobacter* in SOC-poor soil could be enriched by the application of soybean residues and was thought to be associated with C and N cycling as this genus utilizes only a narrow range of organic substrates with nitrate as the electron acceptor (Lian et al., 2017). Similarly, the genus *Steroidobacter* was found to be one keystone in the subsoil in this study. As the concentration of soil total nitrogen and SOC in the subsoil was relative lower than those in the topsoil (**Figure 1**), *Steroidobacter* may contribute to the SOC decomposition for N demand in nutrient poor soil (Lian et al., 2017).

In summary, this study demonstrated that the 32-year long fertilization treatments not only affected the soil physical–chemical parameters but also the composition of the bacterial and archaeal communities both in the top- and subsoil. Our previous study revealed that yields of rice and wheat were comparable and occasionally higher under CFM than under CF treatment. *Proteobacteria* and *Acidobacteria* were the most abundant phyla in topsoil. Compared to CF treatment, different bacterial taxa were enriched under CFM treatment in both topsoil and subsoil. These taxa are widely distributed in soil and involved in soil nutrient transformations. Taken together, the spatial variability of soil properties due to long-term fertilization strongly shaped the bacterial and archaeal community composition and their interactions at both high and low taxonomic levels and resulted in distinct association patterns among the specific taxa in topsoil and subsoil.

AUTHOR CONTRIBUTIONS

YG, ST, and XZ designed the study. YG, QX, and KZ analyzed the data and wrote the manuscript. SL, XY, and LZ collected and analyzed soil samples. ST and QC managed the long-term field experiment. All authors reviewed the manuscript.

ACKNOWLEDGMENT

The study was funded by the Natural Science Foundation of China (grant No. 41201256).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01516/full#supplementary-material

DATA SHEET S1 | All the significantly differential OTUs at the phylum level within the whole soil profile.

DATA SHEET S2 | All the significantly differential OTUs at the genus level within the whole soil profile.

DATA SHEET S3 | Differential type of network nodes in the topsoil (0–20 cm).

DATA SHEET S4 | Differential type of network nodes in the subsoil (20–90 cm).

REFERENCES

Banerjee, S., Baah-Acheamfour, M., Carlyle, C. N., Bissett, A., Richardson, A. E., Siddique, T., et al. (2016). Determinants of bacterial communities in Canadian agroforestry systems. *Environ. Microbiol.* 18, 1805–1816.

Bao, X. L., Yu, J., Liang, W. J., Lu, C. Y., Zhu, J. G., and Li, Q. (2015). The interactive effects of elevated ozone and wheat cultivars on soil microbial community composition and metabolic diversity. *Appl. Soil. Ecol.* 87, 11–18. doi: 10.1016/j.apsoil.2014.11.003

Barton, H. A., Giarrizzo, J. G., Suarez, P., Robertson, C. E., Broering, M. J., and Banks, E. D. (2014). Microbial diversity in a venezuelan orthoquartzite cave is dominated by the *chloroflexi* (class *ktedonobacterales*) and *thaumarchaeota* group i.1c. *Front. Microbiol.* 5:615. doi: 10.3389/fmicb.2014.00615

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* 57, 289–300.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 491–507. doi: 10.1093/biomet/93.3.491

Birkhofer, K., Bezemer, T. M., Bloem, J., Bonkowski, M., Christensen, S., Dubois, D., et al. (2008). Long-term organic farming fosters below and aboveground biota: implications for soil quality, biological control and productivity. *Soil Biol. Biochem.* 40, 2297–2308. doi: 10.1016/j.soilbio.2008.05.007

Burger, M., and Jackson, L. E. (2003). Microbial immobilization of ammonium and nitrate in relation to ammonification and nitrification rates in organic and conventional cropping systems. *Soil Biol. Biochem.* 35, 29–36. doi: 10.1016/S0038-0717(02)00233-X

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Carbonetto, B., Rascovan, N., Álvarez, R., Mentaberry, A., and Vázquez, M. P. (2014). Structure, composition and metagenomic profile of soil microbiomes associated to agricultural land use and tillage systems in Argentine pampas. *PLoS ONE* 9:e99949. doi: 10.1371/journal.pone.0099949

Cederlund, H., Wessen, E., Enwall, K., Jones, C. M., Juhanson, J., Pell, M., et al. (2014). Soil carbon quality and nitrogen fertilization structure bacterial communities with predictable responses of major bacterial phyla. *Appl. Soil Ecol.* 84, 62–68. doi: 10.1016/j.apsoil.2014.06.003

Chen, L. D., Gong, J., Fu, B. J., Huang, Z. L., Huang, Y. L., and Gui, L. D. (2007). Effect of land use conversion on soil organic carbon sequestration in the loess hilly area, loess plateau of China. *Ecol. Res.* 22, 641–648. doi: 10.1007/s10661-014-4131-9

Chen, Q. L., An, X. L., Li, H., Su, J. Q., Ma, Y. B., and Zhu, Y. G. (2016). Long-term field application of sewage sludge increases the abundance of antibiotic resistance genes in soil. *Environ. Int.* 9, 1–10. doi: 10.1016/j.envint.2016.03.026

Chien, S. H., Prochnow, L. I., and Cantarella, H. (2009). Chapter 8 recent developments of fertilizer production and use to improve nutrient efficiency and minimize environmental impacts. *Adv. Agron.* 102, 267–322. doi: 10.1016/S0065-2113(09)01008-6

Cruz-Barrón, M. D., Cruz-Mendoza, A., Navarro–Noya, Y. E., Ruiz-Valdiviezo, V. M., Ortíz-Gutiérrez, D., Ramírez-Villanueva, D. A., et al. (2017). The bacterial community structure and dynamics of carbon and nitrogen when maize (*Zea mays* L.) and its neutral detergent fibre were added to soil from Zimbabwe with contrasting management practices. *Microb. Ecol.* 73, 135–152. doi: 10.1007/s00248-016-0807-8

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Syst.* 1695, 1–9.

Ding, G. C., Heuer, H., and Smalla, K. (2012). Dynamics of bacterial communities in two unpolluted soils after spiking with phenanthrene: soil type specific and common responders. *Front. Microbiol.* 3:290. doi: 10.3389/fmicb.2012.00290

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Eilers, K. G., Debenport, S., Anderson, S., and Fierer, N. (2012). Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biol. Biochem.* 50, 58–65. doi: 10.1016/j.soilbio.2012.03.011

El-Baruni, B., and Olsen, S. R. (1979). Effect of manure on solubility of phosphorus in calcareous soils. *Soil Sci.* 128, 219–225. doi: 10.1097/00010694-197910000-00005

Engel, A. S., Meisinger, D. B., Porter, M. L., Payn, R. A., Schmid, M., Stern, L. A., et al. (2010). Linking phylogenetic and functional diversity to nutrient spiraling in microbial mats from Lower Kane Cave (USA). *ISME J.* 4, 98–110. doi: 10.1038/ismej.2009.91

Fan, T. L., Stewart, B. A., Wang, Y., Lou, J. J., and Zhou, G. Y. (2005). Long-term fertilization effects on grain yield, water-use efficiency and soil fertility in the dry land of Loess Plateau in China. *Agric. Ecosyst. Environ.* 106, 313–329. doi: 10.1016/j.agee.2004.09.003

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832

Fierer, N., Allen, A. S., Schimel, J. P., and Holden, P. A. (2003a). Controls on microbial CO_2 production: a comparison of surface and subsurface soil horizons. *Global. Change Biol.* 9, 1322–1332. doi: 10.1046/j.1365-2486.2003.00663.x

Fierer, N., Bradford, M. A., and Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology* 88, 1354–1364. doi: 10.1890/05-1839

Fierer, N., Schimel, J. P., and Holden, P. A. (2003b). Variations in microbial community composition through two soil depth profiles. *Soil Biol. Biochem.* 35, 167–176.

Freitag, T. E., Chang, L., Clegg, C. D., and Prosser, J. I. (2005). Influence of inorganic nitrogen management regime on the diversity of nitrite-oxidizing bacteria in agricultural grassland soils. *Appl. Environ. Microbiol.* 71, 8323–8334. doi: 10.1128/AEM.71.12.8323-8334.2005

Gaggia, F., Baffoni, L., Di Gioia, D., Accorsi, M., Bosi, S., Marotti, I., et al. (2013). Inoculation with microorganisms of *Lolium perenne* L evaluation of plant growth parameters and endophytic colonization of roots. *New Biotechnol.* 30, 695–704. doi: 10.1016/j.nbt.2013.04.006

Greenblum, S., Turnbaugh, P. J., and Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. U.S.A.* 10, 594–599. doi: 10.1073/pnas.1116053109

Gu, Y. F., Zhang, X. P., Tu, S. H., and Lindström, K. (2008). Soil microbial biomass, crop yields, and bacterial community structure as affected by long-term fertilizer treatments under wheat-rice cropping. *Eur. J. Soil Biol.* 45, 239–246. doi: 10.1016/j.ejsobi.2009.02.005

Guo, J. H., Liu, X. J., Zhang, Y., Shen, J. L., Han, W. X., Zhang, W. F., et al. (2010). Significant acidification in major Chinese croplands. *Science* 327, 1008–1010. doi: 10.1126/science.1182570

Hartman, W. H., Richardson, C. J., Vilgalys, R., and Bruland, G. L. (2008). Environmental and anthropogenic controls over bacterial communities in wetland soils. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17842–17847. doi: 10.1073/pnas.0808254105

He, S. B., Guo, L. X., Niu, M. Y., Miao, F. H., Jiao, S., Hu, T. M., et al. (2017). Ecological diversity and co-occurrence patterns of bacterial community through soil profile in response to long-term switchgrass cultivation. *Sci. Rep* 7, 3608. doi: 10.1038/s41598-017-03778-7

Huang, W. R., Bai, Z. H., Hoefel, D., Hu, Q., Lv, X., Zhuang, G. Q., et al. (2012). Effects of cotton straw amendment on soil fertility and microbial communities. *Front. Environ. Sci. Eng.* 6:336–349. doi: 10.1007/s11783-011-0337-z

Hug, L. A., Castelle, C. J., Wrighton, K. C., Thomas, B. C., Sharon, I., and Frischkorn, K. R. (2013). Community genomic analyses constrain the distribution of metabolic traits across the *chloroflexi* phylum and indicate roles in sediment carbon cycling. *Microbiome* 1:22. doi: 10.1186/2049-2618-1-22

King, G. M. (2014). Urban microbiomes and urban ecology: how do microbes in the built environment affect human sustainability in cities? *J. Microbiol.* 52, 721–728. doi: 10.1007/s12275-014-4364-x

Kumar, U., Shahid, M., Tripathi, R., Mohanty, S., Kumar, A., Bhattacharyya, P., et al. (2017). Variation of functional diversity of soil microbial community in sub-humid tropical rice-rice cropping system under long-term organic and inorganic fertilization. *Ecol. Indic.* 73, 536–543. doi: 10.1016/j.ecolind.2016.10.014

Kuntal, M. H., Anand, S., Mishra, B., Manna, M. C., Wanjari, R. H., Mandal, K. G., et al. (2008). Impact of long-term application of fertilizer, manure and lime under intensive cropping on physical properties and organic carbon content of an Alfisol. *Geoderma* 148, 173–179. doi: 10.1016/j.geoderma.2008.09.015

Legendre, P., and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716

Leite, M. F., Pan, Y., Bloem, J., ten Berge, H., and Kuramae, E. E. (2017). Organic nitrogen rearranges both structure and activity of the soil-borne microbial seedbank. *Sci. Rep* 7:42634. doi: 10.1038/srep42634

Li, C. H., Yan, K., Tang, L. S., Jia, Z. J., and Li, Y. (2014). Change in deep soil microbial communities due to long-term fertilization. *Soil Biol. Biochem.* 75, 264–272. doi: 10.1016/j.soilbio.2014.04.023

Li, F., Chen, L., Zhang, J. B., Yin, J., and Huang, S. M. (2017). Bacterial community structure after long-term organic and inorganic fertilization reveals important associations between soil nutrients and specific taxa involved in nutrient transformations. *Front. Microbiol.* 8:187. doi: 10.3389/fmicb.2017.00187

Lian, T. X., Jin, J., Wang, G. H., Tang, C. X., Yu, Z. H., Li, Y. S., et al. (2017). The fate of soybean residue-carbon links to changes of bacterial community composition in Mollisols differing in soil organic carbon. *Soil Biol. Biochem.* 109, 50–58. doi: 10.1016/j.soilbio.2017.01.026

Liu, J. J., Sui, Y. Y., Yu, Z. H., Shi, Y., Chu, H. Y., Jin, J., et al. (2014). High throughput sequencing analysis of biogeographical distribution of bacterial communities in the black soils of northeast China. *Soil Biol. Biochem.* 70, 113–122. doi: 10.1016/j.soilbio.2013.12.014

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome. Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Lu, R. K. (1999). *Soil and Agro-Chemical Analytical Methods*. Beijing: China Agricultural Science and Technology Press, 146–195.

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., and Vacherie, B. (2010). A nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13479–13484. doi: 10.1073/pnas.1003860107

Luo, F., Zhong, J., Yang, Y., Scheuermann, R. H., and Zhou, J. Z. (2006). Application of random matrix theory to biological networks. *Phys. Lett. A* 357, 420–423. doi: 10.1016/j.physleta.2006.04.076

Madsen, E. (1995). Impacts of agricultural practices on subsurface microbial ecology. *Adv. Agron.* 54, 1–67. doi: 10.1016/S0065-2113(08)60897-4

Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Marschner, P., Kandeler, E., and Marschner, B. (2003). Structure and function of the soil microbial community in a long-term fertilizer experiment. *Soil Biol. Biochem.* 35, 453–461. doi: 10.1007/s00248-008-9372-0

McGenity, T. J., Folwell, B. D., McKew, B. A., and Sanni, G. O. (2012). Marine crude-oil biodegradation: a central role for interspecies interactions. *Aquat. Biosyst.* 8:10. doi: 10.1186/2046-9063-8-10

Narihiro, T., Terada, T., Ohashi, A., Kamagata, Y., Nakamura, K., and Sekiguchi, Y. (2012). Quantitative detection of previously characterized syntrophic bacteria in anaerobic wastewater treatment systems by sequence-specific rRNA cleavage method. *Water Res.* 46, 2167–2175. doi: 10.1016/j.watres.2012.01.034

Naumoff, D. G., and Dedysh, S. N. (2012). Lateral gene transfer between the *Bacteroidetes* and *Acidobacteria*: the case of α-L-Rhamnosidases. *FEBS Lett.* 586, 3843–3851. doi: 10.1016/j.febslet.2012.09.005

Navarrete, A. A., Soares, T., Rossetto, R., van Veen, J. A., Tsai, S. M., and Kuramae, E. E. (2015). *Verrucomicrobial* community structure and abundance as indicators for changes in chemical factors linked to soil fertility. *Antonie Van Leeuwenhoek* 108, 741–752. doi: 10.1007/s10482-015-0530-3

Nemergut, D. R., Cleveland, C. C., Wieder, W. R., Washenberger, C. L., and Townsend, A. R. (2010). Plot-scale manipulations of organic matter inputs to soils correlate with shifts in microbial community composition in a lowland tropical rain forest. *Soil Biol. Biochem.* 42, 2153–2160. doi: 10.1016/j.soilbio.2010.08.011

Newman, M. (2006). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74, 036104. doi: 10.1103/PhysRevE.74.036104

Nowinski, N. S., Trumbore, S. E., Schuur, E. A. G., Mack, M. C., and Shaver, G. R. (2008). Nutrient addition prompts rapid destabilization of organic matter in an Arctic tundra ecosystem. *Ecosystems* 11, 16–25. doi: 10.1007/s10021-007-9104-1

Oksanen, J., Kindt, R., Pierre, L., O'Hara, B., Simpson, G. L., Solymos, P., et al. (2013). *Vegan: Community Ecology Package, R Package Version 2.0*. Available at: http://vegan.r-forge.r-project.org

Palomo, A., Jane Fowler, S., Gülay, A., Rasmussen, S., Sicheritz-Ponten, T., and Smets, B. F. (2016). Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of *Nitrospira* spp. *ISME J.* 10, 2569–2581. doi: 10.1038/ismej.2016.63

Pascault, N., Ranjard, L., Kaisermann, A., Bachar, D., Christen, R., and Terrat, S. (2013). Stimulation of different functional groups of bacteria by various plant residues as a driver of soil priming effect. *Ecosystems* 16, 810–822. doi: 10.1007/s10021-013-9650-7

Philippot, L., Andersson, S. G., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., et al. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat. Rev. Microbiol.* 8, 523–529. doi: 10.1038/nrmicro2367

Pollard, K. S., Gilbert, H. N., Ge, Y., Taylor, S., and Dudoit, S. (2013). *Multtest: Resampling-Based Multiple Hypothesis Testing. R Package Version* 2.17.0.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna: The R Foundation for Statistical Computing.

Ramirez, K. S., Lauber, C. L., Knight, R., Bradford, M. A., and Fierer, N. (2010). Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology* 91, 3463–3470. doi: 10.1890/10-0426.1

Roldán, A., Salinas-García, J. R., Alguacil, M. M., Díaz, E., and Caravaca, F. (2005). Soil enzyme activities suggest advantages of conservation tillage practices in sorghum cultivation under subtropical conditions. *Geoderma* 129, 178–185. doi: 10.1016/j.geoderma.2004.12.042

Salako, F. K., Babalola, O., Hauser, S., and Kang, B. T. (1999). Soil macroaggregate stability under different fallow management systems and cropping intensities in southwestern Nigeria. *Geoderma* 91, 103–123. doi: 10.1016/S0016-7061(99)00006-3

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Singh, B., and Usha, K. (2003). Salicylic acid induced physiological and biochemical changes in wheat seedlings under water stress. *Plant Growth Regul.* 39, 137–141. doi: 10.1023/A:1022556103536

Smit, E., Leeflang, P., Gommans, S., van den Broek, J., van Mil, S., and Wernars, K. (2001). Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Appl. Environ. Microbiol.* 67, 2284–2291. doi: 10.1128/AEM.67.5.2284-2291.2001

Stark, C., Condron, L. M., Stewart, A., Di, H. J., and O'Callaghan, M. (2007). Influence of organic and mineral amendments on microbial soil properties and processes. *Appl. Soil Ecol.* 35, 79–93. doi: 10.1016/j.apsoil.2006.05.001

Stowe, D. C., Lamhamedi, M. S., Carles, S., Fecteau, B., Margolis, H. A., Renaud, M., et al. (2010). Managing irrigation to reduce nutrient leaching in containerized white spruce seedling production. *New For.* 40, 185–204. doi: 10.1007/s11056-010-9193-0

Su, J. Q., Ding, L. J., Xue, K., Yao, H. Y., Quensen, J., Bai, S. J., et al. (2015). Long-term balanced fertilization increases the soil microbial functional diversity in a phosphorus-limited paddy soil. *Mol. Ecol.* 24, 136. doi: 10.1111/mec.13010

Tejeda-Agredano, M. C., Gallego, S., Vila, J., Grifoll, M., Ortega-Calvo, J. J., and Cantos, M. (2013). Influence of the sunflower rhizosphere on the biodegradation of PAHs in soil. *Soil Biol. Biochem.* 57, 830–840. doi: 10.1016/j.soilbio.2012.08.008

Vick-Majors, T. J., Achberger, A., Santibáñez, P., Dore, J. E., Hodson, T., and Michaud, A. B. (2015). Biogeochemistry and microbial diversity in the marine cavity beneath the mcmurdo ice shelf, antarctica. *Limnol. Oceanogr.* 61, 572–586. doi: 10.1002/lno.10234

Wang, M. C., and Yang, C. H. (2003). Type of fertilizer applied to a paddy–upland rotation affects selected soil quality attributes. *Geoderma* 114, 93–108. doi: 10.1016/S0016-7061(02)00356-7

Whitman, T., Pepe-Ranney, C., Enders, A., Koechli, C., Campbell, A., Buckley, D. H., et al. (2016). Dynamics of microbial community composition and

soil organic carbon mineralization in soil following addition of pyrogenic and fresh organic matter. *ISME J* 10, 2918–2930. doi: 10.1038/ismej.2016.68

Will, C., Thürmer, A., Wollherr, A., Nacke, H., Herold, N., Schrumpf, M., et al. (2010). Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. *Appl. Environ. Microbiol.* 76, 6751–6759. doi: 10.1128/AEM.01063-10

Xia, S., Li, J., Wang, R., Li, J., and Zhang, Z. (2010). Tracking composition and dynamics of nitrification and denitrification microbial community in a biofilm reactor by PCR-DGGE and combining fish with flow cytometry. *Biochem. Eng. J.* 49, 370–378. doi: 10.1016/j.bej.2010.01.013

Xu, X., Kim, J. Y., Oh, Y. R., and Park, J. M. (2014). Production of biodiesel from carbon sources of macroalgae, *Laminaria japonica*. *Bioresour. Technol.* 169, 455–461. doi: 10.1016/j.biortech.2014.07.015

Yabe, S., Aiba, Y., Sakai, Y., Hazaka, M., and Yokota, A. (2010). *Thermosporothrix hazakensis* gen. nov., sp. nov., isolated from compost, description of *Thermosporotrichaceae fam*. nov. within the class *Ktedonobacteria Cavaletti* et al., 2007 and emended description of the class *Ktedonobacteria*. *Int. J. Syst. Evol. Microbiol.* 60, 1794–1801. doi: 10.1099/ijs.0.018069-0

Yu, Y., Wei, W., Chen, L. D., Jia, F. Y., Yang, L., Zhang, H. D., et al. (2015). Responses of vertical soil moisture to rainfall pulses and land uses in a typical loess hilly area, China. *Solid Earth* 6, 595–608. doi: 10.5194/se-6-595-2015

Zhang, F., and He, Z. (2012). Simultaneous nitrification and denitrification with electricity generation in dual-cathode microbial fuel cells. *J. Chem. Technol. Biotechnol.* 87, 153–159. doi: 10.1002/jctb.2700

Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 21, 660–666. doi: 10.1038/nbt827

Zhang, Y., Du, B. H., Jin, Z. G., Li, Z. H., Song, H. N., and Ding, Y. Q. (2011). Analysis of bacterial communities in rhizosphere soil of healthy and diseased cotton (*Gossypium* sp.) at different plant growth stages. *Plant Soil* 339, 447–455. doi: 10.1007/s11104-010-0600-2

Zhao, J., Ni, T., Li, Y., Xiong, W., Ran, W., Shen, B., et al. (2014). Responses of bacterial communities in arable soils in a rice-wheat cropping system to different fertilizer regimes and sampling times. *PLoS ONE* 9:e85301. doi: 10.1371/journal.pone.0085301

Zhou, J. Z., Xue, K., Xie, J. P., Deng, Y., Wu, L. Y., Cheng, X. L., et al. (2012). Microbial mediation of carbon-cycle feedbacks to climate warming. *Nat. Clim. Chang* 2, 106–110. doi: 10.1038/nclimate1331

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Gu, Wang, Lu, Xiang, Yu, Zhao, Zou, Chen, Tu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 March 2018
doi: 10.3389/fmicb.2018.00484

Maize Endophytic Bacterial Diversity as Affected by Soil Cultivation History

David Correa-Galeote[1]*, *Eulogio J. Bedmar*[1] *and Gregorio J. Arone*[2]

[1] *Department of Soil Microbiology and Symbiotic Systems, Estación Experimental del Zaidín, Agencia Estatal Consejo Superior de Investigaciones Científicas, Granada, Spain,* [2] *Department of Agricultural Sciences, National University of Huancavelica, Huancavelica, Peru*

The bacterial endophytic communities residing within roots of maize (*Zea mays* L.) plants cultivated by a sustainable management in soils from the Quechua maize belt (Peruvian Andes) were examined using tags pyrosequencing spanning the V4 and V5 hypervariable regions of the 16S rRNA. Across four replicate libraries, two corresponding to sequences of endophytic bacteria from long time maize-cultivated soils and the other two obtained from fallow soils, 793 bacterial sequences were found that grouped into 188 bacterial operational taxonomic units (OTUs, 97% genetic similarity). The numbers of OTUs in the libraries from the maize-cultivated soils were significantly higher than those found in the libraries from fallow soils. A mean of 30 genera were found in the fallow soil libraries and 47 were in those from the maize-cultivated soils. Both alpha and beta diversity indexes showed clear differences between bacterial endophytic populations from plants with different soil cultivation history and that the soils cultivated for long time requires a higher diversity of endophytes. The number of sequences corresponding to main genera *Sphingomonas, Herbaspirillum*, *Bradyrhizobium* and *Methylophilus* in the maize-cultivated libraries were statistically more abundant than those from the fallow soils. Sequences of genera *Dyella* and *Sreptococcus* were significantly more abundant in the libraries from the fallow soils. Relative abundance of genera *Burkholderia, candidatus* Glomeribacter, *Staphylococcus*, *Variovorax, Bacillus* and *Chitinophaga* were similar among libraries. A canonical correspondence analysis of the relative abundance of the main genera showed that the four libraries distributed in two clearly separated groups. Our results suggest that cultivation history is an important driver of endophytic colonization of maize and that after a long time of cultivation of the soil the maize plants need to increase the richness of the bacterial endophytes communities.

Keywords: biodiversity, endophytes, maize, Quechua region, 16S rRNA, pyrosequencing, PGPR bacteria

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Ravindra Soni, Indira Gandhi Agricultural University, India
José A. Herrera-Cervera, University of Granada, Spain

***Correspondence:**
David Correa-Galeote
correa@ccg.unam.mx

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *10 October 2017*
Accepted: *28 February 2018*
Published: *16 March 2018*

Citation:
Correa-Galeote D, Bedmar EJ and Arone GJ (2018) Maize Endophytic Bacterial Diversity as Affected by Soil Cultivation History. Front. Microbiol. 9:484. doi: 10.3389/fmicb.2018.00484

INTRODUCTION

Chacras are small (200–10,000 m^2) plots in the Quechua region of the Peruvian Andes where maize, pea, wheat, potatoes and other vegetables and cereals are cultivated by the native peasants. Commodities that they use for their own consumption and for trade in local and regional markets. Genetic and archeological data indicate that after domestication in Mexico about 8,700 years before the present (cal. BP) (Piperno et al., 2009; Van Heerwaarden et al., 2011; Grobman et al., 2012), maize spread in other Mexican regions and into south of Mexico, reaching the southern Andean highlands by 4,000 before the present (Perry et al., 2006). Since then maize is the staple diet of

the Quechua natives who continue growing it as their ancestors did. Mostly without chemical fertilization, no pesticide application and without irrigation and yet chacras maintain a sustainable production for years.

Bacterial endophytes have been defined as microorganisms that could be isolated from surface-sterilized plant tissues and do not visibly harm host plants (Petrini, 1991; Hallmann et al., 1997; Schulz and Boyle, 2006). It is now considered that endophytism is a universal phenomenon (Kobayashi and Palumbo, 2000) and is likely that all plants harbor endophyte bacteria (Rosenblueth and Martínez-Romero, 2006; Ryan et al., 2008; Compant et al., 2010; Dudeja and Giri, 2014). Endophytes have been involved in plant growth promotion, biological control of plant pathogens, isolation of compounds of pharmaceutical or biotechnological interest (reviewed in Schulz, 2006; Weyens et al., 2009; Li et al., 2012; Malfanova et al., 2013; Hardoim et al., 2015; Ma et al., 2016; Vejan et al., 2016; Sharma et al., 2017).

Previous studies have analyzed bacterial taxa associated with maize. Most of the work has been done by using culture-dependent methods (Rai et al., 2007; Rijavec et al., 2007; Pereira et al., 2009, 2011; Ikeda et al., 2013; Celador-Lera et al., 2016; Menéndez et al., 2016; Sandhya et al., 2017) or assessed bacterial diversity independently of culture approaches (Schmalenberger and Tebbe, 2003; Herschkovitz et al., 2005a,b; Pereira et al., 2011; Correa-Galeote et al., 2016; Liu et al., 2017). The high-throughput pyrosequencing technology introduced by 454 Life Science (Margulies et al., 2005; Rothberg and Leamon, 2008) has been used to assess diversity in cultivar-specific bacterial endophyte communities in potato roots (Manter et al., 2010), leaf vegetables (Jackson et al., 2013), the spermosphere and phyllosphere of spinach (López-Velasco et al., 2013), tomato leaves (Romero et al., 2014), grapevine leaves and stems (Yousaf et al., 2014) and roots and shoots of cucumber (Eevers et al., 2016).

Most of the works has been done toward understand the effect of an engineered varieties of maize in the bacterial diversity in maize rhizosphere (Schmalenberger and Tebbe, 2003; Ikeda et al., 2013; Liu et al., 2017) or to analyse the changes in the rhizosphere of maize after the application of an inoculant (Herschkovitz et al., 2005a,b; Sanguin et al., 2006a,b; Alves et al., 2015). In recent years, sustainability agriculture methods have also been considered as great potential source of new information and perspective in agriculture and food systems (Wezel et al., 2011). Bacterial diversity is central to ecosystem sustainability and soil biological function, for which the role of roots is especially important (Sanguin et al., 2006b). However, the characterization of bacterial endophyte community of agroecology systems as is the Quechua practices has been poorly analyzed.

In this work we have analyzed the bacterial endophytes that inhabit the roots of maize plants grown under a different management history in the Quechua region by pyrosequencing the V4 and V5 hypervariable regions of the 16S rRNA gene. Our hypothesis was that the soil cultivation history plays a pivotal role of structuring the endophytic bacterial communities of maize plants.

MATERIALS AND METHODS

Site Description and Root Sampling

Maize (*Zea mays* L.) plants were grown at 4 chacras located inside the same farm field (Figure S1A) near Allpas (12° 50′ 27″ S, 74° 34′ 14″ W, at 3,537 m above sea level), a village in the province of Acobamba (Huancavelica, Peru), following the traditional agricultural practices of the Quechua natives. The lateral roots (∼2 mm diameter, 2–3 cm long) of the maize plants (morphotype Qarway) were harvested 120 days after sowing. At the sampling time two of the four chacras had been cultivated with maize for at least 5 years (MC soil) and the other two were under fallow conditions before the maize sowing (F soil) for at least 5 years (Figure S1B). For each chacra, roots were sampled from plants grown at three different sites (four plants per site), pooled together, washed with sterile tap water to remove attached soil and stored at −20°C until further processing. Physicochemical analyses of the four chacras indicated they have identical soil characteristics with a sandy loam texture (64.0% sand, 30.0% silt, 6% clay), pH 5.74, 1.6% organic C and 0.11% total N (Table S1).

Surface Sterilization of Maize Roots and Isolation of Endophytes

Unfrozen roots were surface-sterilized as indicated by Liu et al. (2017). Essentially, roots were immersed in 70% ethanol for 3 min, washed with fresh sodium hypochlorite solution (2.5% available Cl^-) for 5 min, rinsed with 70% ethanol for 30 s and finally washed thoroughly with sterile distilled water. To confirm that the sterilization process was successful, small pieces of roots were cut and placed on Petri dishes containing yeast extract-mannitol (YEM) medium (Vincent, 1970). The plates were examined for bacterial growth after incubation at 30°C for 12 days. Maize roots that were not contaminated as detected by culture-dependent sterility test were used for further experiments.

Extraction of DNA From Maize Roots

DNA was extracted from 250 mg of unfrozen tissue as previously indicated (Correa-Galeote et al., 2013). Essentially, after thoroughly cutting with an sterile scalpel, samples were homogenized in 1 ml of extraction buffer containing 100 mM Tris (pH 8.0), 100 mM EDTA, 100 mM NaCl, 1% (w/v) polyvinylpyrrolidone and 2% (w/v) sodium dodecyl sulfate using a 2-ml mini-bead-beater tube containing 0.5 and 0.1 g of 106-μm- and 2-mm-diameter glass beads, respectively, for 60 s at 27 Hz. Cell debris was eliminated by centrifugation (14,000 rpm for 5 min at 4°C). Proteins were removed by treatment with 5 M sodium acetate. After treatment for 12 h with ice-cold isopropanol, nucleic acids were precipitated by centrifugation (14,000 rpm for 30 min at 4°C), washed with 70% ice-cold ethanol, recentrifruged (14,000 rpm for 15 min at 4°C) and air-dried for 30 min. Finally, DNA was purified using GeneClean columns (Qiagen). Quality and size of DNA were checked by electrophoresis on 1% agarose and quantified by spectrophotometry at 260 nm using a Nanodrop spectrophotometer (NanoDrop ND1000).

Amplification and Pyrosequencing of DNA From Maize Roots

Polymerase chain reaction (PCR) amplification of the hypervariable V4-V5 regions of the 16S rRNA gene was performed over each individual DNA extraction from roots of maize plants grown in F and MC soils using universal primers U519F and U926R (Baker et al., 2003) joined to a multiplex identifier sequence (Binladen et al., 2007; Parameswaran et al., 2007). For each sample, amplicons were generated in several replicate PCRs using mixtures (25 μl) that contained 25 pmol of each primer, 1.8 mM $MgCl_2$, 0.2 mM dNTPs, 1 × the corresponding *Taq* buffer, 1 U of *Taq* Master (5 Prime, USA) and 10 ng of the DNA template. The PCR program consisted of an initial denaturation step at 94°C for 4 min, 25 cycles of denaturation at 94°C for 15 s, primer annealing at 55°C for 45 s and extension at 72°C for 1 min, followed by a final step of heating at 72°C for 10 min. Amplicons of the same treatment were pooled together to reduce per-PCR variability and purified using the ultracentrifugal filters Ultracel-100 K membranes (Amicon) according to the manufacturer's instructions. After quantification by Nanodrop ND1000 and visualization of the DNA by agarose electrophoresis, the samples were combined in equimolar amounts and pyrosequenced in a Roche Genome Sequencer FLX system using 454 Titanium chemistry at LifeSequencing S.L. (Valencia, Spain).

Taxonomic Assignment of Sequence Reads and Diversity Indexes

Raw sequences were processed through the Ribosomal Database Project (RDP) pyrosequencing pipeline (http://pyro.cme.msu.edu) release 11 (Cole et al., 2014). Sequences were trimmed for primers, filtered and assigned to four libraries (F1G, F2G, MC1G and MC2G) according to their tags. Sequences shorter than 150 base pair, with quality scores <20 or containing any unresolved nucleotides were removed from the dataset. Chimeras were identified using the Uchime tool from FunGene database (Edgar et al., 2011) and removed from the dataset. Sequences were aligned using the SILVA-based bacterial reference alignment in the MOTHUR program (Schloss and Westcott, 2011). Aligned sequences were clustered into operational taxonomic units (OTUs) defined at 97% similarity cutoff using MOTHUR and their relative abundances calculated. The number of sequences in each OTU was employed to calculate the Good's coverage index, which is considered a relative measure of how well the sequences obtained represent the entire populations (Hughes and Bohannan, 2004). Taxonomic assignation of the sequences was performed using Geneious (Biomatters). Shannon (H′) and Simpson (S′) diversity indexes and Jaccard indexes (J_{class} and J_{abund}) were used to analyze the alpha- and beta-diversity, respectively (Chao et al., 2005).

Statistical Analyses

Relative abundances of the main genera and values of the diversity index were compared using the Student *t*-test in the XLSTAT software (Addinsoft). Multivariate techniques were used to analyze the relative abundance of endophytes using PC-ORD (McCune et al., 2002). A canonical correspondence analysis (CCA) was built to study differences in composition of dominant endophyte genera in roots from plants grown in F and MC soils.

Accession Numbers

Pyrosequencing reads are deposited in GenBank under accession numbers KT764133 to KT764925.

RESULTS

A total of 38,443 sequences were obtained from the four 16S rDNA samples sent to pyrosequencing, of which 11,278 were retained after filtering and removing chimeras. The mean number of total retained sequences per library was 2,819, ranging from 1,770 to 3,718. Average length of retained sequences was 374 ± 5 base pair (mean ± SD). Using the MOTHUR program all the sequences aligned correctly in the expected position of the 16S rDNA sequence of *Escherichia coli* and were grouped at 97% similarity in 244 distinct OTUs. A representative sequence from each OTU was sent to NCBI for identification, after which 10,485 sequences were removed as they were identified as Streptophyta-related sequences. The remaining 793 sequences grouped into 188 bacterial OTUs, of which 17 were supported by 10 or more reads and 91 corresponded to singletons (Table S2). Values of the Good's coverage index were higher than 68% for all the samples. The number of OTUs in libraries F1G and F2G were 48 and 53, respectively, significantly lower than those of 88 and 112 found in libraries MC1G and MC2G (**Table 1**). The Shannon index for OTUs in F1G and F2G showed similar values, 3.32 and 3.62, respectively, that were statistically lower than those of 4.04 and 4.02 for OTUs in MC1G and MC2G, respectively. On the other hand, values of the Simpson index for the OTUS for the four libraries varied between 0.23 and 0.32, and no significant differences were found among them (**Table 1**). The Jaccard index for J_{class} and J_{abund} also showed that the degree of similarity between libraries MC1G and MC2G was higher than that between F1G and F2G (**Table 2**). The number of shared genera between pair to pair libraries is shown in **Table 2**.

Unclassified sequences were 14 (11.48%) and 15 (13.51%) for libraries F1G and F2G, respectively, and 28 (13.73%) and 64 (17.98%) for MC1G and MC2G, respectively (**Table 3**, **Figure 1**). The remaining sequences distributed into 6 and 7 phyla for F1G and F2G, respectively, and 7 and 8 for MC1G and

TABLE 1 | Number of OTUs, values of Good's coverage index and Shannon and Simpson biodiversity index of bacterial endophytes from roots of maize plants grown in fallow (F1 and F2) and maize-cultivated (MC1 and MC2) soils.

Diversity index	F1G	F2G	MC1G	MC2G	*p*-value of *t*-test
Number of OTUs	48b	53b	88a	112a	0.06
Good's coverage	77.05	68.47	75.00	81.74	n.a.
Shannon	3.32b	3.62b	4.04a	4.02a	0.06
Simpson	0.052a	0.031a	0.023a	0.030a	0.34

Values in the same row followed by different letters are statistically different according to the Student's t-test (α ≤ 0.1). n.a., not applicable.

MC2G, respectively (**Table 3**). Phyla Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria, Acidobacteria, Chloroflexi and Cyanobacteria (in decreasing abundance) were found in each one of the four libraries, Deinococcus-Thermus and Gemmatimonadetes were detected only in libraries FG and Verrucomicrobia was found exclusively in endophytes from libraries MCG (**Figure 1**). The number of classes, orders, families and genera are also shown in **Table 3**.

The total 188 OTUs distributed into 82 different genera (Table S2), 12 of which showed a relative abundance higher than 1% and represented the 55% of the total endophytes. Altogether, these genera were (in decreasing order of abundance) *Sphingomonas*, *Burkholderia*, *Candidatus* Glomeribacter, *Dyella*, *Herbaspirillum*, *Bradyrhizobium*, *Staphylococcus*, *Methylophilus*, *Variovorax*, *Streptococcus*, *Bacillus* and *Chitinophaga* (Table S2). The number of sequences corresponding to genera *Sphingomonas, Herbaspirillum*, *Bradyrhizobium* and *Methylophilus* in libraries MCG were statistically ($\alpha \leq 0.1$) more abundant than those in the F libraries, and sequences of genera *Dyella* and *Sreptococcus* were significantly more abundant in the F libraries (**Figure 2**). Relative abundance of genera *Burkholderia, candidatus* Glomeribacter, *Staphylococcus*, *Variovorax, Bacillus* and *Chitinophaga* were similar among libraries (**Figure 2**).

A CCA sample ordination based on the relative abundance of the 12 main genera mentioned above showed that they distributed in two clearly separated groups (**Figure 3**). The two CCA axes explained 93% of the total variance and revealed that cultivation history of the soil was responsible for the grouping of the libraries along the axis 1 (canonical coefficient 1.10).

TABLE 2 | Number of shared genera between clone libraries, and Jaccard similarity index using genera presence/absence (J_{class}) and relative abundances (J_{abund}) of the bacterial endophyte communities from roots of maize plants grown in fallow (F1 and F2) and maize-cultivated (MC1 and MC2) soils.

	Number of shared genera	J_{class}	J_{abund}
F1G-F2G	14	0.30	0.62
F1G-MC1G	18	0.31	0.61
F1G-MC2G	18	0.31	0.66
F2G-MC1G	14	0.23	0.53
F2G-MC2G	19	0.32	0.66
MC1G-MC2G	32	0.70	0.77

DISCUSSION

One of the most successful soil management techniques in agricultural land is the use of fallow periods (Costa et al., 2015), this work is a first approach to understand the role of the soil cultivation history in the bacterial diversity of the endophytic bacteria of maize plants cultivated under sustainably practices. Using 454 next generation sequencing we assessed the composition and abundance of endophytic communities inside roots of amilaceous maize plants grown under fallow and maize-cultivated conditions in Andean chacras. Pyrosequencing revealed an unprecedented number of bacterial endophytes as compared with those of the genera found in previous studies based on culture-dependent and culture-independent methods (McInroy and Kloepper, 1995; Chelius and Triplett, 2001; Rai et al., 2007; Pereira et al., 2011; Ikeda et al., 2013; Sandhya et al., 2017). Altogether, a 15.26% of the total sequences found inside roots corresponded to unclassified bacteria, which indicates the presence of hitherto uncultured bacterial groups. Nevertheless, despite the resolving power of pyrosequencing to detect phylogenetic groups, genera *Pantoea*, *Klebsiella* and *Erwinia* found by other authors (Pereira et al., 2011; Montañez et al., 2012; Ikeda et al., 2013; Liu et al., 2017) after sequencing of the 16S rRNA gene of endophytes isolated from roots of different maize genotypes were not detected in our libraries. This could be due to qualitative differences in endophytic colonization (Ikeda et al., 2013).

A variety of bacteria have been reported to be endophytic, among them mostly Proteobacteria, but also Firmicutes, Actinobacteria and Bacteroidetes (reviewed in Rosenblueth and Martínez-Romero, 2006; Bulgarelli et al., 2013; Malfanova et al., 2013; Hardoim et al., 2015; Liu et al., 2017). In our study, regardless of the cultivation history of the soil, members of phylum Proteobacteria were the most abundant followed by those of Firmicutes, Bacteroidetes and Actinobacteria.

TABLE 3 | Number of taxa and distribution of sequences (%) of bacterial endophytes in roots of maize plants grown in fallow (F1 and F2) and maize-cultivated (MC1 and MC2) soils.

	Fallow soil				Maize-cultivated soil			
	F1		F2		MC1		MC2	
	Number of taxa	Number of sequences (%)	Number of taxa	Number of sequences (%)	Number of taxa	Number of sequences (%)	Number of taxa	Number of sequences (%)
Phylum	6	108 (88.52)	7	96 (86.49)	7	176 (86.27)	8	292 (82.02)
Class	8	107 (87.70)	10	92 (82.88)	8	168 (82.35)	9	285 (80.06)
Order	15	105 (86.07)	15	89 (80.18)	18	154 (75.49)	17	274 (76.97)
Family	24	99 (81.15)	21	79 (71.17)	33	151 (74.02)	33	264 (74.16)
Genus	30	99 (81.15)	30	81 (72.97)	46	147 (72.06)	48	260 (73.03)
Unclassified sequences		14 (11.48)		15 (13.51)		28 (13.73)		64 (17.98)

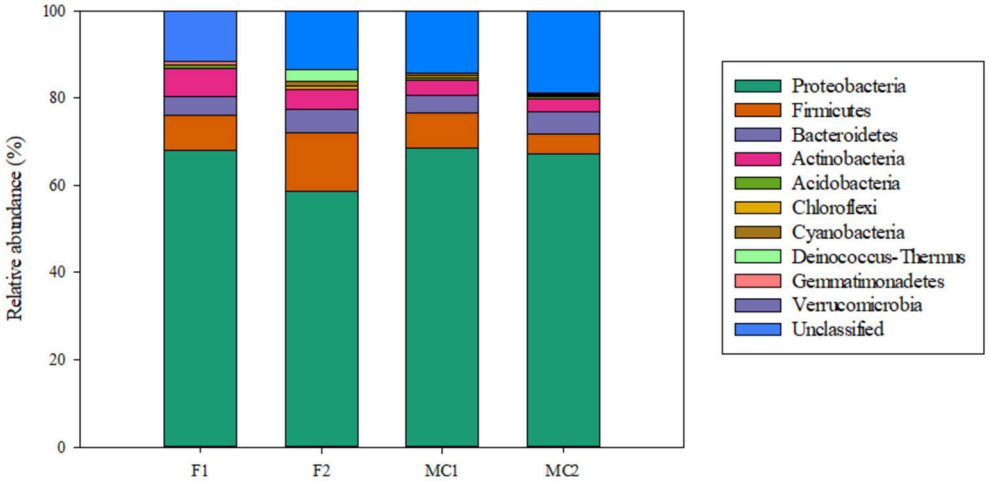


FIGURE 1 | Relative abundance of bacterial endophytes from roots of maize plants grown in fallow (F1 and F2) and maize-cultivated (MC1 and MC2) soils.

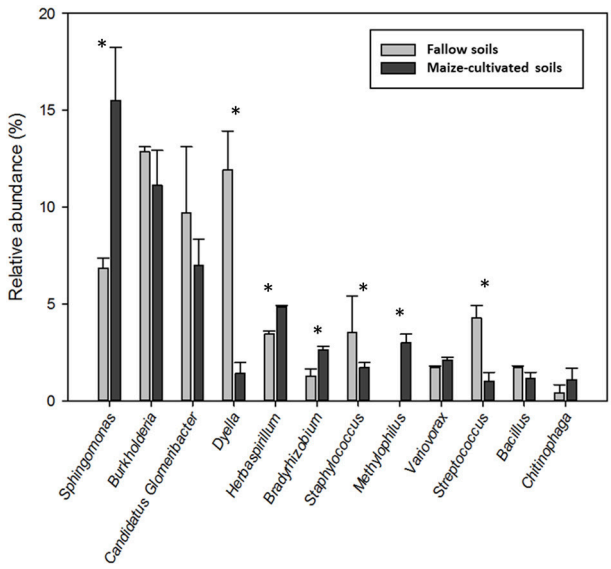


FIGURE 2 | Relative abundance of the 12 main genera found in roots of maize plants grown in fallow and maize-cultivated soils. *Indicates statistically significant differences according to the Student's *t*-test ($\alpha \leq 0.1$).

Number of OTUs and Shannon index values were statistically higher for libraries MC1G and MC2G than those for libraries FG1 and FG2. However, although the J_{class} and J_{abund} indexes were also higher for the MC communities, the bacterial endophytic communities within the plant roots from MC soils were more similar. Eleven out the 12 main genera were present in both MC and F soils and four of them (*Sphingomonas, Herbaspirillum*, *Bradyrhizobium* and *Methylophilus*) had increased relative abundance in the MC soils in comparison with that in the F soils. These results, together with those of the clone libraries diversity, indicate that plant cultivation history could have a pivotal role responsible for selection of roots endophytes from rhizospheric bacterial reservoirs. Also, these results could indicated that the maize plant growth in soils cultivated for long time requires a higher diversity of endophytes than the plants grown in a soil under a fallow time due that the natural resources of the soil are depleted after 5 years of cultivation. For example, excessive cultivation can wreck the structure of soil by reducing the capacity of holding enough moisture for growing plants (FAO, 1994) and also has been demonstrate that after 3 years of cultivation organic C, N and P declined about a 25% (Bowman et al., 1990). According to Wood et al. (2017), the fallow period is a key determinant of vegetation and soil dynamics as this period renew soil fertility, biomass and biodiversity. Therefore, after a long-time cultivation the maize plants needs a higher presence of endophytes to minimize the depletion of the soil resources.

Bacterial endophytes have been shown to modulate plant growth and development through N_2 fixation, solubilization of insoluble phosphorus, production of siderophores, production of phytohormones, lowering of ethylene concentration, production of antibiotics and antifungal metabolites and inducing systemic resistance (Somers et al., 2004; Hardoim et al., 2008, 2015; Ahemad and Kibret, 2014; Vejan et al., 2016). Some genera in this study have been shown to be diazotrophic bacteria (*Sphingomonas*, *Burkholderia*, *candidatus* Glomeribacter, *Herbaspirillum*, *Bradyrhizobium* and *Bacillus*), others solubilize inorganic phosphorus (*Sphingomonas*, *Burkholderia*, *Herbaspirillum*, *Bradyrhizobium*, *Staphylococcus*, *Methylophilus*, *Variovorax*, *Streptococcus*, *Bacillus* and *Chitinophaga*), are siderophore (*Sphingomonas*, *Burkholderia*, *Herbaspirillum*, *Bradyrhizobium*, *Staphylococcus*, *Methylophilus*, *Variovorax*, *Streptococcus*, *Bacillus* and *Chitinophaga*) or indole acetic acid (*Sphingomonas*, *Burkholderia*, *Dyella*, *Herbaspirillum*, *Bradyrhizobium*, *Staphylococcus, Methylophilus*, *Variovorax*, *Bacillus* and *Chitinophaga*) producers, have 1-aminocyclopropane-1-carboxylate (ACC) deaminase activity (*Sphingomonas*, *Burkholderia*, *Dyella*, *Herbaspirillum*, *Bradyrhizobium*, *Staphylococcus*, *Variovorax* and *Bacillus*) and

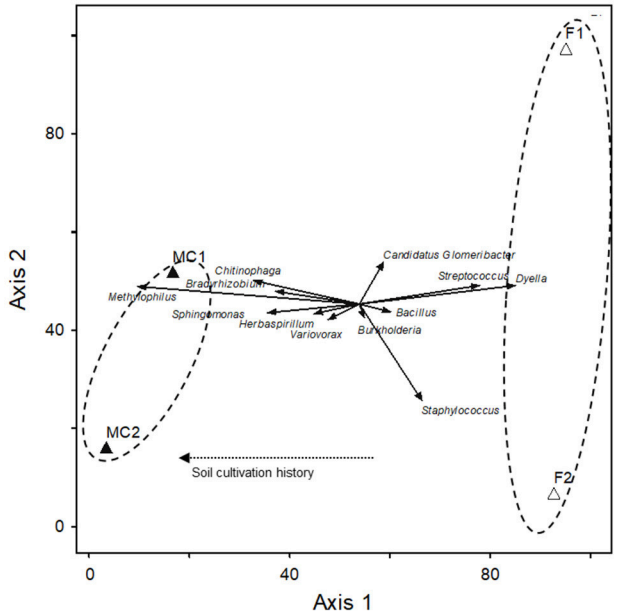


FIGURE 3 | Canonical correspondence analysis (CCA) of the 12 main genera found in roots of maize plants grown in fallow and maize-cultivated soils. Solid arrows represent vector scores for the different genera. Open and closed triangles represent the axes 1 and 2 scores for the main genera found in fallow (F1 and F2) and maize-cultivated (MC1 and MC2) soils, respectively. The dashed arrow represents the biplot vector for cultivation history of the soil.

are involved in biocontrol activity (*Sphingomonas*, *Burkholderia*, *candidatus* Glomeribacter, *Herbaspirillum*, *Methylophilus*, *Variovorax, Bacillus* and *Chitinophaga*; see Table S3).

To our knowledge, 6 of the 12 main genera in this study (*Bradyrhizobium*, *Variovorax*, *Chitinophaga*, *candidatus* Glomeribacter, *Dyella* and *Streptococcus*) have not been reported as endophytes of amilaceous maize. It should be noted, that the bacteria reported as maize endophytes for the first time could present biotechnological implications as is the case of the formulation of new microbial inoculants. The presence of *Streptococcus*, *Dyella* and *Staphylococcus* is intriguing as they are well-known human pathogens; these three genera have been detected in maize seeds, roots of blackberry, grapevine shoots, apple and orange fresh fruits (Liu et al., 2013; Phukon et al., 2013; Pinto et al., 2014; Yousaf et al., 2014; Contreras et al., 2016) and they were reported as the dominant endophytes of legumes (Boine et al., 2008; Becerra-Castro et al., 2011). Moreover, recent works suggest that pathogenic bacteria are common inhabitants of the interior of plants (Szilagyi-Zecchin et al., 2014; Blain et al., 2017; Sandhya et al., 2017).

There is to note, out of the 12 main endophytic genera here described, the genera *Candidatus* Glomeribacter, *Dyella*, *Herbaspirillum* and *Streptococcus* were not found in a previous work (Correa-Galeote et al., 2016) that describe the rizhospherics communities of these chacras and therefore the mechanisms of how these bacteria arrive to the interior of the maize roots is still unclear.

The plant host genotype (Ding et al., 2013), soil type (Rasche et al., 2006; Bulgarelli et al., 2013) and environmental soil conditions (Lundberg et al., 2012) among other factors shape bacterial community composition. Because seeds of amilaceous maize used for planting were the same and the environmental conditions, including soil type, soil psychochemical properties and irrigation, were very much alike for the four chacras used in this study, our results suggest that soil cultivation history could be a main factor controlling colonization of the internal root tissues of the plants.

Taken together our results lend support to the suggestion that cultivation history is an important driver of endophytic colonization of maize and that after a long time of cultivation of the soil the maize plants there grown need to increase the richness of the bacterial endophytes communities. Also, these results point to the importance of the fallow period of the traditional and sustainable Quechua agriculture methods in the maintenance of the soil fertility of Peruvian soils.

As a caveat, since richness of bacteria colonizing maize roots was based on pyrosequencing, it is not known whether the detection of bacteria based on DNA signature alone represent active microbes that are interacting with the host plant. Further experiments should be made in order to isolate the main endophytes described in this work and also analyze their role in the development of the maize host plant cultivated under a sustainable method as is the traditional Quechua agriculture.

AUTHOR CONTRIBUTIONS

DC-G made substantial contributions to the conception and design of the work, acquisition, analysis and interpretation of data, drafting the work and revising it critically for important intellectual content, final approval of the version to be published, agreement to be accountable for all aspects of the work. EB made substantial contributions to the conception or design of the work, interpretation of data, revising it critically for important intellectual content, final approval of the version to be published, agreement to be accountable for all aspects of the work. GA made substantial contributions to the conception or design of the work, revising it critically for important intellectual content, final approval of the version to be published, agreement to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

This study was supported by ERDF-cofinanced grant PE2012-AGR1968 from Consejería de Economía, Innovación y Ciencia (Junta de Andalucía, Spain) and the CSIC-sponsored I-COOP Agrofood project 2014CD0013.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00484/full#supplementary-material

REFERENCES

Ahemad, M., and Kibret, M. (2014). Mechanisms and applications of plant growth promoting rhizobacteria: current perspective. *JKSUES* 26, 1–20. doi: 10.1016/j.jksus.2013.05.001

Alves, G. C., Videira, S. S., Urquiaga, S., and Reis, V. M. (2015). Differential plant growth promotion and nitrogen fixation in two genotypes of maize by several *Herbaspirillum inoculants*. *Plant Soil* 387, 307–321. doi: 10.1007/s11104-014-2295-2

Baker, G. C., Smith, J. J., and Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods* 55, 541–555. doi: 10.1016/j.mimet.2003.08.009

Becerra-Castro, C., Kidd, P. S., Prieto-Fernández, Á., Weyens, N., Acea, M. J., and Vangronsveld, J. (2011). Endophytic and rhizoplane bacteria associated with *Cytisus striatus* growing on hexachlorocyclohexane-contaminated soil: isolation and characterisation. *Plant Soil* 340, 413–433. doi: 10.1007/s11104-010-0613-x

Binladen, J., Gilbert, M. T., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., et al. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2:e197. doi: 10.1371/journal.pone.0000197

Blain, N. P., Helgason, B. L., and Germida, J. J. (2017). Endophytic root bacteria associated with the natural vegetation growing at the hydrocarbon-contaminated Bitumount provincial historic site. *Can. J. Microbiol.* 63, 502–515. doi: 10.1139/cjm-2017-0039

Boine, B., Naujoks, G., and Stauber, T. (2008). Investigations on influencing plant-associated bacteria in tissue cultures of black locust (*Robinia pseudoacacia* L.). *Plant Cell Tissue Organ Cult.* 94, 219–223. doi: 10.1007/s11240-008-9395-8

Bowman, R. A., Reeder, J. D., and Lober, R. W. (1990). Changes in soil properties in a central plains rangeland soil after 3, 20, and 60 years of cultivation. *Soil Sci.* 150, 851–857. doi: 10.1097/00010694-199012000-00004

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838. doi: 10.1146/annurev-arplant-050312-120106

Celador-Lera, L., Menéndez, E., Flores-Félix, J. D., Mateos, P. F., and Rivas, R. (2016). "Analysis of the PGPB potential of bacterial endophytes associated with maize," in *Biological Nitrogen Fixation and Beneficial Plant-Microbe Interaction*, eds F. González-Andrés and E. James (Berlin: Springer), 23–35.

Chao, A., Chazdon, R. L., Colwell, R. K., and Shen, T. J. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* 8, 148–159. doi: 10.1111/j.1461-0248.2004.00707.x

Chelius, M. K., and Triplett, E. W. (2001). The diversity of Archaea and bacteria in association with the roots of *Zea mays* L. *Microb. Ecol.* 41, 252–263. doi: 10.1007/s002480000087

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

Compant, S., Van Der Heijden, M. G., and Sessitsch, A. (2010). Climate change effects on beneficial plant–microorganism interactions. *FEMS. Microbiol. Ecol.* 73, 197–214. doi: 10.1111/j.1574-6941.2010.00900.x

Contreras, M., Loeza, P. D., Villegas, J., Farias, R., and Santoyo, G. (2016). A glimpse of the endophytic bacterial diversity in roots of blackberry plants (*Rubus fruticosus*). *Genet. Mol. Res.* 15:gmr.15038542. doi: 10.4238/gmr.15038542

Correa-Galeote, D., Bedmar, E. J., Fernández-González, A. J., Fernández-López, M., and Arone, G. J. (2016). Bacterial communities in the rhizosphere of amilaceous maize (*Zea mays* L.) as assessed by pyrosequencing. *Fron. Plant. Sci.* 7:1016. doi: 10.3389/fpls.2016.01016

Correa-Galeote, D., Tortosa, G., and Bedmar, E. J. (2013). Determination of denitrification genes abundance in environmental samples. *Metagenomics* 2:235702. doi: 10.4303/mg/235702

Costa, J. L., Aparicio, V., and Cerdà, A. (2015). Soil physical quality changes under different management systems after 10 years in the Argentine humid pampa. *Solid Earth* 6, 361–371. doi: 10.5194/se-6-361-2015

Ding, T., Palmer, M. W., and Melcher, U. (2013). Community terminal restriction fragment length polymorphisms reveal insights into the diversity and dynamics of leaf endophytic bacteria. *BMC Microbiol.* 13:1. doi: 10.1186/1471-2180-13-1

Dudeja, S. S., and Giri, R. (2014). Beneficial properties, colonisation, establishment and molecular diversity of endophytic bacteria in legumes and non-legumes. *Afr. J. Microbiol. Res.* 8, 1562–1572. doi: 10.5897/AJMR2013.6541

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Eevers, N., Beckers, B., de Beeck, M. O., White, J. C., Vangronsveld, J., and Weyens, N. (2016). Comparison between cultivated and total bacterial communities associated with *Cucurbita pepo* using cultivation-dependent techniques and 454 pyrosequencing. *Syst. Appl. Microbiol.* 39, 58–66. doi: 10.1016/j.syapm.2015.11.001

FAO (1994). *Land Degradation in South Asia: Its Severity, Causes and Effects Upon the People*. Roma: World Soil Resources Reports.

Grobman, A., Bonavia, D., Dillehay, T. D., Piperno, D. R., Iriarte, J., and Holst, I. (2012). Preceramic maize from Paredones and Huaca Prieta, Peru. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1755–1759. doi: 10.1073/pnas.1120270109

Hallmann, J., Quadt-Hallmann, A., Mahaffee, W. F., and Kloepper, J. W. (1997). Bacterial endophytes in agricultural crops. *Can. J. Microbiol.* 43, 895–914. doi: 10.1139/m97-131

Hardoim, P. R., van Overbeek, L. S., Berg, G., Pirttilä, A. M., Compant, S., Campisano, A., et al. (2015). The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.* 79, 293–320. doi: 10.1128/MMBR.00050-14

Hardoim, P. R., van Overbeek, L. S., and van Elsas, J. D. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* 16, 463–471. doi: 10.1016/j.tim.2008.07.008

Herschkovitz, Y., Lerner, A., Davidov, Y., Okon, Y., and Jurkevitch, E. (2005a). *Azospirillum brasilense* does not affect population structure of specific rhizobacterial communities of inoculated maize (*Zea mays*). *Environ. Microbiol.* 7, 1847–1852. doi: 10.1111/j.1462-2920.2005.00926.x

Herschkovitz, Y., Lerner, A., Davidov, Y., Rothballer, M., Hartmann, A., Okon, Y., et al. (2005b). Inoculation with the plant-growth-promoting rhizobacterium *Azospirillum brasilense* causes little disturbance in the rhizosphere and rhizoplane of maize (*Zea mays*). *Microb. Ecol.* 50, 277–288. doi: 10.1007/s00248-004-0148-x

Hughes, J. B., and Bohannan, B. J. M. (2004). "Application of ecological diversity statistics in microbial ecology," in *Molecular Microbial Ecology Manual, Section 7*, eds G. A. Kowalchuk, F. J. de Bruijn, I. M. Head, A. D. Akkermans, and J. D. van Elsas (Berlin: Springer), 3223–3246.

Ikeda, A. C., Bassani, L. L., Adamoski, D., Stringari, D., Cordeiro, V. K., Glienke, C., et al. (2013). Morphological and genetic characterization of endophytic bacteria isolated from roots of different maize genotypes. *Microb. Ecol.* 65, 154–160. doi: 10.1007/s00248-012-0104-0

Jackson, C. R., Randolph, K. C., Osborn, S. L., and Tyler, H. L. (2013). Culture dependent and independent analysis of bacterial communities associated with commercial salad leaf vegetables. *BMC Microbiol.* 13:274. doi: 10.1186/1471-2180-13-274

Kobayashi, D. Y., and Palumbo, J. D. (2000). "Bacterial endophytes and their effects on plants and uses in agriculture," in *Microbial Endophytes*, eds C. W. Bacon and J. F. White (New York, NY: Marcel Dekker Inc.), 199–233.

Li, H. Y., Wei, D. Q., Shen, M., and Zhou, Z. P. (2012). Endophytes and their role in phytoremediation. *Fungal Divers.* 54, 11–18. doi: 10.1007/s13225-012-0165-x

Liu, Y., Wang, R., Li, Y., Cao, Y., Chen, C., Qiu, C., et al. (2017). High-throughput sequencing-based analysis of the composition and diversity of endophytic bacterial community in seeds of "Beijing" hybrid maize planted in China. *Plant Growth Regul.* 81, 317–324. doi: 10.1007/s10725-016-0208-5

Liu, Y., Zuo, S., Zou, Y., Wang, J., and Song, W. (2013). Investigation on diversity and population succession dynamics of endophytic bacteria from seeds of maize (*Zea mays* L., Nongda108) at different growth stages. *Ann. Microbiol.* 63, 71–79. doi: 10.1007/s13213-012-0446-3

López-Velasco, G., Carder, P. A., Welbaum, G. E., and Ponder, M. A. (2013). Diversity of the spinach (*Spinacia oleracea*) spermosphere and phyllosphere bacterial communities. *FEMS Microbiol. Lett.* 346, 146–154. doi: 10.1111/1574-6968.12216

Lundberg, D. S., Lebeis, S. L., Herrera-Paredes, S., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488, 86–90. doi: 10.1038/nature11237

Ma, Y., Rajkumar, M., Zhang, C., and Freitas, H. (2016). Beneficial role of bacterial endophytes in heavy metal phytoremediation. *J. Environ. Manage.* 174, 14–25. doi: 10.1016/j.jenvman.2016.02.047

Malfanova, N., Lugtenberg, B., and Berg, G. (2013). "Bacterial endophytes: who and where, and what are they doing there?," in *Molecular Microbial Ecology of the Rhizosphere Volume I*, ed F. J. de Brujin (New York, NY: Wiley-Blackwell), 393–403.

Manter, D. K., Delgado, J. A., Holm, D. G., and Stong, R. A. (2010). Pyrosequencing reveals a highly diverse and cultivar-specific bacterial endophyte community in potato roots. *Microb. Ecol.* 60, 157–166. doi: 10.1007/s00248-010-9658-x

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959

McCune, B., Grace, J. B., and Urban, D. L. (2002). *Analysis of Ecological Communities, Vol. 28*. Gleneden Beach, OR: MjM Software Design.

McInroy, J. A., and Kloepper, J. W. (1995). Survey of indigenous bacterial endophytes from cotton and sweet corn. *Plant Soil* 173, 337–342. doi: 10.1007/BF00011472

Menéndez, E., Ramirez-Bahena, M. H., Peix, A., Tejedor, C., Mulas, R., González-Andrés, F., et al. (2016). "Analysis of cultivable endophytic bacteria in roots of maize in a soil from León province in mainland Spain," in *Biological Nitrogen Fixation and Beneficial Plant-Microbe Interaction*, eds F. González-Andrés and E. James (Berlin: Springer), 45–53.

Montañez, A., Blanco, A. R., Barlocco, C., Beracochea, M., and Sicardi, M. (2012). Characterization of cultivable putative endophytic plant growth promoting bacteria associated with maize cultivars (*Zea mays* L.) and their inoculation effects *in vitro*. *Appl. Soil Ecol.* 58, 21–28. doi: 10.1016/j.apsoil.2012.02.009

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., et al. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res.* 35:e130. doi: 10.1093/nar/gkm760

Pereira, C. J., Pereira, E. J. G., Cordeiro, E. M. G., Della Lucia, T. M. C., Tótola, M. R., and Guedes, R. N. C. (2009). Organophosphate resistance in the maize weevil *Sitophilus zeamais*: magnitude and behavior. *Crop Prot.* 28, 168–173. doi: 10.1016/j.cropro.2008.10.001

Pereira, P., Ibáñez, F., Rosenblueth, M., Etcheverry, M., and Martínez-Romero, E. (2011). Analysis of the bacterial diversity associated with the roots of maize (*Zea mays* L.) through culture-dependent and culture-independent methods. *ISRN Ecol.* 2011:938546. doi: 10.5402/2011/938546

Perry, L., Sandweiss, D. H., Piperno, D. R., Rademaker, K., Malpass, M. A., Umire, A., et al. (2006). Early maize agriculture and interzonal interaction in southern Peru. *Nature* 440, 76–79. doi: 10.1038/nature04294

Petrini, O. (1991). "Fungal endophytes of tree leaves," in *Microbial Ecology of Leaves*, eds J. H. Andrews and S. S. Hirano (New York, NY: Springer-Verlag Inc.), 179–197.

Phukon, M., Sahu, P., Srinath, R., Nithya, A., and Babu, S. (2013). Unusual occurrence of *Staphylococcus warneri* as endophyte in fresh fruits along with usual *Bacillus* spp. *J. Food Saf.* 33, 102–106. doi: 10.1111/jfs.12028

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., and Gomes, A. C. (2014). Unravelling the diversity of grapevine microbiome. *PLoS ONE* 9:e85622. doi: 10.1371/journal.pone.0085622

Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J., and Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5019–5024. doi: 10.1073/pnas.0812525106

Rai, R., Dash, P. K., Prasanna, B. M., and Singh, A. (2007). Endophytic bacterial flora in the stem tissue of a tropical maize (*Zea mays* L.) genotype: isolation, identification and enumeration. *World J. Microbiol. Biotechnol.* 23, 853–858. doi: 10.1007/s11274-006-9309-z

Rasche, F., Velvis, H., Zachow, C., Berg, G., van Elsas, J. D., and Sessitsch, A. (2006). Impact of transgenic potatoes expressing anti-bacterial agents on bacterial endophytes is comparable with the effects of plant genotype, soil type and pathogen infection. *J. Appl. Ecol.* 43, 555–566. doi: 10.1111/j.1365-2664.2006.01169.x

Rijavec, T., Lapanje, A., Dermastia, M., and Rupnik, M. (2007). Isolation of bacterial endophytes from germinated maize kernels. *Can. J. Microbiol.* 53, 802–808. doi: 10.1139/W07-048

Romero, F. M., Marina, M., and Pieckenstain, F. L. (2014). The communities of tomato (*Solanum lycopersicum* L.) leaf endophytic bacteria, analyzed by 16S-ribosomal RNA gene pyrosequencing. *FEMS Microbiol. Lett.* 351, 187–194. doi: 10.1111/1574-6968.12377

Rosenblueth, M., and Martínez-Romero, E. (2006). Bacterial endophytes and their interactions with hosts. *Mol. Plant Microbe Interact.* 19, 827–837. doi: 10.1094/MPMI-19-0827

Rothberg, J. M., and Leamon, J. H. (2008). The development and impact of 454 sequencing. *Nat. Biotechnol.* 26, 1117–1124. doi: 10.1038/nbt1485

Ryan, R. P., Germaine, K., Franks, A., Ryan, D. J., and Dowling, D. N. (2008). Bacterial endophytes: recent developments and applications. *FEMS Microbiol. Lett.* 278, 1–9. doi: 10.1111/j.1574-6968.2007.00918.x

Sandhya, V., Shrivastava, M., Ali, S. Z., and Prasad, V. S. S. K. (2017). Endophytes from maize with plant growth promotion and biocontrol activity under drought stress. *Russ. Agric. Sci.* 43, 22–34. doi: 10.3103/S1068367417010165

Sanguin, H., Herrera, A., Oger-Desfeux, C., Dechesne, A., Simonet, P., Navarro, E., et al. (2006a). Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria. *Environ. Microbiol.* 8, 289–307. doi: 10.1111/j.1462-2920.2005.00895.x

Sanguin, H., Remenant, B., Dechesne, A., Thioulouse, J., Vogel, T. M., Nesme, X., et al. (2006b). Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. *Appl. Environ. Microbiol.* 72, 4302–4312. doi: 10.1128/AEM.02686-05

Schloss, P. D., and Westcott, S. L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl. Environ. Microbiol.* 77, 3219–3226. doi: 10.1128/AEM.02810-10

Schmalenberger, A., and Tebbe, C. C. (2003). Bacterial diversity in maize rhizospheres: conclusions on the use of genetic profiles based on PCR-amplified partial small subunit rRNA genes in ecological studies. *Mol. Ecol.* 12, 251–262. doi: 10.1046/j.1365-294X.2003.01716.x

Schulz, B. (2006). "Mutualistic interactions with fungal root endophytes," in *Microbial Root Endophytes*, eds B. Schulz, C. Boyle, and T. N. Sieber (Berlin: Springer), 261–279.

Schulz, B., and Boyle, C. (2006). "What are endophytes?," in *Microbial Root Endophytes*, eds B. Schulz, C. Boyle, and T. N. Sieber (Berlin: Springer), 1–13.

Sharma, I. P., Chandra, S., Kumar, N., and Chandra, D. (2017). "PGPR: heart of soil and their role in soil fertility," in *Agriculturally Important Microbes for Sustainable Agriculture Volume I*, eds V. S. Meena, P. K. Mishra, J. K. Bisht, and A. Pattanayak (Singapore: Springer), 51–67. doi: 10.1007/978-981-10-5589-8_3

Somers, E., Vanderleyden, J., and Srinivasan, M. (2004). Rhizosphere bacterial signaling: a love parade beneath our feet. *Crit. Rev. Microbiol.* 30, 205–240. doi: 10.1080/10408410490468786

Szilagyi-Zecchin, V. J., Ikeda, A. C., Hungria, M., Adamoski, D., Kava-Cordeiro, V., Glienke, C., et al. (2014). Identification and characterization of endophytic bacteria from corn (*Zea mays* L.) roots with biotechnological potential in agriculture. *AMB Exp.* 4:26. doi: 10.1186/s13568-014-0026-y

Van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., Gonzalez, J. D. J. S., et al. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1088–1092. doi: 10.1073/pnas.1013011108

Vejan, P., Abdullah, R., Khadiran, T., Ismail, S., and Nasrulhaq Boyce, A. (2016). Role of plant growth promoting rhizobacteria in agricultural sustainability-a review. *Molecules* 21:573. doi: 10.3390/molecules21050573

Vincent, J. M. (1970). *A Manual for the Practical Study of the Root-Nodule Bacteria*. Hoboken, NJ: Wiley-Blackwell.

Weyens, N., van der Lelie, D., Taghavi, S., Newman, L., and Vangronsveld, J. (2009). Exploiting plant–microbe partnerships to improve biomass production and remediation. *Trends Biotechnol.* 27, 591–598. doi: 10.1016/j.tibtech.2009.07.006

Wezel, A., Bellon, S., Doré, T., Francis, C., Vallod, D., and David, C. (2011). "Agroecology as a science, a movement and a practice," in *Sustainable Agriculture, Vol. 2*, eds E. Lichtfouse, M. Hamelin, M. Navarrete, and P. Debaeke (Amsterdam: Springer), 27–43.

Wood, S. L., Rhemtulla, J. M., and Coomes, O. T. (2017). Cropping history trumps fallow duration in long-term soil and vegetation dynamics of shifting cultivation systems. *Ecol. Appl.* 27, 519–531. doi: 10.1002/eap.1462

Yousaf, S., Bulgari, D., Bergna, A., Pancher, M., Quaglino, F., Casati, P., et al. (2014). Pyrosequencing detects human and animal pathogenic taxa in the grapevine endosphere. *Front. Microbiol.* 5:327. doi: 10.3389/fmicb.2014.00327

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Correa-Galeote, Bedmar and Arone. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 July 2017
doi: 10.3389/fmicb.2017.01266

Long-Term Rock Phosphate Fertilization Impacts the Microbial Communities of Maize Rhizosphere

*Ubiana C. Silva[1], Julliane D. Medeiros[2], Laura R. Leite[2], Daniel K. Morais[2,3], Sara Cuadros-Orellana[2,4], Christiane A. Oliveira[5], Ubiraci G. de Paula Lana[5], Eliane A. Gomes[5] and Vera L. Dos Santos[1]**

[1] Microbiology Department, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] Biosystems Informatics and Genomics Group, René Rachou Research Center, Fiocruz, Belo Horizonte, Brazil, [3] Microbiology Institute, Czech Academy of Sciences – CAS, Prague, Czechia, [4] Centro de Biotecnología de los Recursos Naturales, Facultad de Ciencias Agrarias y Forestales, Universidad Católica del Maule, Talca, Chile, [5] Embrapa Maize and Sorghum, Sete Lagoas, Brazil

Phosphate fertilization is a common practice in agriculture worldwide, and several commercial products are widely used. Triple superphosphate (TSP) is an excellent soluble phosphorus (P) source. However, its high cost of production makes the long-term use of crude rock phosphate (RP) a more attractive alternative in developing countries, albeit its influence on plant-associated microbiota remains unclear. Here, we compared long-term effects of TSP and RP fertilization on the structure of maize rhizosphere microbial community using next generation sequencing. Proteobacteria were dominant in all conditions, whereas Oxalobacteraceae (mainly *Massilia* and *Herbaspirillum*) was enriched in the RP-amended soil. *Klebsiella* was the second most abundant taxon in the RP-treated soil. *Burkholderia* sp. and *Bacillus* sp. were enriched in the RP-amended soil when compared to the TSP-treated soil. Regarding fungi, Glomeromycota showed highest abundance in RP-amended soils, and the main genera were *Scutellospora* and *Racocetra*. These taxa are already described as important for P solubilization/acquisition in RP-fertilized soil. Maize grown on TSP and RP-treated soil presented similar productivity, and a positive correlation was detected for P content and the microbial community of the soils. The results suggest changes of the microbial community composition associated to the type of phosphate fertilization. Whilst it is not possible to establish causality relations, our data highlights a few candidate taxa that could be involved in RP solubilization and plant growth promotion. Moreover, this can represent a shorter path for further studies aiming the isolation and validation of the taxa described here concerning P release on the soil plant system and their use as bioinoculants.

Keywords: microbial community, maize rhizosphere, rock phosphate

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council (CONICET), Argentina

Reviewed by:
Aymé Spor, INRA, UMR1347 Agroécologie, France
David Dowling, Institute of Technology Carlow, Ireland

****Correspondence:***
Vera L. Dos Santos
verabio@gmail.com

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *27 April 2017*
Accepted: *23 June 2017*
Published: *11 July 2017*

Citation:
Silva UC, Medeiros JD, Leite LR, Morais DK, Cuadros-Orellana S, Oliveira CA, de Paula Lana UG, Gomes EA and Dos Santos VL (2017) Long-Term Rock Phosphate Fertilization Impacts the Microbial Communities of Maize Rhizosphere. Front. Microbiol. 8:1266. doi: 10.3389/fmicb.2017.01266

INTRODUCTION

Maize (*Zea mays* L.) is one of the main cereals produced in the world, with approximately 900 million tons produced annually (United State Department of Agricultural, 2017). Phosphorus (P) is an essential nutrient for this crop, especially in the flowering and grain filling stages (Vasconcellos et al., 2000) and to ensure this production, millions of tons of P fertilizer are added to soils each year

(Koppelaar and Weikard, 2013). However, this overuse of P fertilizers raises the cost of production and can exerts negative impacts on aquatic environment by the eutrophication of surface water (Dodds et al., 2009). In addition, phosphate reserves are a finite resource that are predicted to be depleted in the next few centuries (Cordell and White, 2011). This can become a limiting factor in global food production. Research efforts have been directed to the use of rock phosphate (RP) as P fertilizers as they have lower costs, are agronomically more useful and environmentally more feasible than soluble P. Although RP has a lower reactivity than commercial fertilizers for direct application on the soil, P availability from these rocks can be increased over years of cultivation, for instance, due to the action of the soil microbiota (Coutinho et al., 1991). Actually, P-solubilizer microorganisms have been described at this niche (Whitelaw, 1999; Linu et al., 2009; Silva et al., 2014; Matthews and Adzahar, 2016). Organic acid production has been pointed as the main factor involved in RP microbial solubilization (Rashid et al., 2004; Mendes et al., 2014). Nevertheless, other mechanisms, such as the release of H^+ ions during NH_4^+ assimilation or metabolic reactions that trigger proton excretion such as cellular respiration, can also contribute to P solubilization (Illmer and Schinner, 1995). Furthermore, microorganisms at the rhizosphere can increase P supply by hydrolysis of organic P through the action of phosphatases, especially phytases (Richardson and Simpson, 2011). Important factors for increasing plant P acquisition includes roots architecture alteration in many plant families like root cluster formation, proliferation and increase of lateral length, a higher number of root hairs, and a reduced length and thickness of the primary root (Hammond et al., 2009; Hunter et al., 2014). All these processes have a strong influence on the microbial communities that colonize the rhizosphere of these plants, and studies suggest that plants can attract P solubilizing microorganisms according to their root structure and the compounds released from exudates (Bolan et al., 1997; Grayston et al., 1998; Broeckling et al., 2008; Guo and Wang, 2009; Hartmann et al., 2009; Chaparro et al., 2012; Hunter et al., 2014).

The impact of RP on rhizosphere microbial diversity is still poorly understood. Recently, culture-independent methods, such as "meta-omics" based on next generation sequencing, have helped to shed some light on the diversity associated to plant–microorganism–soil interactions (Aira et al., 2010; Chhabra et al., 2013; Peiffer et al., 2013; Li et al., 2014). These techniques provide access to a collection of microbial groups related to certain environmental conditions and contribute to the description of non-cultivated taxa by the classic methods of microbiology. Based on the exposed, we used a culture-independent approach to analyze the effect of 3 years of maize fertilization with RP on the microbial community diversity of the rhizosphere. We hypothesized that long-term RP fertilization will drive the selection of efficient P solubilizing or uptake microorganisms to the rhizosphere of maize comparing to conventional use of soluble fertilizers. The results may contribute to understanding of the microbial ecosystem services to the plants. In the future, the use of microbial inoculants able to improve P release/acquisition by plants along with RP fertilization can become a sustainable agriculture practice that will reduce of the environmental impacts besides of the costs production.

MATERIALS AND METHODS

Experimental Design

The field site is located at Embrapa Maize and Sorghum in Sete Lagoas, Minas Gerais, Brazil (19°28′S 44°15′W). The experiment was conducted in an agricultural Oxisol of the Brazilian Savanna biome (Cerrado), classified as a low P soil (Supplementary Table S1). This area was cultivated for 3 years with the 30F35YH hybrid maize (Pioneer, Brazil) on a system of two annual cultivations, one in summer and the other in winter. Each cultivation was conducted using the same P fertilization conditions, namely, soil fertilized with RP (rock phosphate of Araxa), soil with triple superphosphate (TSP, a commercial P fertilizer), and soil without added P. The experimental design was composed of these three types of P fertilization in three replicates. P_2O_5 concentration on the control soil was 4 kg/ha and was adjusted to 100 kg/ha on the RP or TSP fertilized soil. Each experimental plot consisted of six rows that were 5 m long and spaced at 70 cm. The useful experimental area consisted of 14 m^2 of the four central rows of 5 m, eliminating the two border rows.

Sample Collection for Metataxonomic and Growth Parameters Analysis

Rhizospheric soil was collected at 60 days after sowing (flowering time) from the plants in the summer period of the 3rd year of cultivation. Each triplicate consisted of five randomly chosen plants. Initially, the roots of the plants collected in the field were washed with water under pressure to remove excess soil. Then, approximately 5 g (fresh weight) of the fine roots with adhering rhizosphere soil from each replicate were transferred to 50 mL Falcon tubes with 30 mL of phosphate buffer (per liter: 6.33 g of $NaH_2PO_4.H_2O$ and 16.5 g of $Na_2HPO_4.7H_2O$). Samples were shaken for 30 s and the roots were transferred to new Falcon tubes containing 30 mL of phosphate buffer. The soil obtained from this procedure corresponded to the first soil fraction. Roots were homogenized in the phosphate buffer again and sonicated at low frequency (50–60 Hz) for 5 min, followed by five cycles of sonication for 30 s and rest for 30 s. Roots were removed and the second fraction of the obtained soil was mixed with the first soil fraction and centrifuged. The pellet obtained after this procedure, considered rhizosphere soil, was frozen in liquid nitrogen and stored at −80°C until DNA extraction.

Subsamples of shoots and grains corresponding to a stand of 45 plants at the end of the experiment (approximately 90 days of culture) were collected for measuring parameters related to plant growth, such as nutrients content, plant biomass and grain productivity. Also, the rhizosphere soil of five plants in three replicates for each treatment was collected at flowering time for determination of P content.

PCR Amplification and Sequencing

Metagenomic DNA was extracted from rhizosphere soil samples using a Max Power Soil DNA Kit (MO BIO Laboratories, Inc., Carlsbad, United States). DNA was quality-checked using agarose gel electrophoresis and quantified by absorbance at 260 nm on a spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States). The V3–V4 region of the 16S rRNA gene was amplified from DNA using the primer pair 341F and 806R (Klindworth et al., 2013) and the ITS region was amplified using the primer pair ITS3_KYO1F and ITS4_KYO1R (Toju et al., 2012) (Supplementary Table S2). The primers were modified to contain an overhang sequence complementary to the Nextera Index (Illumina, San Diego, CA, United States). A PCR reaction was performed in 25 μL of final volume, which contained, briefly, 12.5 μL of PCR buffer (2× KAPA HiFi HotStart ReadyMix), 5 μM of primer and 30 ng of DNA. PCR amplification was performed using 95°C for 3 min of initial denaturation, followed by 30 cycles at 95°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 30 s and a final extension at 72°C for 5 min. The amplicons were purified with AMPure XT beads (Beckman Coulter Genomics, Danvers, MA, United States). Then, amplicons were ligated to an index sequence during a second PCR, containing 12.5 μL of PCR buffer (2× KAPA HiFi HotStart ReadyMix), 3 μL of each Nextera XT index, 2.5 μL of the PCR purified product and 7 μL of ultrapure water (25 μL final volume). Amplification was performed at 95°C for 3 min, eight cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 30 s and a final extension at 72°C for 5 min. Amplicons were purified using AMPure XP beads and the size of the libraries was checked using Bioanalyzer DNA 1000 Assay (Agilent, Santa Clara, CA, United States) and quantified using the KK4824 Kapa kit (Biosciences, Woburn, MA, United States). The paired-end (2 × 300 bp) libraries were sequenced using a MiSeq Reagent V3 Kit.

DNA Sequence Analysis

The DNA sequences were analyzed using a modified version of the pipeline recommended by the Brazilian Microbiome Project (http://www.brmicrobiome.org/; Pylro et al., 2014). Briefly, quality control of the sequence data was performed with Trimmomatic v0.32 (Bolger et al., 2014), with an average Phred score of 15 into a sliding window composed of 4 bp. Then, the sequences were truncated to a size of 400 bp for bacteria and 300 bp for fungi, and unique sequences (singletons) were removed. Chimeric sequences were filtered using USEARCH (Edgar, 2010). Taxonomy was assigned using Greengenes_13_08 for the bacteria database and UNITE_2016_08_04 for the fungi database at 97% similarity and was performed using the QIIME package (Caporaso et al., 2010). To improve the taxonomic assignment of the Oxalobacteraceae and Gigasporaceae families, their sequences were also aligned using the SINA Alignment Service from Silva (Pruesse et al., 2012) and NCBI BLASTn (Altschul et al., 1990) from GenBank, respectively.

The sequence data have been submitted to the GenBank databases under accession number PRJNA379083 for bacterial community and PRJNA379918 for fungal community.

AMF Community Analysis by T-RFLP

For arbuscular mycorrhizal fungi (AMF) community analysis by terminal restriction fragment length polymorphism (T-RFLP), the 28S rRNA gene was amplified using the general fungal LR1 and FLR2 primers (Van Tuinen et al., 1998; Trouvelot et al., 1999), followed by a nested PCR reaction employing the AMF-specific primers FLR3 (5′ labeled with 6-FAM) and FLR4 (5′ labeled with NED) (Gollotte et al., 2004). The PCR contained 1× reaction buffer, 0.2 μM of each primer, 2.5 mM of $MgCl_2$, 0.125 mM of dNTPs, 2.5 U of Taq DNA polymerase (Invitrogen, Carlsbad, CA, United States) and 50 ng of DNA in a final volume of 50 μL. A nested PCR reaction was performed with 2.5 μL of the products from the first PCR reaction in the same conditions reported above. Thermal cycling for all reactions included an initial denaturing step of 95°C for 5 min, 35 cycles consisting of 1 min at 95°C, 1 min at 58°C and 1 min at 72°C, followed by a final extension step of 72°C for 10 min. The amplified fragments were digested using the restriction enzyme *Taq*I (New England Biolabs, Beverly, MA, United States) and were incubated for 6 h at 65°C. To evaluate the generated fragments, 2 μL of the digestion was mixed to 9.8 μL of deionized formamide (Applied Biosystems, Foster

TABLE 1 | Read numbers obtained from sequencing, OTU number, richness estimators, and diversity indices of the P treatments for bacteria and fungi community, adjusted for 37,000 and 100,000 reads, respectively.

				Richness estimator		Diversity index	
Group	**Treatment[1]**	**N° reads[2]**	**N° OTU[3]**	**Chao1[4]**	**ACE[5]**	**Shannon[6,7]**	**Simpson[8]**
Bacteria	Control	84, 224 ± 23, 034	1258 ± 140 a	1329 ± 145 a	1315 ± 140 a	6.4 ± 0.4 b	0.92 ± 0.03 b
	RP	77, 972 ± 19, 846	1441 ± 128 a	1513 ± 116 a	1501 ± 116 a	7.6 ± 0.35 a	0.98 ± 0.01 a
	TSP	60, 211 ± 12, 328	1463 ± 88 a	1536 ± 79 a	1519 ± 82 a	8.3 ± 0.03 a	0.99 ± 0.01 a
Fungi	Control	257, 511 ± 97, 148	628 ± 44 a	649 ± 40 a	652 ± 42 a	5.4 ± 0.4 a	0.92 ± 0.03 a
	RP	269, 627 ± 56, 806	724 ± 84 a	759 ± 80 a	756 ± 83 a	6.1 ± 0.1 a	0.96 ± 0.01 a
	TSP	249, 197 ± 53, 537	658 ± 55 a	696 ± 54 a	696 ± 53 a	5.2 ± 0.8 a	0.86 ± 0.01 a

[1]*Treatments: control, without added P; RP, rock phosphate added; TSP, triple superphosphate added.* [2]*Mean and standard error of the read numbers for triplicates in each condition.* [3]*Analysis of variance adjusted to the negative binomial distribution followed of the contrast test at 5% for bacteria and fungi community.* [4,5,6,8]*Analysis of variance followed by a Scott Knott test at 5% for bacteria and fungi.* [7]*Analysis of variance using distribution of Poisson followed of the contrast test at 5% for fungi, according to the Shannon index. Means with the same letter is equal according to Scott Knott test or Contrast test at 5%.*

City, CA, United States) and 0.2 μL standard ROX 500 (Applied Biosystems). Analysis was performed on the 3500 XL Genetic Analyzer (Applied Biosystems) using the GeneMapper 5.0 software. T-RF profiles were selected by the software if their minimum peak height was above the noise observed, usually above 50 relative fluorescence units. Only peaks between 50 and 500 bp were considered to avoid peaks caused by primer-dimers and to obtain fragments within the linear range of the internal size standard. T-RF length profiles for AMF were loaded into the online T-RFLP processing software T-REX (Culman et al., 2009) for noise filtering and peak alignment.

Statistical Analysis

Nutrient content, plant biomass and grain productivity of the plants were measured and submitted to variance analysis. Means of different response factors were compared by a Scott Knott test ($p < 0.05$).

T-RFLP data were analyzed using non-metric multidimensional scaling (NMDS) with a Jaccard similarity matrix. Permutational multivariate analysis of variance of the distance matrix was then performed by an Adonis test using the Vegan package in the R program (R Development Core Team, 2008).

Taxonomic classification and alpha and beta diversity analyses were performed using the core diversity pipeline of the QIIME package. Rarefaction curves were constructed based on the observed operational taxonomic units (OTUs) number per reads for bacteria and fungi, and these curves were evaluated by good coverage analysis. To assess the diversity of the samples, we used the coverage estimator Ace, the richness estimator Chao1 and the Shannon and Simpson diversity indices.

Analyses of variance were performed to indices, followed by a Scott Knott test at 5% probability for Ace, Chao1, and Simpson (for bacteria and fungi) and for Shannon (for bacteria). Analysis of variance of the Shannon index for fungi was performed using a Poisson distribution, followed by a contrast test at 5% probability. Additionally, for comparing the OTU number, we used the negative binomial distribution, followed by a contrast test at 5% probability. Variance analysis, followed by a Scott Knott test at 5%, was also used to determine differences at the levels of phyla and families (only for groups with at least 3% relative abundance). Moreover, significance analysis of the relative abundance of the bacterial and fungal taxa was held at a family level (values greater than 0.1%) using the edgeR package. All of the above analysis were performed using R program.

Principal coordinates analysis (PCoA) was performed using the weighted and unweighted UniFrac distance metrics for bacteria and the Bray–Curtis metric for fungi to show the influence of P fertilization on beta diversity. Similarity analysis (ANOSIM) of the clustering was used to evaluate the contribution of the P treatments on microbial communities between samples. In addition, we conducted an NMDS analysis using the Rho metric for bacteria and fungi to assess the sources of variation in the bacteria and fungi community matrices (constructed using OTU number, taxa richness, and family abundance) due to plant biomass, grain yield and P accumulated in the soil using the Past program.

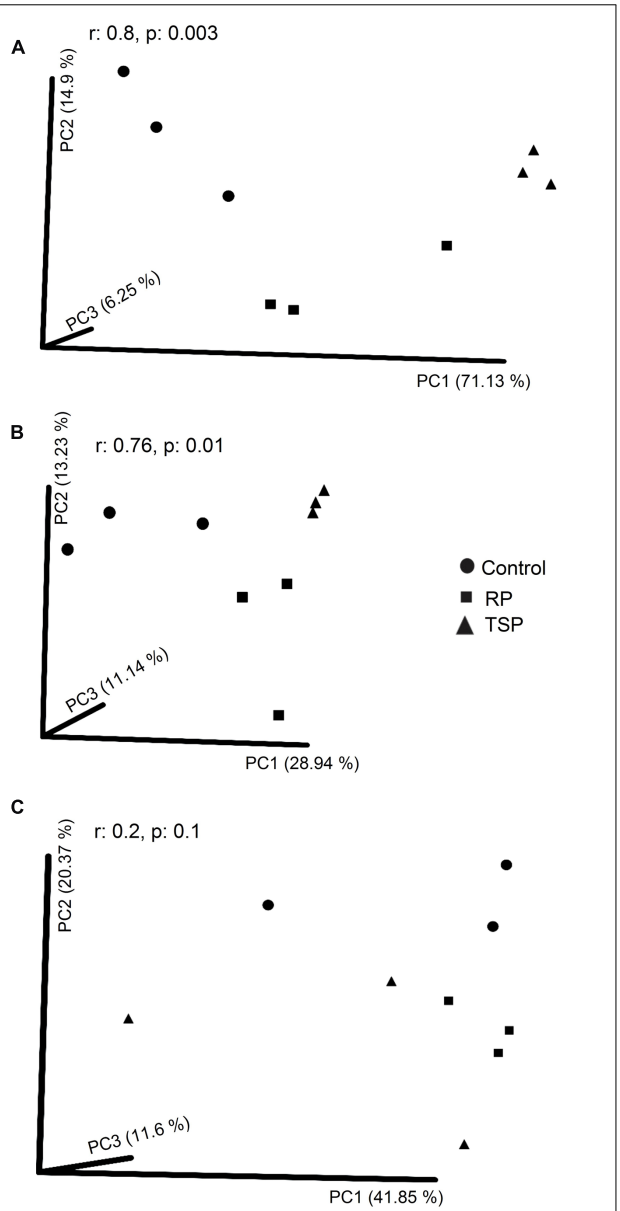


FIGURE 1 | Principal coordinate analysis (PCoA) of the bacterial and fungal communities of the maize rhizosphere grown without added P (control), with rock phosphate (RP), and with triple superphosphate (TSP). Analysis of similarity (ANOSIM) was made between among the P treatments. **(A)** Clustering similarity of bacteria using weighted UniFrac and **(B)** using unweighted UniFrac. **(C)** Clustering similarity of the fungal community using the Bray–Curtis metric.

RESULTS

P Fertilization Affects the Structure and Composition of the Maize Rhizosphere Microbial Community

Following assembly and quality filtering, a total of 667,222 high-quality reads of the bacterial 16S rRNA gene were obtained

from nine samples of rhizospheric soil with an average of 74,136 reads/sample. For fungal amplicon libraries, 2,329,006 high-quality reads were also obtained from the nine samples of rhizospheric soil (an average of 258,778 reads per sample). OTUs having at least 97% sequence similarity with reads deposited in the used databases (Greengenes and Unite) was represented in rarefaction curves within each treatment (**Supplementary Figures S1A,C**). Rarefaction curves tend to an asymptote reaching 0.99 of Good Coverage values (**Supplementary Figures S1B,D**), indicating that the depth of the sequencing was sufficient to accurately characterize both the bacterial and fungal communities in the samples. Bacterial richness (Chao1 and Ace) was greater than fungal richness in all samples (**Table 1**), but there were no significant differences between the treatments ($p = 0.10$; $\alpha = 0.05$). According to Shannon's and Simpson's indices, the bacterial diversity was higher in the treatments RP and TSP added ($p = 0.02$; $\alpha = 0.05$) when compared to the control (**Table 1**), which showed lower equitability. For fungi, the distribution of OTUs among species (equitability) was favored to a higher extent by the RP treatment (Simpson's index = 0.96), followed by the control (0.92) and the TSP treatment (0.86) (**Table 1**).

We used the distance matrix of weighted UniFrac (**Figure 1A**) and found a clustering of the bacterial communities into three groups, depending on the P treatments. However, the transformations based on the unweighted UniFrac distance showed a reduction in variance related to the P treatments (**Figure 1B**). For fungi, there was no separation of the samples at different P fertilization treatments using Bray–Curtis in the PCoA (**Figure 1C**). However, T-RFLP allowed grouping the samples ($p = 0.003$) depending on the given P treatments for the AMF (**Supplementary Figure S2**).

In general, differences in the abundance of bacterial community taxa were observed between P treatments (**Figure 2**). Regarding RP fertilized soil, the most abundant phylum was Proteobacteria (52%), followed by Planctomycetes (11%), Actinobacteria (7.8%) and Chloroflexi (5.8%) (**Figure 2A**). In the TSP, Proteobacteria (41%) was also the most abundant, followed by Planctomycetes (13.6%), Actinobacteria (8.6%), Chloroflexi (8.5%) and Acidobacteria (6.4%). On the control, the phylum Proteobacteria corresponded to 64.7% of the reads, followed by Chloroflexi with 6.2%, Planctomycetes with 5.6% and Actinobacteria with 5.4%. The major alteration corresponded to the decrease on the abundance of Proteobacteria with the use of TSP and RP in relation to control. For the other phyla, with exception for Firmicutes and Chloroflexi, it was observed an inverse effect: the relative abundance increased with the use of these P sources. **Figure 2B** summarizes the alterations observed in the bacterial community at family level. The dominant families were Enterobacteriaceae and Oxalobacteraceae. In the RP fertilized soil, there was a predominance of Oxalobacteraceae compared to the control and TSP treatments. Enterobacteriaceae was stimulated on the control and RP treatments (**Figure 2B**). The *Massilia* and *Herbaspirillum* genera were the most abundant of Oxalobacteraceae in the RP soil (**Figure 2C**). In addition, *Klebsiella* genus was the most abundant OTU found in the bacterial community, indicating its contribution to the differentiation of samples between P sources. The main alterations between the bacterial and fungal communities on the maize rhizosphere cultivated with TSP and RP were further explored by the Volcano plot (**Figure 3A**) that shown a general perspective of the OTUs' fold changes between treatments according to the significance analysis of the relative abundance of the taxa using edgeR analysis. It was also observed predominance of Oxalobacteraceae in the RP fertilized soil compared to the TSP added soil, following by Burkholderiaceae and Bacillaceae (**Figure 3B**), comprised by *Burkholderia* and *Bacillus* genera (**Figure 2C**).

For the fungi, variation of the taxa abundance was also observed as a function of the P source. Ascomycota dominated in all experimental setup systems (**Figure 2D**). The second most abundant phylum was Glomeromycota, followed by Basidiomycota, Zygomycota, and Chytridiomycota. In the RP fertilized soil, the Glomeromycota OTUs were enriched in comparison to other treatments (**Figure 2D**), and Gigasporaceae was the dominant family in this niche (**Figure 2E**). *Scutellospora* and *Racocetra* were the most representative genera of this family (**Figure 2F**). Other families were abundant in the RP treatment in relation to TSP, according to edgeR analysis showed in the Volcano Plot (**Figures 3C,D**), such as Mortierellaceae and Acaulosporaceae, which showed *Mortierella* (**Figure 2F**) and *Acaulospora* (not showed data) as representative genera, respectively. Moreover, Saccharomycetales predominated in the control soil.

Soil P Content and Productivity of the Maize are Related to the Microbial Community

We used NMDS analysis to assess whether any parameters such as plant biomass, maize productivity, and P content on the soil (Supplementary Table S3) could be associated with the taxonomic variation of the microbial communities amidst P treatments (**Figure 4**). The formation of two groups was observed, one composed of the samples of the RP- and TSP-fertilized soils and the control soil, suggesting that the bacterial and fungal communities profile was altered in the control soil. For RP and TSP treatments, we observed positive correlation of the samples with grain yield, maize biomass and P content in the soil. Additionally, the values of all these parameters for the control decreased in relation to the RP- or TSP-fertilized soil (Supplementary Table S3).

DISCUSSION

We used a metataxonomic approach to characterize the structure of the microbial community at maize rhizosphere under different P sources. The P fertilization type modified the taxa profile of the microbial community of maize rhizosphere (**Figure 1**). The abundance of OTUs from bacteria were more affected by P sources than richness, as noted in the separation of the bacterial community by weighted UniFrac, which considers phylogenetic affiliations and OTUs abundance (**Figure 1A**). These results are different from the unweighted UniFrac

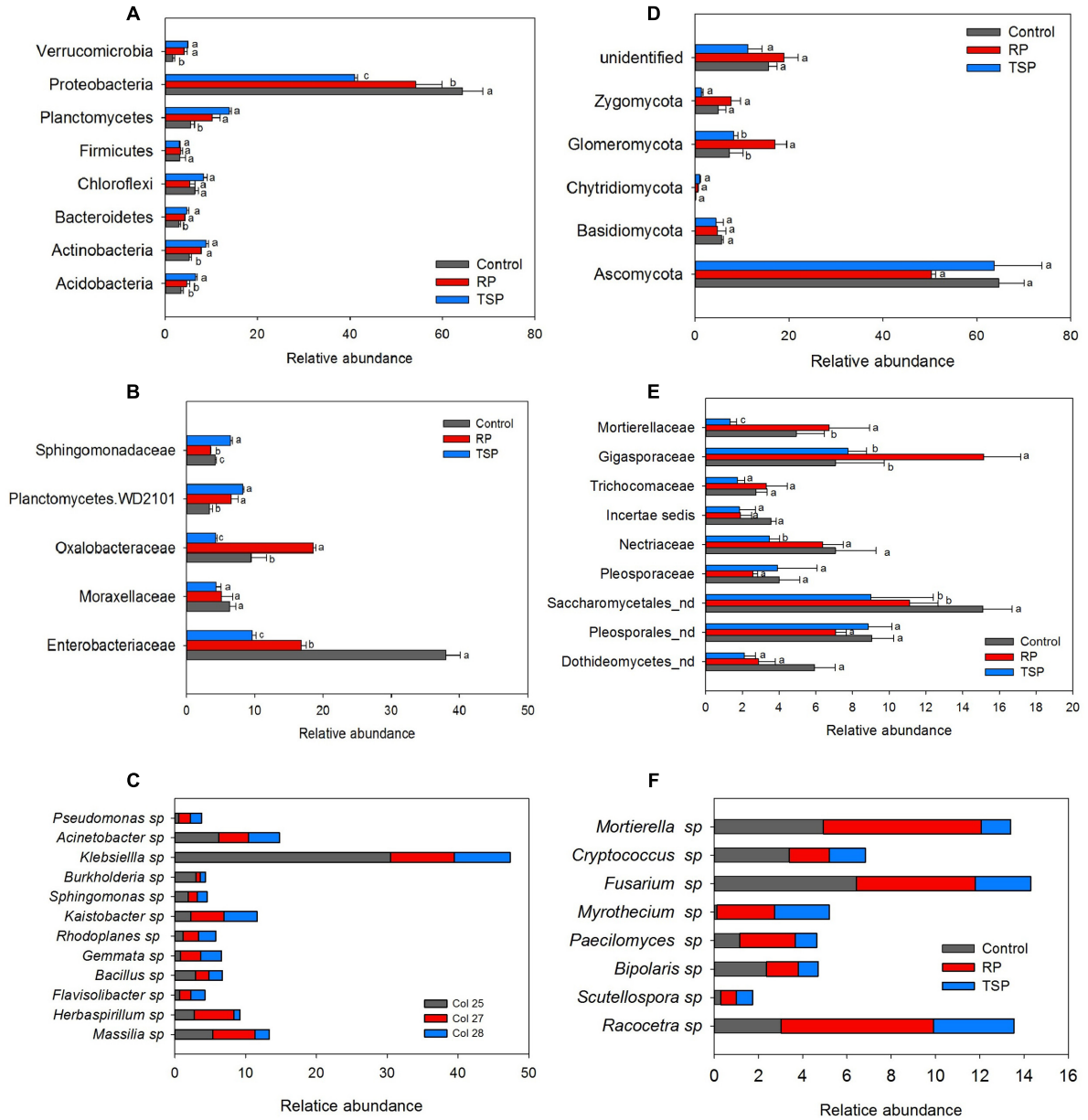


FIGURE 2 | Relative abundance of **(A)** bacterial phyla, **(B)** families, and **(C)** genera. **(D)** Fungal phyla, **(E)** families, and **(F)** genera. Taxa occurring in at least one condition with an abundance greater than 3% were considered, and variance analysis was performed between P treatments, followed by Scott Knott test at 5%. For each taxon, means with the same letter are not significantly different. Control, treatment without added P; RP, treatment with rock phosphate; TSP, treatment with triple superphosphate; nd, identification was not possible at the family level.

(**Figure 1B**), which is sensitive only to the presence or absence of taxa, but not to OTUs abundance. Additionally, the Shannon and Simpson indices, which consider OTUs abundance, showed significantly higher values for P added samples (RP and TSP) in relation to the control for bacteria (**Table 1**). These results are different from previous studies that described a increase of the bacterial diversity in the P unfertilized soil (da Silva and Nahas, 2002; Toljander et al., 2008). It is possible that the effects of P sources on bacterial diversity are variable and are likely to be site-dependent. Moreover, such effects can be somehow associated to the soil and rhizosphere microbiota capacity of recovery from environmental perturbations, mainly after long periods of exposition to such condition. Factors like soil chemistry, plant genotype, management techniques, and plant growth stage also has modified the bacterial community associated with the maize (Castellanos et al., 2009; Cavaglieri et al., 2009; Aira et al., 2010; Li et al., 2014).

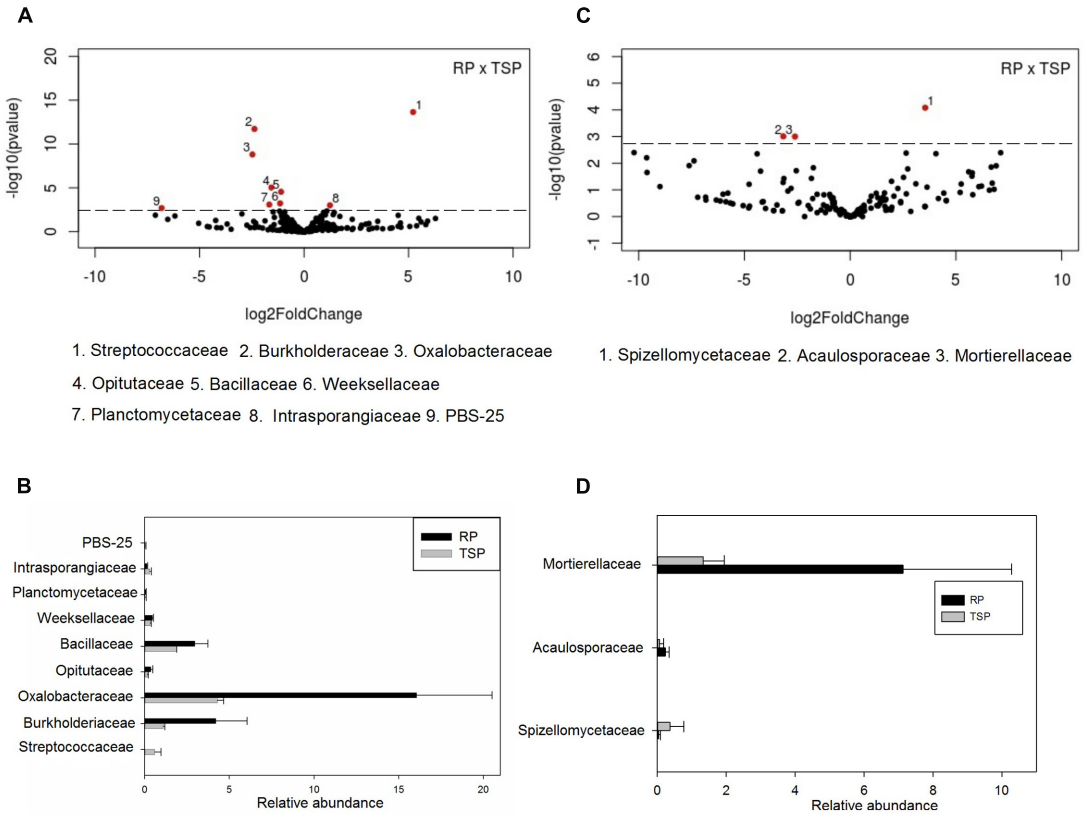


FIGURE 3 | Significance analysis of the relative abundance of the taxa until family level (threshold of 0.1%) for bacterial **(A)** and fungal **(C)** communities using the edgeR analysis. In the volcano plot (**A** and **C**), the taxa that differ between RP and TSP treatments are showed above dashed line of the *y*-axis, corresponding to 0.05 of *p*-value and *x*-axis represents the fold changes in relative abundance (log2) of the taxa between treatments. Relative abundance only of the significant taxa in the edgeR analysis (volcano plot) for bacteria community **(B)** and **(D)** fungi community of the RP compared to TSP.

We also evaluated which taxa were enriched or decreased in the RP treatment in relation to the TSP treatment and the control. Proteobacteria was the dominant phylum in all P treatments (**Figure 2A**), and was found to be predominant in the microbial community of the maize rhizosphere (Chauhan et al., 2011; Peiffer et al., 2013; Li et al., 2014; Yang et al., 2017). Different taxa related to this phyla responded differently to P treatments. While Enterobacteriacea taxa (gammaproteobacteriacee) decreased with TSP and RP, Oxalobacteraceae and Burkholderiaceae (betaproteobacteria) increased with RP addition (**Figures 2B**, **3B**). Bacillaceae (Firmicutes) showed significantly higher abundance in RP when compared to TSP-fertilized soil (**Figure 3B**). The taxa stimulated in response to cultivation using RP share similarity with those found in the *Phaseolus vulgaris* rhizosphere such as Oxalobacteraceae, Enterobacteriaceae, besides Actinobacteria, Comamonadaceae, Bradyrhizobacteriaceae, and Pseudomonodaceae (Trabelsi et al., 2017).

Klebsiella sp., the most abundant genera of Enterobacteriaceae (**Figure 2C**), has been frequently found in association with the maize crop (Brusetti et al., 2005; Roesch et al., 2007; Arruda et al., 2013), and it comprises bacterial species that are able to improve P release (Rajput et al., 2013; Walpola et al., 2014). Our group isolated *Klebsiella* from the rhizosphere and from the endophytic microbiota of maize cultivated in low P soil (Vieira et al., unpublished data). *Massilia* and *Herbaspirillum* are the most abundant genera of the Oxalobacteraceae family, and have already been described as P solubilizing bacteria (Estrada et al., 2013; Wang et al., 2016). *Burkholderia* sp. (Burkholderiaceae) and *Bacillus* sp. (Bacillaceae) are also reported to be effective RP solubilizers, both *in vitro* assays (Gomes et al., 2014; Ghosh et al., 2016) or inoculated in plants (Baig et al., 2014; Stephen et al., 2015; Wahid et al., 2016). Furthermore, *Burkholderia* and *Herbaspirillum* are both diazotrophs.

Regarding fungal community, although clustering related to P source were not significant on PCoA (**Figure 1C**), the AMF community was divided according to P treatments using T-RFLP analysis (**Supplementary Figure S2**). Additionally, the Glomeromycota phylum showed the most evident change between treatments, being the most abundant phylum on RP added soil (**Figure 2D**). Among the AMF families, Gigasporaceae was significantly more abundant on the RP added soil (**Figure 2E**). *Scutellospora* and *Racocetra* genera were predominant in this family (**Figure 2F**). Our research group has also found Gigasporaceae in the rhizosphere of maize cultivated in low P level soil, especially *Racocetra* sp. (unpublished data).

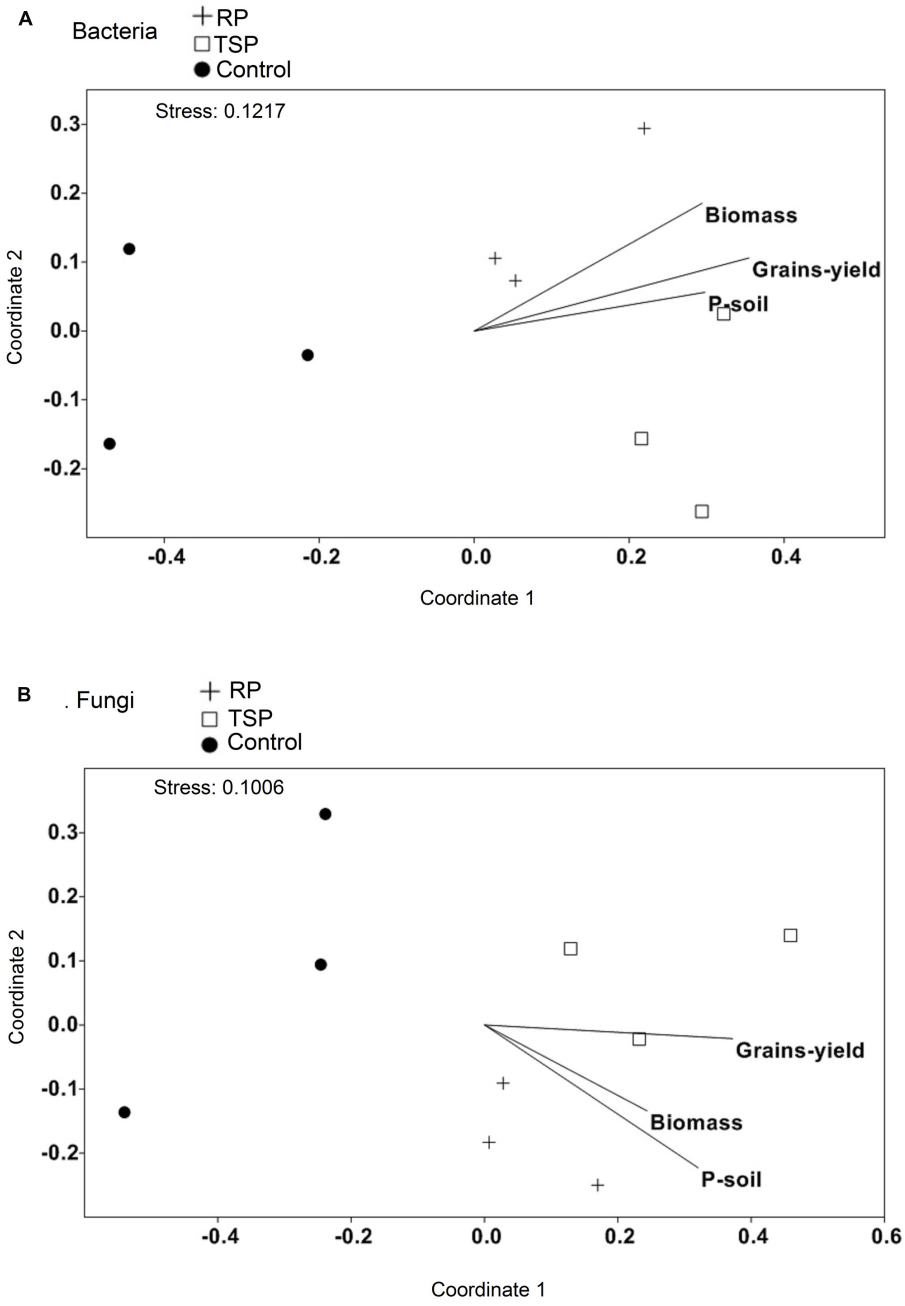


FIGURE 4 | Non-metric multidimensional scaling (NMDS) of the composition of the **(A)** bacterial community and **(B)** the fungal community using a Rho metric. The samples were plotted with fitted environmental variables (biomass of the plant, grain yield, and P in the soil at the end of the experiment). Arrows indicate the direction of the correlation with the microbial community. For community composition, an OTU richness and phyla abundance with more than 1% of relative abundance were used. Treatments: soil without added P (control), soil treated with rock phosphate (RP), and soil with added triple superphosphate (TSP).

Acaulosporaceae family also showed a significant increase in their abundance in the RP added soil when compared to soil with TSP ($p < 0.05$) (**Figures 3C,D**). P absorption by maize can be related to greater AMF presence on RP soil, since plant–AMF symbiosis supports P supply to plants on low P availability soils (Nouri et al., 2014). Thus, this indicates that these fungi may have contributed to maize productivity in RP treatment (Supplementary Table S3). *Mortierella* sp. also showed major abundance on RP-fertilized soil (**Figure 2F**), and was already reported as RP solubilizing fungi (Osorio and Habte, 2013; Vega et al., 2015). Moreover, synergistic effects of the inoculation of *Mortierella* sp. with AMF for plant P uptake and growth have been described (Zhang et al., 2011; Osorio and Habte, 2013).

Some studies have reported bacteria as symbionts of AMF hyphae. These were defined as "mycorrhiza helper bacteria," which stimulate mycelial growth and/or contribute to inhibition of competitors and antagonists (Frey-Klett et al., 2007; Offre et al., 2008; Scheublin et al., 2010). Varied strains found in this work show symbiose with AMF. *Massilia* sp., for example, was reported by Cruz et al. (2008) to be associated to spores of AMF. *Burkholderia* and *Bacillus* genera were also reported as AMF symbionts from hyphosphera (Levy et al., 2003; Frey-Klett et al., 2007), and some *Burkholderia* strains isolated from AMF hyphae showed RP solubilization capacity (Taktek et al., 2015, 2017; Trabelsi et al., 2017). Thus, this result suggests that the high abundance of *Massilia* (Oxalobacteraceae), *Burkholderia* (Burkholderiaceae) and *Bacillus* (Bacillaceae) in the RP-fertilized soil could be associated to the greater abundance of AMF in this treatment when compared to the others (**Figures 2C**, **3B**). Moreover, some bacterial genera enriched in the RP treatment have been described as auxin producers including *Herbaspirillum* spp. (Brusamarello-Santos et al., 2012), *Burkholderia* sp. and *Bacillus* sp. (Galdiano et al., 2011). This phytohormone is known to stimulate root growth that can also help P uptake by the plant once it increases its absorption surface area.

A lower AMF abundance in the TSP supplemented soil was expected since there is a significant decrease in the percentage of AMF colonization in the soil with high P level available (Smith et al., 1992). However, there was also a lower AMF abundance in the soil with no P addition, possibly due to the occurrence of P competition between the AMF hyphae and roots in the rhizosphere (Smith et al., 2011). Some studies have demonstrated that plants cultivated in very low P doses have decreased AMF colonization (Amijee et al., 1989; Koide and Li, 1990).

Our results are consistent with the long-term selection of P solubilizing microbial community. These microorganisms could be contributing to an increase on the P content in the RP-fertilized soil at the end of the 3 years of cultivation and to the facilitated P absorption by plants. This is evidenced by the maize plants growth profile cultivated in soil added RP, a low solubility phosphorus source, that was comparable to that of the plants cultivated in the TSP-added soil (**Figure 4** and Supplementary Table S3). However, we do not neglect the plant's ability in acquiring nutrients by itself and the effects of physical–chemical processes over the 3-year cultivation, which may also contribute for soil P dynamics and consequently help its release in RP-fertilized soil (Goedert and Lobato, 1980; Coutinho et al., 1991). While it is not possible to infer cause-effect relationship in microbiota-phosphate acquisition, our study points to a few candidate groups possibly involved RP solubilization, which could be further evaluated as inoculants in greenhouse and field studies.

AUTHOR CONTRIBUTIONS

US, SC-O, CO, EG, and VD designed the experiment. US, JM, and UL obtained and processed the data. US, JM, LL, and DM analyzed the data. US and VD wrote the paper with contribution of all co-authors.

FUNDING

The authors are grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), to Pró-Reitoria de Pesquisa da UFMG, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant no 477349/2013-7 and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) grant no. Apq-01819-13 for financial support.

ACKNOWLEDGMENTS

The authors are grateful to the CPqRR platform by support in bioinformatics data processing (http://www.cpqrr.fiocruz.br/pg/plataformas/), Brazilian Microbiome Project for suggesting data analysis pipelines, to Anna Christina de Matos Salim and Flávio Marcos Gomes Araújo for their assistance in library construction and sequencing, and to Marcus V. Dias-Souza, for his suggestions on the manuscript. The Brazilian group belongs to the MCTI/CNPq/CAPES/FAPS (INCT-MPCPAgro).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01266/full#supplementary-material

FIGURE S1 | Rarefaction curves of **(A)** bacteria and **(C)** fungi using QIIME software were calculated according observed OTU number in the treatment means. Control, without added P; RP, rock phosphate addition; TSP, addition of triple superphosphate. Furthermore, 0.99 of Good Coverage values from 10,000 reads for **(B)** bacteria and 20,000 for **(D)** fungi was observed in all treatments.

FIGURE S2 | Community of mycorrhizal fungi detected in the soil fertilized with RP, TSP, and without the addition of P (control). Non-metric multidimensional scaling (NMDS) was based on the matrix of Jaccard and it was performed variance analysis of the distance matrix using the multivariate permutation test (Adonis).

REFERENCES

Aira, M., Gomez-Brandon, M., Lazcano, C., Baath, E., and Dominguez, J. (2010). Plant genotype strongly modifies the structure and growth of maize rhizosphere microbial communities. *Soil Biol. Biochem.* 42, 2276–2281. doi: 10.1016/j.soilbio.2010.08.029

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Amijee, F., Tinker, P. B., and Stribley, D. P. (1989). The development of endomycorrhizal systems VII. A detailed study of the effects of soil phosphorus on colonization. *New Phytol.* 111, 435–446. doi: 10.1111/j.1469-8137.1989.tb00706.x

Arruda, L., Beneduzi, A., Martins, A., Lisboa, B., Lopes, C., Bertolo, F., et al. (2013). Screening of rhizobacteria isolated from maize (*Zea mays* L) in Rio Grande do Sul State (south Brazil) and analysis of their potential to improve plant growth. *Appl. Soil Ecol.* 63, 15–22. doi: 10.1016/j.apsoil.2012.09.001

Baig, K. S., Arshad, M., Khalid, A., Hussain, S., Abbas, M. N., and Imran, M. (2014). Improving growth and yield of maize through bioinoculants carrying auxin production and phosphate solubilizing activity. *Soil Environ.* 33, 159–168.

Bolan, N. S., Elliott, J., Gregg, P. E. H., and Weil, S. (1997). Enhanced dissolution of phosphate rocks in the rhizosphere. *Biol. Fertil. Soils* 24, 169–174. doi: 10.1007/s003740050226

Bolger, Anthony, M., Marc Lohse, and Bjoern Usadel. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Broeckling, C. D., Broz, A. K., Bergelson, J., Manter, D. K., and Vivanco, J. M. (2008). Root exudates regulate soil fungal community composition and diversity. *Appl. Environ. Microbiol.* 74, 738–744. doi: 10.1128/AEM.02188-07

Brusamarello-Santos, L. C. C., Pacheco, F., Aljanabi, S. M. M., Monteiro, R. A., Cruz, L. M., Baura, V. A., et al. (2012). Differential gene expression of rice roots inoculated with the diazotroph *Herbaspirillum seropedicae*. *Plant Soil* 356, 113–125. doi: 10.1007/s11104-011-1044-z

Brusetti, L., Francia, P., Bertolini, C., Pagliuca, A., Borin, S., Sorlini, C., et al. (2005). Bacterial communities associated with the rhizosphere of transgenic Bt 176 maize (*Zea mays*) and its non transgenic counterpart. *Plant Soil* 266, 11–21. doi: 10.1007/s11104-005-5399-x

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., and Costello, E. K. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Castellanos, T., Dohrmann, A. B., Imfeld, G., Baumgarte, S., and Tebbe, C. C. (2009). Search of environmental descriptors to explain the variability of the bacterial diversity from maize rhizospheres across a regional scale. *Eur. J. Soil Biol.* 45, 383–393. doi: 10.1016/j.ejsobi.2009.07.006

Cavaglieri, L., Orlando, J., and Etcheverry, M. (2009). Rhizosphere microbial community structure at different maize plant growth stages and root locations. *Microbiol. Res.* 164, 391–399. doi: 10.1016/j.micres.2007.03.006

Chaparro, J., Sheflin, A., Manter, D., and Vivanco, J. (2012). Manipulating the soil microbiome to increase soil health and plant fertility. *Biol. Fertil. Soils* 48, 489–499. doi: 10.1007/s00374-012-0691-4

Chauhan, P. S., Chaudhry, V., Mishra, S., and Nautiyal, C. S. (2011). Uncultured bacterial diversity in tropical maize (*Zea mays* L.) rhizosphere. *J. Basic Microbiol.* 51, 15–32. doi: 10.1002/jobm.201000171

Chhabra, S., Brazil, D., Morrissey, J., Burke, J. I., O'Gara, F., and N Dowling, D. (2013). Characterization of mineral phosphate solubilization traits from a barley rhizosphere soil functional metagenome. *Microbiologyopen* 2, 717–724. doi: 10.1002/mbo3.110

Cordell, D., and White, S. (2011). Peak phosphorus: clarifying the key issues of a vigorous debate about long-term phosphorus security. *Sustainability* 3, 2027–2049. doi: 10.3390/su3102027

Coutinho, E. L., Natale, W., Nova, A. S. V., and Sitta, D. S. (1991). Eficiência agronômica de fertilizantes fosfatados para a cultura da soja. *Pesqui. Agropecu. Bras.* 26, 1393–1399.

Cruz, A. F., Horii, S., Ochiai, S., Yasuda, A., and Ishii, T. (2008). Isolation and analysis of bacteria associated with spores of *Gigaspora margarita*. *J. Appl. Microbiol.* 104, 1711–1717. doi: 10.1111/j.1365-2672.2007.03695.x

Culman, S. W., Bukowski, R., Gauch, H. G., Cadillo-Quiroz, H., and Buckley, D. H. (2009). T-REX: software for the processing and analysis of T-RFLP data. *BMC Bioinformatics* 10:171. doi: 10.1186/1471-2105-10-171

da Silva, P., and Nahas, E. (2002). Bacterial diversity in soil in response to different plans, phosphate fertilizers and liming. *Braz. J. Microbiol.* 33, 304–310. doi: 10.1590/S1517-83822002000400005

Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., et al. (2009). Eutrophication of U.S. freshwaters: analysis of potential economic damages. *Environ. Sci. Technol.* 43, 12–19. doi: 10.1021/es801217q

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Estrada, G. A., Baldani, V. L. D., de Oliveira, D. M., Urquiaga, S., and Baldani, J. I. (2013). Selection of phosphate-solubilizing diazotrophic *Herbaspirillum* and *Burkholderia* strains and their effect on rice crop yield and nutrient uptake. *Plant Soil* 369, 115–129. doi: 10.1007/s11104-012-1550-7

Frey-Klett, P., Garbaye, J. A., and Tarkka, M. (2007). The mycorrhiza helper bacteria revisited. *New Phytol.* 176, 22–36. doi: 10.1111/j.1469-8137.2007.02191.x

Galdiano, R. F. Jr., Pedrinho, E. A. N., Castellane, T. C. L., and Lemos, E. G. D. M. (2011). Auxin-producing bacteria isolated from the roots of *Cattleya walkeriana*, an endangered Brazilian orchid, and their role in acclimatization. *Rev. Bras. Ciênc. Solo* 35, 729–737.

Ghosh, R., Barman, S., Mukherjee, R., and Mandal, N. C. (2016). Role of phosphate solubilizing *Burkholderia* spp. for successful colonization and growth promotion of *Lycopodium cernuum* L. (Lycopodiaceae) in lateritic belt of Birbhum district of West Bengal. India. *Microbiol. Res.* 183, 80–91. doi: 10.1016/j.micres.2015.11.011

Goedert, W. J., and Lobato, E. (1980). Eficiência agronômica de fosfatos em solo de cerrado. *Pesqui. Agropecu. Bras.* 15, 311–318.

Gollotte, A., van Tuinen, D., and Atkinson, D. (2004). Diversity of arbuscular mycorrhizal fungi colonising roots of the grass species *Agrostis capillaris* and *Lolium perenne* in a field experiment. *Mycorrhiza* 14, 111–117. doi: 10.1007/s00572-003-0244-7

Gomes, E. A., Silva, C. U., Marriel, I. E., Oliveira, C. A., and Lana, U. P. (2014). Rock phosphate solubilizing microorganisms isolated from maize rhizosphere soil. *Rev. Bras. Milho Sorgo* 13, 69–81. doi: 10.18512/1980-6477/rbms.v13n1p69-81

Grayston, S. J., Wang, S., Campbell, C. D., and Edwards, A. C. (1998). Selective influence of plant species on microbial diversity in the rhizosphere. *Soil Biol. Biochem.* 30, 369–378. doi: 10.1016/S0038-0717(97)00124-7

Guo, H. C., and Wang, G. H. (2009). Phosphorus status and microbial community of paddy soil with the growth of annual Ryegrass (*Lolium multiflorum* Lam.) under different phosphorus fertilizer treatments. *J. Zhejiang Univ. Sci. B* 10, 761–768. doi: 10.1631/jzus.B0920101

Hammond, J. P., Broadley, M. R., White, P. J., King, G. J., Bowen, H. C., Hayden, R., et al. (2009). Shoot yield drives phosphorus use efficiency in *Brassica oleracea* and correlates with root architecture traits. *J. Exp. Bot.* 60, 1953–1968. doi: 10.1093/jxb/erp083

Hartmann, A., Schmid, M., Van Tuinen, D., and Berg, G. (2009). Plant-driven selection of microbes. *Plant Soil* 321, 235–257. doi: 10.1007/s11104-008-9814-y

Hunter, P. J., Teakle, G., and Bending, G. D. (2014). Root traits and microbial community interactions in relation to phosphorus availability and acquisition, with particular reference to *Brassica*. *Front. Plant Sci.* 5:27. doi: 10.3389/fpls.2014.00027

Illmer, P., and Schinner, F. (1995). Solubilization of inorganic calcium phosphates-Solubilization mechanisms. *Soil Biol. Biochem.* 27, 257–263. doi: 10.1016/0038-0717(94)00190-C

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, e1. doi: 10.1093/nar/gks808

Koide, R. T., and Li, M. (1990). On host regulation of the vesicular-arbuscular mycorrhizal symbiosis. *New Phytol.* 114, 59–65. doi: 10.1111/j.1469-8137.1990.tb00373.x

Koppelaar, R. H. E. M., and Weikard, H. P. (2013). Assessing phosphate rock depletion and phosphorus recycling options. *Glob. Environ. Change* 23, 1454–1466. doi: 10.1016/j.gloenvcha.2013.09.002

Levy, A., Chang, B. J., Abbott, L. K., Kuo, J., Harnett, G., and Inglis, T. J. (2003). Invasion of spores of the arbuscular mycorrhizal fungus *Gigaspora decipiens* by *Burkholderia* spp. *Appl. Environ. Microbiol.* 69, 6250–6256. doi: 10.1128/AEM.69.10.6250-6256.2003

Li, X. Z., Rui, J. P., Mao, Y. J., Yannarell, A., and Mackie, R. (2014). Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. *Soil Biol. Biochem.* 68, 392–401. doi: 10.1016/j.soilbio.2013.10.017

Linu, M. S., Stephen, J., and Jisha, M. S. (2009). Phosphate solubilizing *Gluconacetobacter* sp., *Burkholderia* sp. and their potential interaction with cowpea (*Vigna unguiculata* (L.) Walp.). *Int. J. Agric. Res.* 4, 79–87. doi: 10.3923/ijar.2009.79.87

Matthews, S., and Adzahar, M. S. (2016). Application of phosphate solubilising microorganisms to increase the solubilisation of rock phosphates in soil. *J. Trop. Agric. Food Sci.* 44, 9–18.

Mendes, G. O., de Freitas, A. L. M., Pereira, O. L., da Silva, I. R., Vassilev, N. B., and Costa, M. D. (2014). Mechanisms of phosphate solubilization by fungal isolates when exposed to different P sources. *Ann. Microbiol.* 64, 239–249. doi: 10.1007/s13213-013-0656-3

Nouri, E., Breuillin-Sessoms, F., Feller, U., and Reinhardt, D. (2014). Phosphorus and nitrogen regulate arbuscular mycorrhizal symbiosis in *Petunia hybrida*. *PLoS ONE* 9:e90841. doi: 10.1371/journal.pone.0090841

Offre, P., Pivato, B., Mazurier, S., Siblot, S., Berta, G., Lemanceau, P., et al. (2008). Microdiversity of *Burkholderiales* associated with mycorrhizal and nonmycorrhizal roots of *Medicago truncatula*. *FEMS Microbiol. Ecol.* 65, 180–192. doi: 10.1111/j.1574-6941.2008.00504.x

Osorio, N. W., and Habte, M. (2013). Synergistic effect of a phosphate-solubilizing fungus and an arbuscular mycorrhizal fungus on leucaena seedlings in an Oxisol fertilized with rock phosphate. *Botany* 91, 274–281. doi: 10.1139/cjb-2012-0226

Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., et al. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6548–6553. doi: 10.1073/pnas.1302837110

Pruesse, E., Peplies, J., and Glockner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. doi: 10.1093/bioinformatics/bts252

Pylro, V. S., Roesch, L. F. W., Ortega, J. M., do Amaral, A. M., Tótola, M. R., Hirsch, P. R., et al. (2014). Brazilian microbiome project: revealing the unexplored microbial diversity—challenges and prospects. *Microb. Ecol.* 67, 237–241. doi: 10.1007/s00248-013-0302-4

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org

Rajput, M. S., Naresh Kumar, G., and Rajkumar, S. (2013). Repression of oxalic acid-mediated mineral phosphate solubilization in rhizospheric isolates of *Klebsiella pneumoniae* by succinate. *Arch. Microbiol.* 195, 81–88. doi: 10.1007/s00203-012-0850-x

Rashid, M., Khalil, S., Ayub, N., Alam, S., and Latif, F. (2004). Organic acids production and phosphate solubilization by phosphate solubilizing microorganisms (PSM) under in vitro conditions. *Pak. J. Biol. Sci.* 7, 187–196. doi: 10.3923/pjbs.2004.187.196

Richardson, A. E., and Simpson, R. J. (2011). Soil microorganisms mediating phosphorus availability update on microbial phosphorus. *Plant Physiol.* 156, 989–996. doi: 10.1104/pp.111.175448

Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., et al. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1, 283–290. doi: 10.1038/ismej.2007.53

Scheublin, T. R., Sanders, I. R., Keel, C., and Van der Meer, J. R. (2010). Characterisation of microbial communities colonising the hyphal surfaces of arbuscular mycorrhizal fungi. *ISME J.* 4, 752–763. doi: 10.1038/ismej.2010.5

Silva, U. C., Mendes, G. O., Silva, N. M. R., Duarte, J. L., Silva, I. R., Tótola, M. R., et al. (2014). Fluoride-tolerant mutants of *Aspergillus niger* show enhanced phosphate solubilization capacity. *PLoS ONE* 9:e110246. doi: 10.1371/journal.pone.0110246

Smith, S. E., Jakobsen, I., Gronlund, M., and Smith, F. A. (2011). Roles of arbuscular mycorrhizas in plant phosphorus nutrition: interactions between pathways of phosphorus uptake in arbuscular mycorrhizal roots have important implications for understanding and manipulating plant phosphorus acquisition. *Plant Physiol.* 156, 1050–1057. doi: 10.1104/pp.111.174581

Smith, S. E., Robson, A. D., and Abbott, L. K. (1992). The involvement of mycorrhizas in assessment of genetically dependent efficiency of nutrient uptake and use. *Plant Soil* 146, 169–179. doi: 10.1007/BF00012010

Stephen, J., Shabanamol, S., Rishad, K. S., and Jisha, M. S. (2015). Growth enhancement of rice (*Oryza sativa*) by phosphate solubilizing *Gluconacetobacter* sp. (MTCC 8368) and *Burkholderia* sp. (MTCC 8369) under greenhouse conditions. *Biotech* 5, 831–837. doi: 10.1007/s13205-015-0286-5

Taktek, S., St-Arnaud, M., Piché, Y., Fortin, J. A., and Antoun, H. (2017). Igneous phosphate rock solubilization by biofilm-forming mycorrhizobacteria and hyphobacteria associated with *Rhizoglomus irregulare* DAOM 197198. *Mycorrhiza* 27, 13–22. doi: 10.1007/s00572-016-0726-z

Taktek, S., Trépanier, M., Servin, P. M., St-Arnaud, M., Piché, Y., Fortin, J. A., et al. (2015). Trapping of phosphate solubilizing bacteria on hyphae of the arbuscular mycorrhizal fungus *Rhizophagus irregularis* DAOM 197198. *Soil Biol. Biochem.* 90, 1–9. doi: 10.1016/j.soilbio.2015.07.016

Toju, H., Tanabe, A. S., Yamamoto, S., and Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* 7:e40863. doi: 10.1371/journal.pone.0040863

Toljander, J. F., Santos-González, J. C., Tehler, A., and Finlay, R. D. (2008). Community analysis of arbuscular mycorrhizal fungi and bacteria in the maize mycorrhizosphere in a long-term fertilization trial. *FEMS Microbiol. Ecol.* 65, 323–338. doi: 10.1111/j.1574-6941.2008.00512.x

Trabelsi, D., Cherni, A., Zineb, A. B., Dhane, S. F., and Mhamdi, R. (2017). Fertilization of *Phaseolus vulgaris* with the Tunisian rock phosphate affects richness and structure of rhizosphere bacterial communities. *Appl. Soil Ecol.* 114, 1–8. doi: 10.1016/j.apsoil.2016.11.014

Trouvelot, S., van Tuinen, D., Hijri, M., and Gianinazzi-Pearson, V. (1999). Visualization of ribosomal DNA loci in spore interphasic nuclei of glomalean fungi by fluorescence in situ hybridization. *Mycorrhiza* 8, 203–206. doi: 10.1007/s005720050235

United State Department of Agricultural (2017). Available at: https://www.worldcornproduction.com/ [accessed June 12, 2017].

Van Tuinen, D., Jacquot, E., Zhao, B., Gollotte, A., and Gianinazzi-Pearson, V. (1998). Characterization of root colonization profiles by a microcosm community of arbuscular mycorrhizal fungi using 25S rDNA-targeted nested PCR. *Mol. Ecol.* 7, 879–887. doi: 10.1046/j.1365-294x.1998.00410.x

Vasconcellos, C. A., Pitta, G. V. E., Franca, G. E., and de Alves, V. M. C. (2000). Fósforo para o milho? *Cultivar* 12, 8–9.

Vega, N. W. O., Habte, M., and León Peláez, J. D. (2015). Effectiveness of a rock phosphate solubilizing fungus to increase soil solution phosphate impaired by the soil phosphate sorption capacity. *Rev. Fac. Nacl. Agron.* 68, 7627–7636. doi: 10.15446/rfnam.v68n2.50950

Wahid, F., Sharif, M., Steinkellner, S., Khan, M. A., Marwat, K., and Khan, S. (2016). Inoculation of arbuscular mycorrhizal fungi and phosphate solubilizing bacteria in the presence of rock phosphate improves phosphorus uptake and growth of maize. *Pak. J. Bot.* 48, 739–747.

Walpola, B. C., Arunakumara, K. K. I. U., and Yoon, M. H. (2014). Isolation and characterization of phosphate solubilizing bacteria (*Klebsiella oxytoca*) with enhanced tolerant to environmental stress. *Afr. J. Microbiol. Res.* 31, 2970–2978. doi: 10.5897/AJMR2013.5771

Wang, F., Shi, N., Jiang, R., Zhang, F., and Feng, G. (2016). In situ stable isotope probing of phosphate-solubilizing bacteria in the hyphosphere. *J. Exp. Bot.* 67, 1689–1701. doi: 10.1093/jxb/erv561

Whitelaw, M. A. (1999). Growth promotion of plants inoculated with phosphate-solubilizing fungi. *Adv. Agron.* 69, 99–151. doi: 10.1016/S0065-2113(08)60948-7

Yang, Y., Wang, N., Guo, X., Zhang, Y., and Ye, B. (2017). Comparative analysis of bacterial community structure in the rhizosphere of maize by high-throughput pyrosequencing. *PLoS ONE* 12:e0178425. doi: 10.1371/journal.pone.0178425

Zhang, H., Wu, X., Li, G., and Qin, P. (2011). Interactions between arbuscular mycorrhizal fungi and phosphate-solubilizing fungus (*Mortierella* sp.) and their effects on *Kostelelzkya virginica* growth and enzyme activities of rhizosphere and bulk soils at different salinities. *Biol. Fertil. Soils* 47, 543–554. doi: 10.1007/s00374-011-0563-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Silva, Medeiros, Leite, Morais, Cuadros-Orellana, Oliveira, de Paula Lana, Gomes and Dos Santos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 May 2017
doi: 10.3389/fmicb.2017.00945

Microbial Community and Functional Structure Significantly Varied among Distinct Types of Paddy Soils But Responded Differently along Gradients of Soil Depth Layers

*Ren Bai[1], Jun-Tao Wang[1], Ye Deng[2,3], Ji-Zheng He[1,4], Kai Feng[1,3] and Li-Mei Zhang[1,3]**

[1] State Key Laboratory of Urban and Regional Ecology, Research Centre for Eco-environmental Sciences, Chinese Academy of Sciences, Beijing, China, [2] Key Laboratory for Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China, [3] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, [4] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Melbourne, VIC, Australia

Paddy rice fields occupy broad agricultural area in China and cover diverse soil types. Microbial community in paddy soils is of great interest since many microorganisms are involved in soil functional processes. In the present study, Illumina Mi-Seq sequencing and functional gene array (GeoChip 4.2) techniques were combined to investigate soil microbial communities and functional gene patterns across the three soil types including an Inceptisol (Binhai), an Oxisol (Leizhou), and an Ultisol (Taoyuan) along four profile depths (up to 70 cm in depth) in mesocosm incubation columns. Detrended correspondence analysis revealed that distinctly differentiation in microbial community existed among soil types and profile depths, while the manifest variance in functional structure was only observed among soil types and two rice growth stages, but not across profile depths. Along the profile depth within each soil type, *Acidobacteria*, *Chloroflexi,* and *Firmicutes* increased whereas *Cyanobacteria*, β-*proteobacteria,* and *Verrucomicrobia* declined, suggesting their specific ecophysiological properties. Compared to bacterial community, the archaeal community showed a more contrasting pattern with the predominant groups within phyla *Euryarchaeota*, *Thaumarchaeota,* and *Crenarchaeota* largely varying among soil types and depths. Phylogenetic molecular ecological network (pMEN) analysis further indicated that the pattern of bacterial and archaeal communities interactions changed with soil depth and the highest modularity of microbial community occurred in top soils, implying a relatively higher system resistance to environmental change compared to communities in deeper soil layers. Meanwhile, microbial communities had higher connectivity in deeper soils in comparison with upper soils, suggesting less microbial interaction in surface soils. Structure equation models were developed and the models indicated that pH was the most representative characteristics of soil type and identified as the key driver in shaping both bacterial and archaeal community structure, but did not directly affect microbial functional structure. The distinctive pattern of microbial taxonomic and functional composition along soil profiles implied functional redundancy within these paddy soils.

Keywords: paddy soil, GeoChip, Mi-Seq sequencing, microbial community, soil profile, soil type, network analysis

OPEN ACCESS

Edited by:
Florence Abram,
NUI Galway, Ireland

Reviewed by:
Hamed Azarbad,
Institut National de la Recherche Scientifique – Institut Armand-Frappier, Canada
Weidong Kong,
Institute of Tibetan Plateau Research (CAS), China

****Correspondence:***
Li-Mei Zhang
zhanglm@rcees.ac.cn

Specialty section:
This article was submitted to Terrestrial Microbiology, a section of the journal Frontiers in Microbiology

Received: *16 February 2017*
Accepted: *11 May 2017*
Published: *29 May 2017*

Citation:
Bai R, Wang J-T, Deng Y, He J-Z, Feng K and Zhang L-M (2017) Microbial Community and Functional Structure Significantly Varied among Distinct Types of Paddy Soils But Responded Differently along Gradients of Soil Depth Layers. Front. Microbiol. 8:945. doi: 10.3389/fmicb.2017.00945

INTRODUCTION

Soils cover most of the natural and artificial habitats of terrestrial ecosystems. Due to the high spatial heterogeneity in soil particles and large variation of soil physiochemical properties among soil types, soils are considered harboring the most diverse microbial groups in comparison with other ecosystems (Schimel and Schaeffer, 2012). For a long time, soil scientists have noticed that the soil natural properties determined by soil parent materials during soil formation period, such as pH, texture, and base saturation, etc., sustain soil biodiversity in nature and greatly affect the basic fertility and productivity of soil to a large degree (Anderson, 1988). Furthermore, anthropogenic activities such as tillage, fertilization, irrigation, and cultivation, etc., in soils exert considerable influence on the structure and functional performance of microbial communities via changing the soil properties, thus subsequently influence soil quality in the long term (Tripathi et al., 2015). To understand the diversity of microbial community and their function structure in soils, and their correlations with soil natural properties and human activities are therefore essential to evaluate the crops productivity and environmental sustainability of soil ecosystems, since microbes are responsible for the fertility and productivity of different soil types to a large degree (Anderson, 1988; Martiny et al., 2006).

Numerous recent studies based on culture-independent techniques have suggested that the diversity and community composition of soil microorganisms on large scales were greatly driven by soil pH, and some other soil properties such as organic matter and salinity (Lauber et al., 2009; Griffiths et al., 2011; Tripathi et al., 2015). These studies mainly concentrated on the microbial distribution pattern across large geographical distances, especially focusing on the surface soil, but rarely paid attention to subtle difference of microbial community among soil types. As soil type represents a consequence of the complex influences from soil parent materials and historical and present climatic conditions, it is difficult to attribute the influence of soil type on microbial community to a single factor (Cao et al., 2012; Zhao et al., 2016). For example, some studies suggested that certain microbial taxa would prefer to specific soil types, and soil bacterial community composition was distinct among soil types but could be hardly explained by a single soil chemical parameter (Nie et al., 2012; Tripathi et al., 2012). Moreover, it was observed that soils with different parent materials overwhelmingly supported distinct bacterial community structure after similar long term cultivation, as land use or management practices may mostly shift microbial community structures in top soils (Girvan et al., 2003; Sheng et al., 2015; Sun et al., 2015). Hence it is of interest to understand how much human disturbance could affect soil microbial community comparing to the effects from soil parent materials. Furthermore, the current knowledge on the differentiation of soil microorganism among various soil types and their potential significance are still very limited and deserved to be well depicted, considering the large body of pedodiversity and microbial diversity.

In addition to soil properties determined by soil parent materials, nutrient and oxygen fluctuating along soil profile depth can subsequently lead to a change in microbial communities with soil depth (Will et al., 2010; Wang et al., 2014; Stone et al., 2015). Compared to topsoil, sub-soil volume is much greater, and thus the microbial community and its function in sub-soil is not negligible. Some investigations focusing on specific microbes such as methane-oxidizing bacteria (Reim et al., 2012), ammonium oxidizing microbes (Wang et al., 2014; Lu et al., 2015), nitrifers and denitrifiers (Qin et al., 2016), and the rates of carbon and nitrogen cycling processes suggested great variation of functional microbes and the process they mediated along soil depth gradients (Durán et al., 2017), while few works have made efforts to compare the overall pattern of microbial functional genes in different soil depth layers. Paddy soils occupied large agricultural areas in China. Although the distribution of contrasting microbial community in different paddy soils and the functional analysis of microbes have been described recently (Sheng et al., 2015; Su et al., 2015), comparative investigations linked both taxonomic and functional structure are limited. Also, some recent studies attempted to exploring the influence of different land uses on both taxonomic and functional community of microbes combining high-throughput sequencing technique with functional gene array (Paula et al., 2014; Mendes et al., 2015), but these studies were mainly carried out in the same region within similar soil types.

Comparing to different management practices and land use types, soils with water-logged paddy rice cultivation receive relatively higher uniformity and similarity of management practice. It is still less understood whether uniform management would assimilate microbial community across different soil types, and the extent of inherent influences of parent materials on shaping microbial communities. Along a paddy soil profile, water would replace the gaseous phase and thus the oxygen status varies dramatically along the soil profile (Kögel-Knabner et al., 2010), but succession of microbial community structure along this oxygen gradient has rarely been studied (Noll et al., 2005). In our previous investigation, activity and diversity of the anaerobic ammonium oxidation (anammox) bacteria were examined in three paddy soil types, and the results revealed distinct pattern of anammox activity, diversity and abundance along depth gradient in different paddy soils (Bai et al., 2015). While the study only focused on the aspect of anammox process rather than a comprehensive observation on soil microbes and their functional genes. Therefore, in the present study, Miseq high-throughput sequencing and GeoChip techniques were combined to characterize the bacterial and archaeal community composition and functional structures across four soil profile depths among three distinct paddy soil types including Inceptisol, Oxisol, and Ultisol soil orders. This study aimed to (1) elaborately depict microbial community succession among different parent materials and explore the linkage between microbial community taxonomic and functional structure, and (2) to understand how much do soil inherent properties affect microbial community composition and its functional structure, and (3) if uniform flooding management would assimilate microbial community among different soil types.

MATERIALS AND METHODS

Mesocosm Incubation and Soil Sampling

The paddy soils used for this study were freshly sampled from a greenhouse mesocosm incubation system as described in our previous study (Bai et al., 2015). Briefly, three paddy soils originally collected from Binhai (BH, 119.84° E, 34.01° N, Inceptisol), Leizhou (LZ, 110.04° E, 20.54° N, lateritic Oxisol), and Taoyuan (TY, 111.48° E, 28.90° N, Ultisol) in three rice production areas in Southeast China with the spatial distance more than 1000 km from each other, were incubated in mesocosms columns (50 cm in diameter and 70 cm in height) and received similar water, fertilization, and rice plantation management as in the field. For each soil type, two replicate columns were constructed, and one profile were sampled from each column at four soil depth intervals (A, 0–5 cm; B, 5–20 cm; C, 20–40 cm; D, 40–60 cm) at the tillering and heading growth stages of rice.

Soil Physicochemical Determinations

Soil pH and EC were measured with a pH meter and a conductivity meter, respectively. NO_3^- and NH_4^+ were extracted with 1 M KCl and determined by using a flow analyzer (AA3, SEAL analytical, Germany). Soil hot water-extractable carbon (HWC) was extracted with water at 70°C and determined by a carbon-nitrogen analyzer. Soil total C, N, and S were measured using an elemental analyzer (Vario EL III-Elementar, Germany). The concentrations of dissolved O_2 along the soil profiles were also measured with an O_2 microsensor electrode (Unisense, Denmark) precisely positioned by a micromanipulator (Unisense).

DNA Extraction and Purification

The total genomic DNA of microbes was extracted from 5 g of dry soil by using a protocol that included liquid nitrogen grinding and sodium dodecyl sulfate as previously described (Zhang et al., 2013). To remove humus and protein components, the DNA was purified with 0.5% low melting point agarose gel, and further purified by a phenol-chloroform-butanol extraction procedure. DNA quantity and quality were evaluated using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA), and all the DNA sample reached final A260/A280 and A260/A230 ratios of >1.7 and 1.8. All DNA samples were stored at −40°C before downstream analysis.

DNA Microarray Hybridization, Scanning and Data Processing

Geochip 4.2 was utilized for analyzing functional structures of soil microbes. The experiments were carried out as previously described (Yang et al., 2013). Briefly, DNA samples were firstly labeled with Cy-5 fluorescent dye using a random priming method. Then the labeled DNA samples were purified with a QIA purification kit (Qiagen, Valencia, CA, USA) and further dried in a SpeedVac (ThermoSavant, Milford, MA, USA) at 45°C for 45 min. Hybridization buffer containing 40% formamide, 25% SSC, 1% SDS, 10 μg of unlabeled herring sperm DNA (Promega, Madison, WI, USA) was added to the dried DNA samples, then vortexed, spun down and incubated at 95°C for 5 min. Hybridizations were performed with a MAUI hybridization station (BioMicro, Salt Lake City, UT, USA). Subsequently, the microarray was scanned by a 100% laser power and 100% photomultiplier tube with a NimbleGen MS 200 Microarray Scanner (Roche, Madison, WI, USA) and signal intensities were quantified.

The obtained raw data was pre-analyzed and denoised with the online pipeline provided by Microarray Data manager[1]. Each sample was treated separately in this procedure. Briefly, values with signal-noise ratio less than 2 were firstly removed. Subsequently, the values of the signals detected by the probes within each sample were normalized with the lnMR method that ln(x+1) was divided by mean of total signal intensity of each sample, where x denotes a detected value by each probe within each sample.

Illumina Mi-Seq Sequencing and Sequence Analysis

The V4–V5 region of bacterial and archaeal 16S rRNA genes were sequenced with primer sets 515F (5′-GTGCCAGCMGCCGCGGTAA-3′)/806R (5′-GGACTACHVG GGTWTCTAAT-3′) and Arch519F (5′-CAGCCGCCGCGG TAA-3′)/Arch 915R (5′-GTGCTCCCCCGCCAATTCCT-3′), respectively. Based on the requirement of Illumina sequencing, Illumina adaptor A was added to the 5′-ends, and Illumina adaptor B and barcode were added to the 3′-ends of primers. The raw sequencing data were further processed with QIIME software. The FLASH method was firstly carried out to assemble the obtained sequences, and then the UPARSE method was utilized to filter chemira and repetitive sequences. The OTUs of bacterial and archaeal sequences were defined according to a similarity of 97%, and OTU tables were established for further analysis. The identification of bacterial and archaeal taxa was conducted based on the Greengenes database.

Statistical Analysis

Pearson correlation analysis and analysis of differences was carried out with Duncan analysis of one-way ANOVA by using SPSS software (IBM, version 19.0). The vegan package of R software (version 3.2.2) was utilized for conducting detrended correspondence analysis (DCA) which was used to study community and functional structures of soil microbes. Structural equation models (SEM) were developed according to an *a priori* model to determine the direct and indirect contribution of soil type, soil depth, pH, salinity, and HWC on microbial community and functional structures (as assessed by the first principal coordinate of the Bray–Curtis dissimilarity matrix). Soil type and depth were set as exogenous variables that latitude and longitude (first principal coordinate of the Bray–Curtis dissimilarity matrix) of the three soil sampling sites were substituted for soil type and soil depth was set as 5, 20, 40, and 60. SEM analysis was conducted by utilizing AMOS 22.0 (Amos, Development Corporation, Meadville, PA, USA).

[1]http://ieg.ou.edu/microarray/

Adequate model fits were examined according to non-significant Chi-square test ($P > 0.05$), goodness fit index (GFI), Akaike value (AIC), and root mean square error of approximation (RMSEA).

Construction of Phylogenetic Molecular Ecological Networks

Phylogenetic molecular ecological networks (pMENs) were constructed based on random matrix theory with the bacterial and archaeal OTUs obtained in profile A and D from the three paddy soils in this study (Deng et al., 2012). Network construction was carried out by utilizing the online pipeline provided by the Institute of Environmental Genomics, University of Oklahoma[2].

The average connectivity index (avgK) and modularity index were utilized to describe the topology structure of the networks. The avgK was used to describe the complexity of the networks, and modularity was utilized as a measurement of system resistance. The topological roles of different nodes were divided into four sub-categories based on the within-module connectivity (Z_i) and the among module connectivity (P_i): (1) nodes with $Z_i > 2.5$ and $P_i < 0.62$ were defined as module hubs; (2) nodes with $Z_i > 2.5$ and $P_i > 0.62$ were defined as network hubs; (3) nodes with $Z_i < 2.5$ and $P_i < 0.62$ were defined as peripherals; (4) nodes with $Z_i < 2.5$ and $P_i > 0.62$ were defined as connectors (Olesen et al., 2006).

Networks Analysis of Soil Microbial Community and Functional Genes

The associations of microbial taxa and functional genes were analyzed using the Cytoscape plug-in CoNet (Soffer et al., 2015). The detected taxa and genes with a minimum occurrence of four across all the samples within certain depth layer were discarded in prior to calculation in order to minimize the artificial association bias (Li et al., 2015). The pairwise calculation was performed at the phylum level for archaea and bacteria expect that *Proteobacteria* were analyzed on the level of class, while functional genes were employed according to the gene category provided by functional gene array (Bai et al., 2013). Combination of correlation scores and *P*-values of Spearman correlation, Pearson correlation, Kullback–Leibler dissimilarity, and Bray–Curtis dissimilarity were utilized for all the pairwise correlation. Potential false-positive correlations and compositionality biases were eliminated by ReBoot procedure with 100 permutations, and the resultant distribution was further refined with 100 bootstraps. Then *Brown* method was utilized to combine the *P*-values for the four correlations measurement, and correlations found to be significant by less than two methods were discarded (Soffer et al., 2015). Only correlations with a coefficient above 0.8 and a significance level below 0.05 were considered statistically robust which were finally displayed as previously reported (Hu et al., 2016). The obtained pairwise correlations were used to construct the co-association networks. Network topology was explored using Cytoscape software and the Network Analyzer plug-in and was illustrated on the open-source interactive platform Gephi (Bastian and Heymann, 2009).

[2]http://ieg.ou.edu

RESULTS

Soil Physical and Chemical Properties

Chemical properties of the three soil columns are shown in **Table 1**. Significant differences with respect to properties were observed between soil types and along profiles within each soil type. Soil water content decreased dramatically along soil profiles. The pH ranged from 8.0 to 8.6 in BH columns, ranged from 6.8 to 7.2 in LZ columns, and ranged from 5.8 to 6.0 in TY soils. Soil salinity which was reflected by soil electrical conductivity (EC) was approximately 10-fold lower in TY than in BH and LZ soils. HWC peaked in the first two soils layers of the three soil types, and was 2- to 8-folds higher in BH and TY soils than in the LZ soil. DCA for measured soil properties in the three paddy soils showed a clear division in soil characteristics among the three paddy soils. In addition, a succession of soil chemical parameters along soil depth was observed along soil depth within each soil type (Supplementary Figure 1).

Diversity Based on the 16S rDNA and Microarray Analysis

A total of 48 samples covering 3 soil types, 4 depth layers, 2 mesocosm replicates, 2 growth stages were subjected to MiSeq sequencing and microarray assay to explore the diversity of species and functional genes. A total of 26075 bacterial OTUs and 3404 archaeal OTUs were generated after resampling with 42364 bacterial reads and 9456 archaeal reads per sample, respectively. No significant difference was found in alpha diversity of bacterial community except sample LZ-A presented the lowest Shannon indices. While Shannon indices of archaeal community was the lowest in layer A of BH and TY soils, and the highest and lowest Simpson indices were detected in TY-C and LZ-D, respectively (**Table 2**).

The number of detected genes, Shannon index, inverse Simpson index and Simpson evenness by Geochip were calculated to evaluate the functional diversity and structure of microbial communities. In total, 44350 genes were detected and ranged from 36860 to 40612 in different samples. Shannon and Simpson diversity in LZ samples were significantly lower than in BH and TY samples, and showed no significant difference between BH and TY samples. Approximately 80% of functional genes were shared by three soil types while around 5% of the genes were unique within each soil type (**Table 3**).

Community Similarity Based on the 16S rDNA and Microarray Analyses

Detrended correspondence analysis analysis of 16S rDNA Mi-seq sequences data indicated that bacterial and archaeal communities were clearly separated into three groups according to soil types, indicating the effects of soil type on microbial community composition (**Figures 1A,B**). A clear trend of community succession along soil profiles was also observed in the three soil types, with surface soils (0–5 cm) separated from sub-soils (5–20 cm, 20–40 cm, and 40–60 cm) markedly, and community structure in surface soils was much more

TABLE 1 | Physicochemical properties of soil samples from greenhouse mesocosm incubation*.

Sample Name	Profile	HWC	pH	NH4	NO3	EC	OM
BH-A	0–5 cm	181.24 ± 20.89^{b}	7.96 ± 0.07^{c}	15.58 ± 7.19^{a}	0.7 ± 0.77^{b}	0.33 ± 0.13^{ab}	2.36 ± 0.27^{cde}
BH-B	5–20 cm	155.98 ± 16.23^{bc}	8.23 ± 0.02^{b}	7 ± 3.66^{b}	0.18 ± 0.13^{b}	0.38 ± 0.09^{a}	2.86 ± 0.13^{b}
BH-C	20–40 cm	88.43 ± 26.11^{cd}	8.51 ± 0.07^{a}	7.75 ± 2.14^{b}	0.2 ± 0.2^{b}	0.29 ± 0.03^{ab}	1.63 ± 0.37^{f}
BH-D	40–60 cm	62.13 ± 14.3^{d}	8.62 ± 0.04^{a}	6.65 ± 0.89^{b}	0.21 ± 0.18^{b}	0.26 ± 0.02^{ab}	1.48 ± 0.27^{f}
LZ-A	0–5 cm	63.33 ± 40.85^{d}	7.02 ± 0.21^{de}	5.81 ± 3.78^{b}	1.28 ± 1.38^{ab}	0.35 ± 0.2^{ab}	2.61 ± 0.07^{bc}
LZ-B	5–20 cm	51.6 ± 18.62^{d}	6.79 ± 0.09^{f}	7.54 ± 3.98^{b}	1.17 ± 1.18^{ab}	0.25 ± 0.03^{ab}	2.5 ± 0.21^{cd}
LZ-C	20–40 cm	45.93 ± 6.95^{d}	6.85 ± 0.14^{ef}	9.36 ± 2.75^{ab}	0.9 ± 0.82^{ab}	0.23 ± 0.03^{b}	2.32 ± 0.17^{cde}
LZ-D	40–60 cm	63.79 ± 26.6^{d}	7.15 ± 0.23^{d}	7.48 ± 4.33^{b}	1.11 ± 1.22^{ab}	0.25 ± 0.05^{ab}	2.49 ± 0.22^{cd}
TY-A	0–5 cm	409.93 ± 139.76^{a}	5.96 ± 0.21^{gh}	9.25 ± 5.97^{ab}	2.6 ± 2.83^{a}	0.06 ± 0.05^{c}	3.74 ± 0.17^{a}
TY-B	5–20 cm	402.87 ± 40.98^{a}	5.81 ± 0.13^{h}	9.35 ± 1.6^{ab}	0.79 ± 0.63^{ab}	0.03 ± 0.02^{c}	3.51 ± 0.08^{a}
TY-C	20–40 cm	117.03 ± 31.37^{bcd}	6.09 ± 0.1^{g}	8.61 ± 4.59^{b}	0.76 ± 0.5^{ab}	0.02 ± 0.01^{c}	2.27 ± 0.10^{de}
TY-D	40–60 cm	110.2 ± 27.61^{bcd}	6.04 ± 0.09^{g}	9.24 ± 3.5^{ab}	0.8 ± 0.45^{ab}	0.02 ± 0.01^{c}	2.15 ± 0.11^{e}

Sample Name	Profile	AOFe	DCBFe	S	interMn	reducMn	Water content
BH-A	0–5 cm	1.24 ± 0.12^{d}	4.66 ± 0.11^{d}	0.031 ± 0.004^{d}	0.12 ± 0.011^{c}	0.0067 ± 0.0005^{a}	1.08 ± 0.42^{b}
BH-B	5–20 cm	1.55 ± 0.38^{c}	5.19 ± 0.61^{d}	0.032 ± 0.002^{d}	0.12 ± 0.004^{c}	0.0068 ± 0.0005^{a}	0.5 ± 0.04^{c}
BH-C	20–40 cm	0.84 ± 0.11^{e}	4.38 ± 0.12^{d}	0.028 ± 0.004^{d}	0.15 ± 0.008^{c}	0.0057 ± 0.0007^{b}	0.33 ± 0.01^{c}
BH-D	40–60 cm	0.68 ± 0.04^{e}	4.2 ± 0.06^{d}	0.024 ± 0.002^{d}	0.16 ± 0.01^{c}	0.0053 ± 0.0006^{b}	0.34 ± 0.01^{c}
LZ-A	0–5 cm	2.6 ± 0.07^{a}	121.63 ± 4.51^{a}	0.043 ± 0.005^{a}	0.31 ± 0.032^{a}	0.0051 ± 0.0005^{b}	1.2 ± 0.25^{b}
LZ-B	5–20 cm	2.34 ± 0.07^{b}	123.63 ± 8.99^{a}	0.048 ± 0.003^{a}	0.24 ± 0.016^{b}	0.0036 ± 0.0005^{c}	0.57 ± 0.01^{c}
LZ-C	20–40 cm	2.34 ± 0.12^{b}	118.31 ± 1.75^{a}	0.043 ± 0.003^{a}	0.26 ± 0.032^{ab}	0.0041 ± 0.0003^{c}	0.52 ± 0.01^{c}
LZ-D	40–60 cm	2.28 ± 0.07^{b}	123.73 ± 10.26^{a}	0.041 ± 0.006^{a}	0.31 ± 0.12^{a}	0.0051 ± 0.0006^{b}	0.49 ± 0.04^{c}
TY-A	0–5 cm	1.28 ± 0.05^{d}	19.04 ± 0.5^{c}	0.039 ± 0.001^{c}	0.026 ± 0.004^{d}	0.0014 ± 0.0001^{d}	2.18 ± 0.81^{a}
TY-B	5–20 cm	1.15 ± 0.08^{d}	18.63 ± 0.77^{c}	0.037 ± 0.003^{c}	0.021 ± 0.001^{d}	0.0014 ± 0.0002^{d}	0.58 ± 0.02^{c}
TY-C	20–40 cm	2.48 ± 0.14^{ab}	30.27 ± 0.93^{b}	0.022 ± 0.002^{b}	0.046 ± 0.002^{d}	0.0039 ± 0.0003^{c}	0.38 ± 0.03^{c}
TY-D	40–60 cm	2.41 ± 0.06^{ab}	29.99 ± 0.59^{b}	0.021 ± 0.002^{b}	0.045 ± 0001^{d}	0.0036 ± 0.0005^{c}	0.41 ± 0.02^{c}

A, B, C, and D represent soil sampling layers 0–5 cm, 5–20 cm, 20–40 cm, and 40–60 cm (the same below); HWC, hot water-extractable organic carbon (mg kg^{-1}); NH4, NH_4^+concentration (mg kg^{-1}); NO3, NO_3^- concentration (mg kg^{-1}); EC, electrical conductivity (d Sm^{-1}); OM, organic matter (g kg^{-1}); N, total nitrogen (%); C, total carbon (%); S, total sulfur (%); C/N, carbon to nitrogen ratio; AO-Fe, ammonium oxalate extractable Fe (mg kg^{-1}); DCB-Fe, DCB extractable Fe (mg kg^{-1}); interMn, interchangeable Mn (mg kg^{-1}); reducMn, reduceable Mn (mg kg^{-1}); Water content, water content.
The different letters indicate significant differences of measured values within all the samples by one-way ANOVA.

TABLE 2 | Number of observed species, indices of Shannon and Simpson diversity of bacterial and archaeal community in four soil profiles of three paddy soils.

	Bacteria			Archaea		
Sample	Observed species	Shannon	Simpson	Observed species	Shannon	Simpson
BH-A	2906	10.32 ab	1.00 a	61	2.72 e	0.74 de
BH-B	3834	10.99 a	1.00 a	294	4.23 bc	0.84 bc
BH-C	3965	10.97 a	1.00 a	319	4.99 a	0.91 ab
BH-D	3644	10.76 a	1.00 a	315	5.11 a	0.92 ab
LZ-A	2518	8.42 b	0.90 a	125	3.88 cd	0.86 abc
LZ-B	2672	9.86 ab	1.00 a	152	3.84 cd	0.84 bc
LZ-C	2633	9.86 ab	1.00 a	161	4.03 cd	0.84 abc
LZ-D	2643	9.91 ab	1.00 a	144	3.55 d	0.79 d
TY-A	2985	9.28 ab	0.93 a	192	3.46 d	0.66 cd
TY-B	3106	10.16 ab	0.99 a	222	4.70 ab	0.90 ab
TY-C	2668	9.96 ab	1.00 a	232	5.00 a	0.93 a
TY-D	2507	9.57 ab	0.99 a	207	4.84 ab	0.92 ab

The different letters indicate significant differences of measured values within all the samples by one-way ANOVA.

heterogeneous compared to deeper soil layers (Supplementary Figure 2). Given the relative low variation in bacterial and archaeal community structure at the two paddy rice growth stages, samples within the same depth of each soil at the two time points were considered as replicates in the following analysis for the 16S rDNA data.

TABLE 3 | Shared and endemic genes among soil types, and diversity indice of functional genes.

Sample name	BH	LZ	TY
BH	**2264 (5.57%)**	*34397 (79.85%)*	*37842 (87.58%)*
LZ		**1141 (3.10%)**	*35213 (83.67%)*
TY			**1275 (3.15%)**
Genes detected	40612	36860	40439
Shannon index	10.36 a	10.22 b	10.34 a
Inverse Simpson index	31536.83 a	27739.00 b	31171.78 a

Bold numbers denote the number and percentage of endemic genes within each soil types, while italic numbers denote the number and percentage of shared genes between soil types.
The different letters indicate significant differences of measured values within all the samples by one-way ANOVA.

Detrended correspondence analysis based on functional gene array data showed that samples from BH, LZ, and TY soils were well separated from each other by DCA I and DCA II, and the distance between BH and TY samples was closer than to LZ samples (**Figure 1C**). Unlike community pattern based on 16S rDNA, however, there was no clear differentiation in the structure of functional community between profile layers within each of the soil types. By contrast, samples of tillering and heading stages formed two clusters, which implied the significant difference of soil microbial functional structure between the two growth stages of paddy rice (**Figure 1C** and Supplementary Figure 2).

Phylogenetic Molecular Network Analysis of Microbial Communities

To further understand the similarity of microbial communities, networks of bacteria and archaea from the three soil profiles were constructed with OTU data obtained from Mi-seq sequencing (**Table 4**). In respect of the bacterial networks, the network size indicated by the numbers of nodes was highest (206) in bacteria community from layer B (**Table 4**). The coefficient of positive correlation among bacterial community in layer A was up to 93%, and 18–27% higher than in other layers. However, the network for bacteria in layer A had the lowest average degree (avgK = 3.68), which suggested higher connectivity of the nodes in deeper soil layers. Also, bacterial community in layer A presented the highest level of modularity, indicated its higher system resistance to changes in comparison with other layers. Meanwhile, module hubs, representing key nodes species in the networks, were detected in all the layers. Surface soil harbored the most module hubs (nine hubs) among the soil layers and layer B and C harbored six and eight hubs, while only one hub was defined in the bottom soils, suggesting a lower amount of generalization in this soil depth (**Figure 2A** and **Table 4**).

Depth-effects on the connection among microbes were also observed in the networks of the archaeal community. Similar to bacteria, the lowest avgK (6.12) but highest modularity (0.45) in the networks of archaeal community were recorded in layer A (**Table 4**), indicating that archaea had a lower connectivity in surface soils than in deeper soil layers and possessed a higher system resistance in comparison with other profile depths. In contrast to bacterial networks, module hub was hardly detected in all the archaeal networks (**Table 4**), suggesting the lack of generality in the archaeal community in these soils. However, more connector nodes between modules were identified in the archaeal community, especially in the bottom layer (**Figure 2B**). Although there was a slight decrease in the coefficient of positive correlation among archaeal nodes from profiles A to C, the highest coefficient was detected in the bottom layer of the archaeal community (**Table 4**).

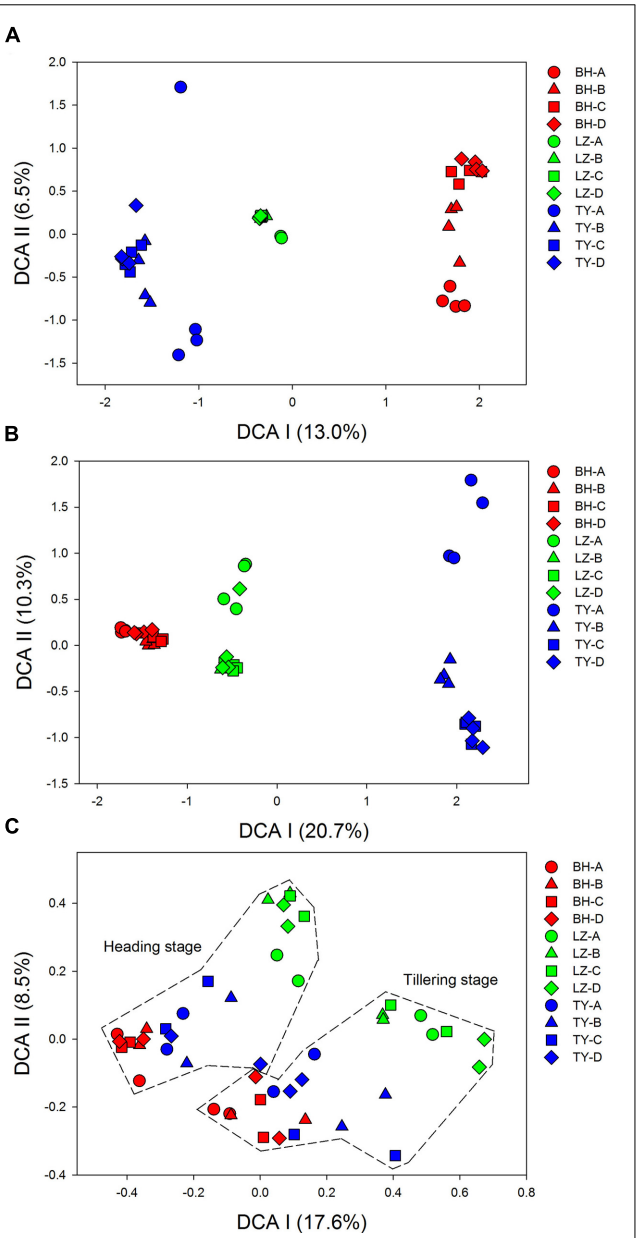


FIGURE 1 | Detrended correspondence analysis (DCA) for microbial taxonomic and functional community. Results of **(A)** bacterial community, **(B)** archaeal community, and **(C)** all of the functional genes from four soil profiles (A, B, C, D) of three paddy soils (BH, LZ, TY).

TABLE 4 | Major topological properties of phylogenetic molecular ecological networks of bacterial and archaeal communities in four profile layers (A to D) of the three paddy soils.

Community	Similarity threshold	Network size	Total links	No. of module hubs	Percentage of positive links	Percentage of negative links	avgK	Modularity
Bacteria (A)	0.82	191	351	9	93%	7%	3.68	0.68
Bacteria (B)	0.85	166	452	6	74%	26%	5.45	0.63
Bacteria (C)	0.83	206	527	8	66%	34%	5.12	0.61
Bacteria (D)	0.83	159	478	1	75%	25%	6.01	0.54
Archaea (A)	0.76	51	156	0	63%	37%	6.12	0.45
Archaea (B)	0.75	77	344	2	61%	39%	8.94	0.38
Archaea (C)	0.77	84	446	0	58%	42%	10.62	0.31
Archaea (D)	0.74	103	479	0	74%	26%	9.30	0.41

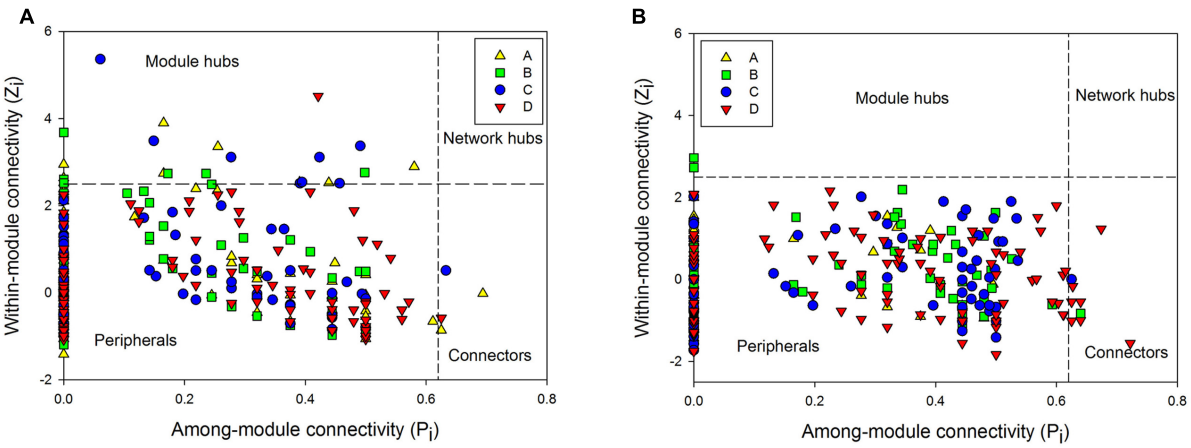


FIGURE 2 | The Z-P plot indicating categories of nodes according to Z_i (within-module connectivity) and P_i (among-module connectivity) of bacterial **(A)** and archaeal community **(B)** in four soil profile layers.

Bacterial and Archaeal Community Composition Based on Mi-Seq Sequencing

Totally, 34 bacteria phyla were identified in all 48 samples of three soil types. The dominant phyla, including *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Firmicutes*, *Planctomycetes*, *Proteobacteria,* and *Verrucomicrobia*, accounted for more than 95% of bacterial sequences (**Figure 3A**). The relative abundance of individual bacterial taxa varied distinctly among different paddy soil types and between surface and subsurface layers. Of the eight abundant bacteria phyla, the relative abundance of *Acidobacteria*, *Chloroflexi,* and *Firmicutes* were the lowest in the three surface soils (11.2–14%, 3.5–7.4%, and 3.6–5.8%, respectively) but were relatively higher in all the deep soil layers (11.2–29%, 7.7–17.8%, and 5.5–12.6%, respectively), and positively correlated with profile depth ($P < 0.01$, $n = 48$, Supplementary Table 1), also negatively correlated to soil HWC ($P < 0.01$, $n = 48$, Supplementary Table 1). Conversely, the abundance of *Verrucomicrobia* and β-*proteobacteria* decreased along soil depth in all the soil profiles and both positively correlated to the fluctuation of soil HWC along depth ($P < 0.01$ and $P < 0.05$, $n = 48$, Supplementary Tables 1, 2). *Cyanobacteria* mainly presented in surface soil layer with a proportion of 2.2–5.8% rather than 0.2–0.9% in deeper soils (**Figure 3A**).

As for archaea, total 3404 OTUs were retrieved and classified into four phyla, including *Thaumarchaeota, Euryarchaeota, Crenarchaeota,* and *Parvarchaeota* (**Figure 3B**). The *Thaumarchaeota*, which composed of *Nitrososphaerales*, *Cenarchaeaceae,* and SAGMA-X, was the most abundant phylum and predominated in the surface layers of three soil types with a relative abundance of 51.8–76.7%, but accounted for a relatively lower proportion between 29.1 and 70% in subsurface layers of three soil types (**Figure 3B**). *Nitrososphaerales* accounted for 30–45.1%, and 31.8–69.7% of *Thaumarchaeota*-affiliating sequences in BH and LZ profile soils, respectively, while it only accounted for 3.3–23.2% in the TY profile soil (**Figure 3B**). *Euryarchaeota* was the second abundant archaeal phylum and mainly composed of classes *Methanobacteria*, *Methanomicrobia,* and *Thermoplasmata*. *Methanobacteria* was approximately 2–3 times higher in BH soils than in LZ and TY soils, *Methanomicrobia* and *Thermoplasmata* were more frequently detected in the sub layers of the TY soil (**Figure 3B**). Converse to *Thaumarchaeota* phylum, the phylum *Crenarchaeota* accounted for a much lower proportion in surface layer than deep layers in

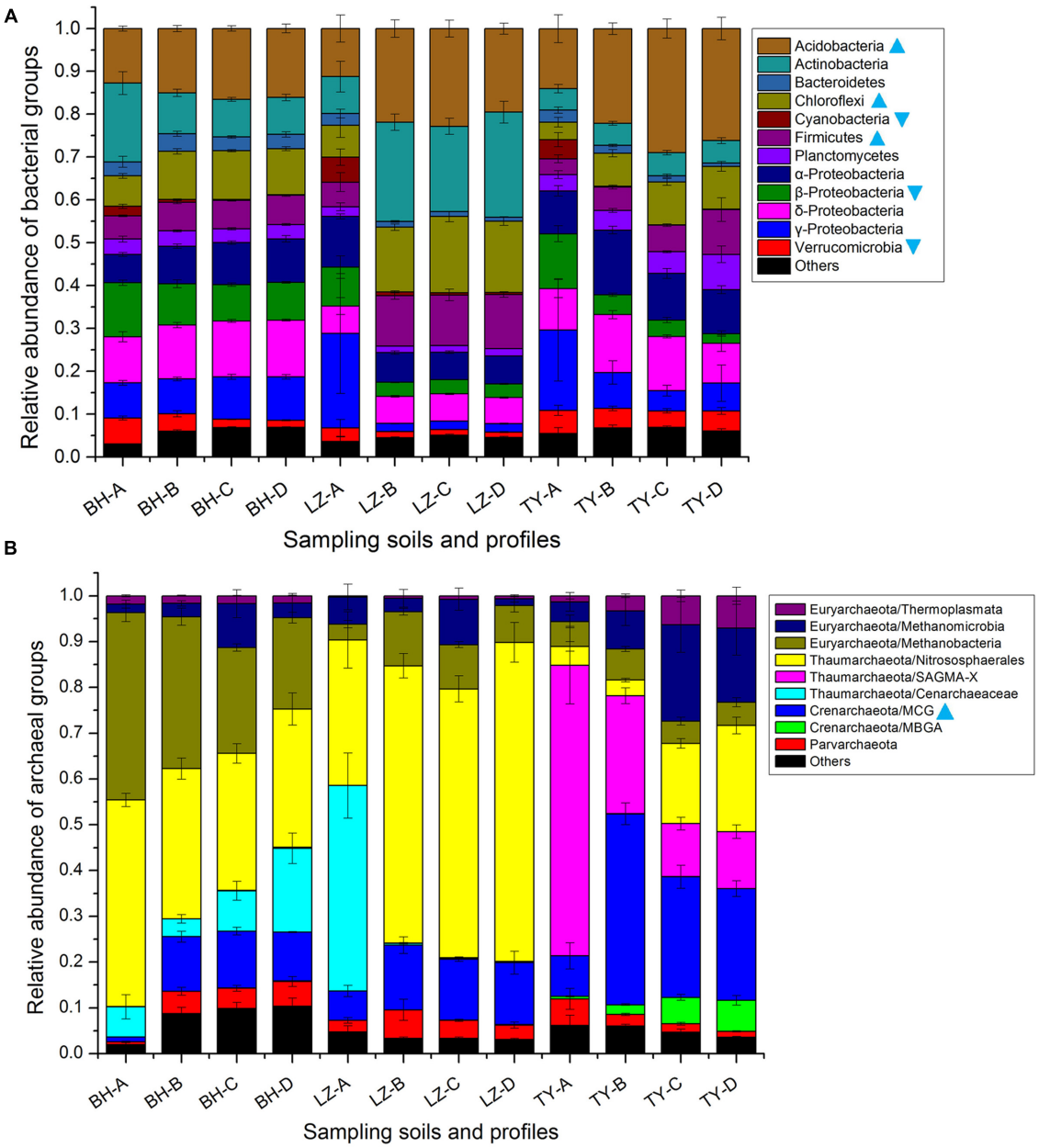


FIGURE 3 | Dominant taxonomic groups of bacterial **(A)** and archaeal **(B)** groups in four soil profiles of three paddy soils. Cyan triangles denote that the relative abundance of the group uniformly decreased along depth layers within each soil type, cyan inverted triangles denote that the relative abundance of the group uniformly increased along depth layers.

three soil types, and was more abundant in TY deep layers than in corresponding layers of BH and LZ columns (**Figure 3B**). The phylum *Crenarchaeota* was composed of MCG group and MBGA group. MBGA group was only detected in TY profiles with an increasing trend from 3 to 56% along the depths (**Figure 3B**), while MCG group occupied much lower abundance (1–8.9%) in surface soils than in sub-soils (10.7–41.7%) in all three soil types. In contrast, the group SAGMA-X of *Thaumarchaeota* was predominant in the TY surface soil (63.4%) but sharply decreased in deep layers (11.5–25.8%), while only occupied 0.01–0.06% proportion of BH and LZ archaeal communities (**Figure 3B**). All these further suggested distinct bacterial and archaeal community composition among the three soil types and the depth-dependent distribution pattern of some microbial groups.

Pattern of Functional Gene Categories

Microbial functional genes were categorized based on the major metabolic processes to understand the pattern of functional microbial communities in different soil types and depth layers.

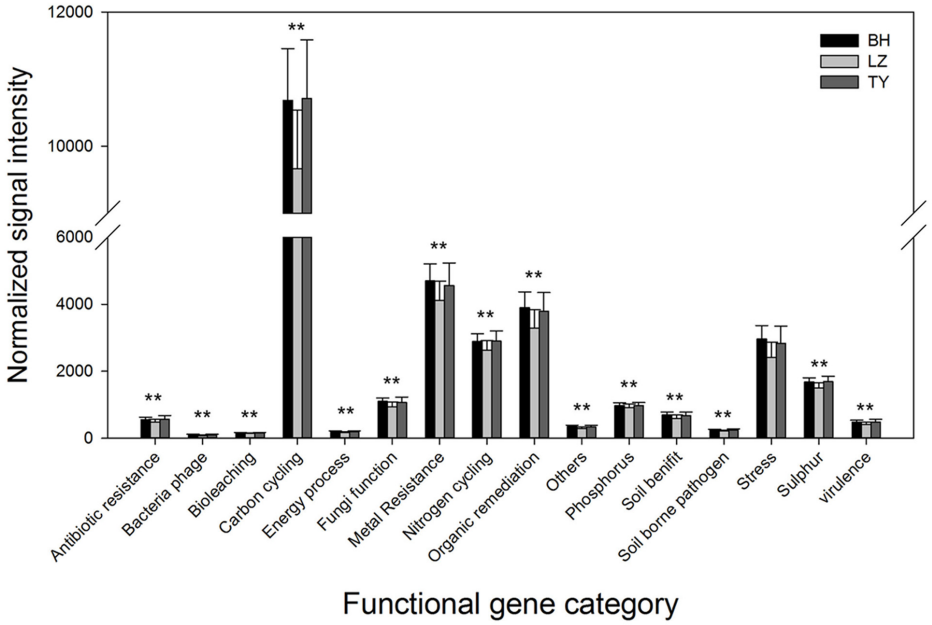


FIGURE 4 | Abundance of different of categories of microbial functional genes in three paddy soils. ** Denotes that the abundance is significantly higher in BH and TY than in LZ soils.

Genes involved in carbon cycling were most abundant in all the samples, followed by genes related to metal resistance, organic remediation, and nitrogen cycling which also presented at a high level (**Figure 4**). DCA analysis did not identify significant differences in the gene abundance of different functional categories among depth layers within each soil type (data not shown). Normalized gene signal intensity of all the functional gene categories from different soil layers within each column were therefore combined. Results from one-way ANOVA indicated that the normalized signal intensity of most of the gene categories in BH and TY columns were significantly higher than in LZ columns (**Figure 4**).

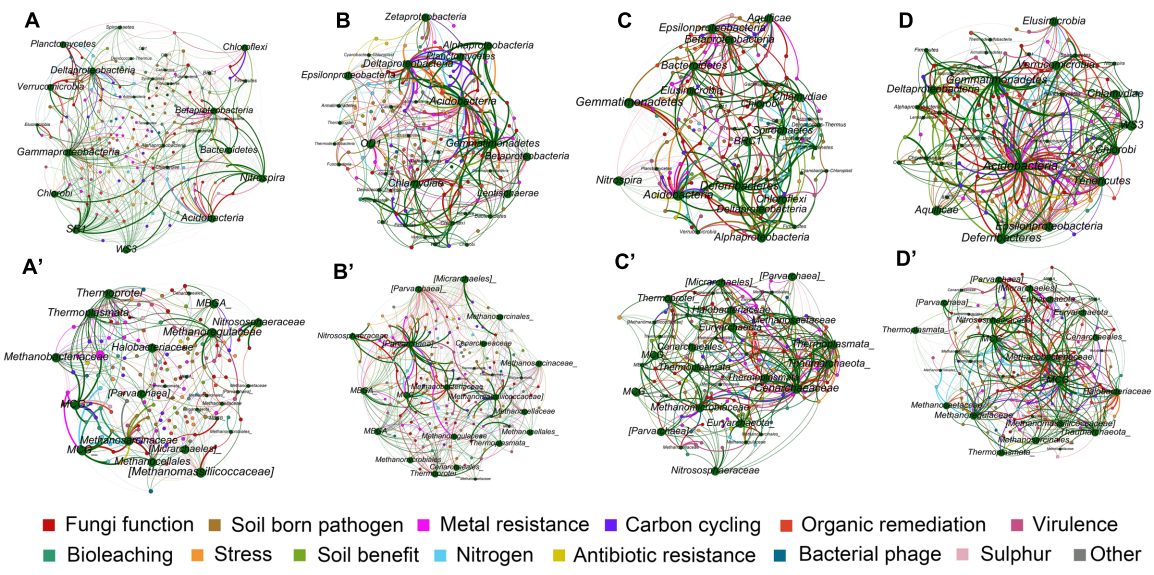


FIGURE 5 | Mutualistic networks of interaction between bacterial community and functions of the four soil profile layers **(A–D)**, and archaeal community and functions **(A'–D')**. Edges in green colors represent influence of microbes to ecological functions, and edges in other colors represent influence of each functional category to microbes.

TABLE 5 | Major topological properties of mutualistic networks between archaeal communities and functions, and between bacterial communities and functions from profiles A to D of the three paddy soils.

Community	No. of nodes	No. of edges	Average degree	Modularity
Bacteria (A)	194	541	2.79	0.57
Bacteria (B)	194	436	2.25	0.45
Bacteria (C)	179	413	2.31	0.53
Bacteria (D)	184	466	2.53	0.44
Archaea (A)	163	353	2.09	0.55
Archaea (B)	186	528	2.84	0.48
Archaea (C)	182	396	2.18	0.49
Archaea (D)	185	504	2.72	0.43

Linkage among Microbial Community, Functional Community, and Soil Properties

Network analysis of microbial taxa and functional genes were generated to estimate the potential linkages between the microbial taxonomic community and overall functional structure obtained from GeoChip analysis (**Figure 5**). The average degree of the networks indicated the amount of total connections between microbes and their functions. For bacterial community, the highest average degree was detected in layer A (2.79), while archaeal community had lowest average degree in layer A (2.09) but relatively higher degree in both layers B and D (**Table 5**), indicating that bacterial community had complex connection with their functions in surface soil while archaea performed complex connections in deeper soils.

The number of edges that linked microbial taxonomic nodes with functional nodes reflects the linkage between microbial groups and functions, and the thickness of the edges showed the intensity of connection between nodes. Although layers B, C, and D of bacterial networks, and layers A and C of archaea presented overall lower average degrees (**Figure 5** and **Table 5**), stronger connections indicated by number of edges between nodes were observed in these networks (Supplementary Table 3). Generally, the numbers of edges of the main dominant bacterial phyla detected by sequencing were similar among the profile layers (**Figure 5** and **Table 5**). Specifically, *Acidobacteria* had relatively strong connections to ecological functions across all soil profile layers, while intensive connections between functions and some microbial groups such as δ-*Proteobacteria*, γ-*Proteobacteria*, *Planctomycetes,* and *Nitrospira* peaked in surface soils or upper layers (Supplementary Table 3). Most of the archaea groups presented weak connection with functions in the surface soils, while *Methanomicrobia* and *Parvarchaeota* in layers B and D, and *Methanobacteria* and MCG in layer D had strong connections with functions, and occupied nearly half of the archaeal connections in the two layers (Supplementary Table 4).

We further developed structure equation models (SEM) to explore how microbial communities and function associated. Parameters in the models included geographical distance of the soil originally located, soil depth, HWC, pH, and EC. The model for bacteria explained 98 and 38% of the variation in bacterial community and functional structure, respectively, along four depth layers of three paddy soils (**Figure 6A**). Soil pH, HWC, and geographical location were identified as the significant factors that shape the bacterial community structure, while microbial functional structure was solely affected by bacterial community (**Figure 6A**). Standardized total effects (including direct and indirect effects) obtained from standardized SEM indicated that soil pH contributed more to bacterial community structure than geographical distance and HWC (Supplementary Figure 3A). The model for archaea explained 91 and 25% of the variation in archaeal community and microbial functions, respectively. Archaeal community was mainly influenced by spatial distance, soil pH and EC (**Figure 6B**). Similar to bacteria, standardized SEM revealed that pH had stronger impact on archaeal community structure than geographical distance and EC (Supplementary Figure 3B).

DISCUSSION

Distinct Differentiation of Microbial Community Structure Within Soil Types and Depth Layers Revealed by 16S rDNA and Network Analysis

One of the main purposes of this study was to reveal the effects of soil parent material and profile depth on both microbial community and functional structures. In particular, Illumina Mi-Seq sequencing combining GeoChip techniques and further network analysis facilitated the work and revealed manifest differentiation of bacterial and archaeal community structure among three paddy soil types and between surface and deeper soils within each soil type. Some previous investigations have concerned the microbial community characteristics across different sampling sites and depths and found that layer depths had overwhelming influence on microbial community structure than sampling location (Fierer et al., 2003; Hartmann et al., 2009; Eilers et al., 2012; Steven et al., 2013). However, these studies were mainly carried out in sites with similar soil properties. The three soil types tested in the present study represented distinct pedogenic history and therein possessed varied soil chemical properties. Further SEM analysis identified soil pH, HWC and salinity as the critical influencing factors significantly influencing microbial community structure (**Figure 6**). Specifically, soil types had a greater direct effect on pH ($P < 0.001$) comparing to soil salinity ($P < 0.01$) and HWC ($P > 0.05$) while pH was also significantly adjusted by HWC and salinity, and pH had stronger impact on bacterial and archaeal community than HWC and salinity in SEMs (**Figure 6** and Supplementary Figure S3). Generally, pH is the most representative characteristics of various parent materials of the soil chemical parameters (Imaya et al., 2005). Accordingly, it is not surprising to see that soil parent materials significantly affect microbial community structure via the determination of soil pH in this study. It either well explained why the observed variation of bacterial and archaeal community

structure in this study were less among profile depths than among soil types, since the fluctuation of pH was smaller along profiles than that among soil types (7.96–8.62 in BH profile; 6.79–7.15 in LZ profile; 5.96–6.09 in TY profile). The results were consistent with numerous previous studies showing that soil pH is a critical factor in determining soil microbial community structure on large scale (Lauber et al., 2009; Griffiths et al., 2011; Kuramae et al., 2012; Thomson et al., 2015; Tripathi et al., 2015).

Furthermore, although the main division was caused preferentially by soil types, bacterial and archaeal community structure significantly changed along the four profile depths in this study. Particularly, a large separation of microbial community between surface (0–5 cm) and deep layers (5–60 cm) within each soil was indicated in DCA analysis (**Figure 2** and Supplementary Figure S1), which was most likely caused by more uniform physicochemical characteristics shared by deeper soil layers (Eilers et al., 2012). SEM analysis clearly showed that soil HWC was greatly driven by depth and significantly affected bacterial community in this study (**Figure 6A**). Similarly, Hartmann et al. (2009) also suggested that the availability of resources is one of the main factors determining soil microbial community composition along depth gradients, and copiotrophic microbes favored surface soils while oligotrophs preferred deeper soils. These results further revealed a nutrient-effects on microbial community along soil depth layers. In addition to pH and HWC, many previous studies also suggested that salinity has considerable impacts on activity, biomass and community structure of soil microbes (Asghar et al., 2012; Yan and Marschner, 2013; Teng et al., 2014; Morrissey and Franklin, 2015). In relation to that, our SEMs indicated archaeal community structure was significantly shaped by soil salinity while salinity was significantly influenced by both soil types and depth gradients (**Figure 6B**). Such pattern implied parent material and depth determined archaeal status via the changing of salinity in these soils.

Network analysis of 16S rDNA data further suggested depth-effects on the connection among microbes. Firstly, the modularity of bacterial and archaeal communities was the highest in the top soils and decreased with soil depth (**Table 4**), indicating a higher degree of habitat heterogeneity for microbes in upper soils (Wang et al., 2015). Also, modularity could serve as an indicator of system resistance (Carpenter et al., 2012). Therefore, the highest modularity of bacterial and archaeal network in top layers was suggestive of a relatively higher system resistance to changes compared to networks in deeper soil layers. Similarly, the lowest avgK value of the network in top layers indicated a low intensive interaction within bacteria and archaea (**Table 4**). Consequently, the higher modularity and lower avgK value in top layers both suggested that the bacterial and archaeal community in the surface soils would be more resistant and less influenced by disturbance than deeper soils according to network theory (Saavedra et al., 2011). Interestingly, despite bacterial and archaeal community had similar patterns of modularity along soil depth, the higher percentages of negative links within the bacterial community in layers B, C, and D were observed, while the higher proportions of negative links within archaeal community were detected in upper soils (**Table 4**), suggesting the competition on resources within bacterial and archaeal community was fiercer in deeper and surface soils, respectively. Such a phenomenon might imply a division of positive interactions of the two kingdoms within soils and archaea might be more adapted to the extreme environments in deeper soils.

Variation of Main Bacterial and Archaeal Groups in Different Soil Types and Depth Layers Revealed by 16S rDNA Analysis

The microbial taxonomy analysis based on 16S rRNA gene suggested that *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, *Chloroflexi,* *Firmicutes,* and *Verrucomicrobia* occupied

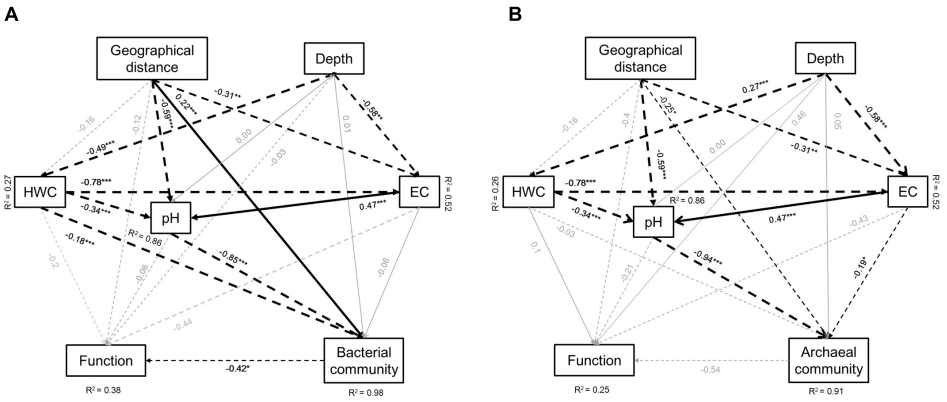


FIGURE 6 | Effects of soil type (geographic distance), depth, soil HWC, pH and salinity (EC) on bacterial community and functional **(A)** and archaeal community and functional **(B)** structure. Geographical distance includes longitude and latitude of the soil sampling sites. Solid lines denote positive effects, and broken lines denote negative effects. Thickness of the arrows denotes significance and strength of the influence. R^2 represent percentage of explanation of the models on the chosen factors. Significant level: $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Goodness-of-fit statistics are as following: **(A)** Chi-square = 0.000, degrees of freedom = 1, RMSEA = 0.000, AIC = 54, GFI = 1.000; **(B)** Chi-square = 0.000, degrees of freedom = 1, RMSEA = 0.000, AIC = 54, GFI = 1.000.

approximately 80–90% of the bacterial communities in the three soil types, and some microbial taxa showed specific distribution patterns according to soil types or profiles. For instance, the highest relative abundance of *Acidobacteria* was recorded in the acidic TY Ultisols (14–29%), while the lowest was in the alkaline BH Inceptisols (12.7–16.6%). Pearson correlation analysis further proved that the relative abundance of *Acidobacteria* negatively correlated with soil pH ($P < 0.01$, $r = -0.48$, $n = 48$). The results were consistent with the previous studies' conclusions that *Acidobacteria* are sensitive to pH change and this phylum thrives in relatively acidic soils (Will et al., 2010; Mendes et al., 2015).

Except to the bacterial groups such as *Planctomycetes*, α-*Proteobacteria,* and δ-*Proteobacteria* evenly distributed in each layer, the other groups showed consistently preference to a certain depth layer in the three soil types. Particularly, *Cyanobacteria, Verrucomicrobia,* and β-*proteobacteria*, were more abundant in all surface samples than in any other sub-soils. *Cyanobacteria* presented in the surface abundantly but hardly detected in sub-soils, which could be explained by their phototrophic metabolic characteristic. A higher relative abundance of β-*proteobacteria* was also reported in an early study in the oxic zone of paddy soils (Lüdemann et al., 2000), indicated the class prefers a relatively oxygen-rich condition. The highest relative abundance of *Verrucomicrobia* were observed in surface soils for three soil types in the present study. Similarly, *Verrucomicrobia* was found to peak at middle depth layers (20–40 cm) in upland soils and it was suggested that *Verrucomicrobia* might prefer a relative anoxic circumstance rather than oxic and extreme anoxic conditions in previous studies (Hansel et al., 2008; Eilers et al., 2012). For the paddy soils used in this study, the surface samples at two sampling points were flooded or at least water-saturated and therein represented an oxic/anoxic interface. The highest relative abundance of *Verrucomicrobia* in this layer and decreasing trend along profile depth further suggested its preference to such oxic/anoxic habitat.

By contrast, *Firmicutes*, *Chloroflexi,* and *Acidobacteria* were more abundant in sub-soils than in surfaces in all three soil types. Increasing of *Firmicutes* along soil profiles was observed in all three soil types, which is consistent with the observations in previous studies (Hansel et al., 2008; Li et al., 2014), and could be explained by their adaptation to low-nutrient environments in deeper soil layers as *Firmicutes* usually thrive in extreme conditions through spore-forming (Li et al., 2014). The decline tendency of *Chloroflexi* and *Acidobacteria* along soil depth coincided with a recent study in a colluvial soil in which the relative abundance of two phyla decreased along the profile up to 80 cm depth but fluctuated in deeper soils of 80–380 cm (Sagova-Mareckova et al., 2016). Moreover, the strong inverse relationship between carbon availability and the abundance of *Acidobacteria* were frequently observed in various soil systems and *Acidobacteria* was supposed to be "oligotrophic" (Fierer et al., 2007; Hansel et al., 2008). The clear decrease in abundance of *Acidobacteria* in carbon-poor deep profiles observed in this study further supported this hypothesis. Besides these groups showing consistent trend in three soil types, some predominant bacterial phyla such as α-, γ-*Proteobacteria,* and *Actinobacteria* did not show consistent shifts in relative abundance along the profiles of three soil types, suggesting the specific microbial community characteristics in different soil types (Eilers et al., 2012).

Taxonomy results of archaeal 16S rRNA gene reads showed that *Thaumarchaeota* was the most predominant phyla (36.7–76.7%) across each layer and three soil types in this study, while *Crenarchaeota* only accounted for 1.03–43.8% of archaeal reads, which was consistent with the observation that *Thaumarchaeota* dominated archaeal community in soil systems (Hu et al., 2013; Chronakova et al., 2015; Schneider et al., 2015; Tripathi et al., 2015). Generally, *Thaumarchaeota* was more abundant in all surface soils, and such a pattern was similar to a previous finding that *Thaumarchaeota* was the only archaeal group in aerobic top soils, in contrast to the more diverse archaeal community in deeper soil layers (Mikkonen et al., 2014). Likewise, proteins from *Thaumarchaeota* were mainly found in relatively oxic environments along a gradient of oxygen and redox in oxygen minimum zones (Hawley et al., 2014), suggesting the distribution of *Thaumarchaeota* was probably oxygen dependent. It was not surprise as the archaeal ammonia oxidizers are mainly affiliated within this phylum. Specifically, *Nitrososphaerales* within the phylum *Thaumarchaeota* was widely detected, with high relative abundance especially in all profile layers of BH and LZ soils, suggesting high abundance of ammonia oxidizers. Interestingly, *Cenarchaeaceae* and SAGMA-X within *Thaumarchaeota* showed contrasting distribution patterns across soil samples in the present study. *Cenarchaeaceae* was only found in BH soils and top soils of LZ columns (6.6–44.8%) while SAGMA-X was mainly detected in TY profiles and predominated in surface of TY soils (63.4%) (**Figure 3**). However, these two groups were rarely recorded in previous reports except that *Cenarchaeaceae* was detected in the fluid of a borehole in South Africa and the SAGMA cluster was found in a study in rivers (Herfort et al., 2009; Hernández-Torres et al., 2015). The ecological significance and potential function of these two groups deserved to be further investigation in the future.

The phylum *Crenarchaeota*, accounted for 9.4–43.8% in TY soils, while only accounted for 1.03–14.1% in BH and LZ soils. The higher relative abundance of *Crenarchaeota* in sub-soil layers was mainly determined by the presence of the MCG group, which was previously found more abundant in subsurface soils with anoxic environments (Mikkonen et al., 2014), presumably due to their anaerobic characteristic. Similarly, some subgroups under this class were recently detected in river sediments and were reported to preferentially occupy deeper layers of the sediments with reducing conditions (Lazar et al., 2015).

The phylum *Euryarchaeota* was detected in all profile depths with the proportion varying between 9.7 and 44.6% and mainly composed of *Methanobacteria* and *Methanomicrobia* in this study. The proportions in the present study were clearly higher than the proportion (∼15%) in upland soils according to previously reported (Hu et al., 2013), as the representative *Euryarchaeota*-affiliating class *Methanobacteria* and *Methanomicrobia* were commonly detected in anaerobic environments such as aquatic ecosystems, wetlands, and paddy soils and responsible for methanogenesis (Mao et al., 2015; Tong

et al., 2015). Although the two classes were both detected in all the samples, the ratio of *Methanobacteria* and *Methanomicrobia* proportion varied markedly among soils types, ranging from 22.22 in BH soils to 0.23 in TY soils. Similarly, it was observed that the relative abundance of *Methanomicrobia* increased largely in the anoxic zone in a lake while *Methanobacteria* was absent in a previous report (Hugoni et al., 2015). Conversely, *Methanobacteria* was widely identified in another study in a water-flooded oil reservoir without detecting *Methanomicrobia* (Hernández-Torres et al., 2015). All these imply that the two classes probably compete for niche and substrate and replace each other under specific soil conditions.

Microbial Functional Structure in Different Paddy Soils and Linkage with Community Structure

Analyzing microbial functional genes of key enzymes related to biogeochemistry and metabolic pathways is essential to link microbial communities to their ecological functions. Functional gene array has been proved being a more effective approach to accomplish this task compared to conventional molecular ecology techniques (He et al., 2010, 2012; Kang et al., 2013). The detected genes in this study involved in diverse functions such as carbon degradation, methane oxidation and production, nitrogen and sulfur cycling, phosphorus utilization, heavy metal and antibiotic resistance, organic remediation and other categories, suggesting versatile ecological processes occurrence in these paddy soils. Revealing the pattern of these genes would be helpful to understand and predict relative functional processes performed by these gene categories (Bai et al., 2013; Zhang et al., 2013). However, DCA and SEMs analysis did not figure out similar effects from environmental factors on functional structure as the effect on bacterial community, except that the functional structure separated between the two rice growth stages in DCA plot (**Figure 1C**), suggesting the factors relative with plant growth potentially affected the structure of soil microbial functions to a greater degree even than soil type. Although significant effect of soil properties on microbial functional structure was not detected by SEMs, microbial functional gene diversity and abundance in the LZ soils (Oxisol) was significantly lower than in the BH soils (Inceptisol) and the TY soils (Ultisol) (**Figure 4** and **Table 3**), and the functional structure in the BH and TY soils were more similar compared to the LZ samples in DCA plot (**Figure 1**). Such pattern might be attributed to the complex influences from soil type but not to a single factor as suggested before (Cao et al., 2012; Zhao et al., 2016). On the other hand, significant effects of bacterial community structure on functional structure while the weak relationship between archaeal community and functional structure were detected by SEM (**Figure 6**), possibly due to the low amount of archaeal probes in the functional gene array analysis.

To establish the linkage for microbial community composition and their function, network analysis between microbial taxa and functional genes were generated in the present study, and the analysis has further proved the known abundant microbial groups such as α- and β-*proteobacteria* and *Acidobacteria* could play essential roles in regulating ecological functions. However, the networks also suggested that the predominated microbial groups in soils did not necessary provide intense functional potentials, while less abundant microbial groups might take greater parts in soil ecological processes. For instance, the two dominant Phyla, *Firmicute* and *Actinobacteria* were detected in all soils but their numbers of connections with functions only ranged between 2 and 11 in the networks. On the contrary, some groups with lower relative abundance such as ε-*Proteobacteria* and *Parvarchaota* had more than 32 and 157 connections with functions (Supplementary Tables 3, 4).

Although both bacterial and archaeal community composition shifted greatly along soil profiles, the general functional structure only varied among soil types but not across the profile depths in the present study. The possible reason for the distinctive pattern of microbial taxonomic and functional composition along soil profiles could be that some microbial groups with less abundant take greater parts in soil ecological processes as mentioned above. Furthermore, the result was consistent with previous observation that the change of microbial community composition does not necessarily connect to a change of microbial functions, and could be explained as the functional redundancy of soil microorganisms (Comte and Giorgio, 2010). Given the possibility that most of microbes inhabiting in soils possessed similar functional genes, the fluctuation in taxonomic structure along soil profile gradients would not necessarily alter the microbial function structure. Such ability could serve as a fundamental property of soil microbes which is essential to environmental perturbation (Comte and Giorgio, 2010). The similar weak linkages between microbial taxonomic and functional community structure were previously observed in Antarctic soils (Yergeau et al., 2007, 2012; Chong et al., 2015), subtropical broadleaf forests (Ding et al., 2015), fen peatlands (Haynes et al., 2015), and in microbial stream biofilms (Dopheide et al., 2015), which further corroborated the theory.

CONCLUSION

By combining Illumina Mi-Seq sequencing with Geochip techniques, the study demonstrated manifest separation of bacterial and archaeal community and function structure among soil types, and visible but relatively slight shift of community structure along soil depth within each soil type, suggesting the overwhelming effect of soil parent material characteristics mainly via determination of soil pH, even under uniform rice cultivation management. Bacterial community showed significantly shifted along soil depth gradients mainly driven by organic carbon, while archaeal community variation along soil depth was mainly determined by salinity. Moreover, bacterial and archaeal taxa showed specific distribution patterns according to soil types or profiles, dependent on their ecophysiological properties. Especially, archaeal community composition showed contrasting patterns among the three paddy soils. Network analysis of bacterial and archaeal community indicated paddy soils harbored a higher degree of habitat heterogeneity for microbes in

upper soils, and thus endowed microbes with higher system resistance and less intensive interaction in the surface soils compare to deeper soils. The relatively weak alignment between microbial community and functional structure suggested possible functional redundancy, and implied the potential resistance and resilience of microbial communities to environmental disturbance within these paddy soils.

AUTHOR CONTRIBUTIONS

RB was responsible to most of the laboratorial works, data processing, and article writing. J-TW conducted the construction of network analysis of microbial community and functional genes, and contributed to the analysis of sequencing data and construction of structure equation models. YD provided essential ideas to the article writing, as well as assistance in constructing phylogenetic molecular ecological networks. J-ZH provided essential ideas to the experimental design and article writing. KF contributed to the construction of phylogenetic molecular ecological networks. L-MZ provided essential ideas to the experimental design, and was responsible for mesocosm setup, article writing and revising.

ACKNOWLEDGMENTS

This work was financially supported by the National Science Foundation of China (No. 41322007), Chinese Academy of Sciences (XDB15020200), MOST (2014BAD14B00), and Youth Innovation Promotion Association of Chinese Academy of Sciences. We'd like to thank Dr. Ying Gao's assistance in the analysis of GeoChip data, and Prof. Phillip Chalk for the language polishing for the manuscript. We are also grateful to Dr. Yu Dai, Dr. Chaolei Yuan, and Dr. Guiyou Zhang's help for soil collection and the analysis of soil chemical properties.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00945/full#supplementary-material

REFERENCES

Anderson, D. (1988). The effect of parent material and soil development on nutrient cycling in temperate ecosystems. *Biogeochemistry* 5, 71–97. doi: 10.1007/BF02180318

Asghar, H. N., Setia, R., and Marschner, P. (2012). Community composition and activity of microbes from saline soils and non-saline soils respond similarly to changes in salinity. *Soil Biol. Biochem.* 47, 175–178. doi: 10.1016/j.soilbio.2012.01.002

Bai, R., Xi, D., He, J.-Z., Hu, H.-W., Fang, Y.-T., and Zhang, L.-M. (2015). Activity, abundance and community structure of anammox bacteria along depth profiles in three different paddy soils. *Soil Biol. Biochem.* 91, 212–221. doi: 10.1016/j.soilbio.2015.08.040

Bai, S., Li, J., He, Z., Van Nostrand, J. D., Tian, Y., Lin, G., et al. (2013). GeoChip-based analysis of the functional gene diversity and metabolic potential of soil microbial communities of mangroves. *Appl. Microbiol. Biotechnol.* 97, 7035–7048. doi: 10.1007/s00253-012-4496-z

Bastian, M., and Heymann, S. (2009). "Gephi: an open source software for exploring and manipulating networks," in *Proceedings of the International AAAI Conference on Weblogs and Social Media*, San Jose, CA.

Cao, P., Zhang, L. M., Shen, J. P., Zheng, Y. M., Di, H. J., and He, J. Z. (2012). Distribution and diversity of archaeal communities in selected Chinese soils. *FEMS Microbiol. Ecol.* 80, 146–158. doi: 10.1111/j.1574-6941.2011.01280.x

Carpenter, S., Arrow, K., Barrett, S., Biggs, R., Brock, W., Crépin, A.-S., et al. (2012). General resilience to cope with extreme events. *Sustainability* 4, 3248–3259. doi: 10.3390/su4123248

Chong, C. W., Pearce, D. A., and Convey, P. (2015). Emerging spatial patterns in Antarctic prokaryotes. *Front. Microbiol.* 6:1058. doi: 10.3389/fmicb.2015.01058

Chronakova, A., Schloter-Hai, B., Radl, V., Endesfelder, D., Quince, C., Elhottova, D., et al. (2015). Response of archaeal and bacterial soil communities to changes associated with outdoor cattle overwintering. *PLoS ONE* 10:e0135627. doi: 10.1371/journal.pone.0135627

Comte, J., and Giorgio, P. A. D. (2010). Linking the patterns of change in composition and function in bacterioplankton successions along environmental gradients. *Ecology* 91, 1466–1476. doi: 10.1890/09-0848.1

Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics* 13:113. doi: 10.1186/1471-2105-13-113

Ding, J., Zhang, Y., Wang, M., Sun, X., Cong, J., Deng, Y., et al. (2015). Soil organic matter quantity and quality shape microbial community compositions of subtropical broadleaved forests. *Mol. Ecol.* 24, 5175–5185. doi: 10.1111/mec.13384

Dopheide, A., Lear, G., He, Z., Zhou, J., and Lewis, G. D. (2015). Functional gene composition, diversity and redundancy in microbial stream biofilm communities. *PLoS ONE* 10:e0123179. doi: 10.1371/journal.pone.0123179

Durán, J., Morse, J. L., Rodríguez, A., Campbell, J. L., Christenson, L. M., Driscoll, C. T., et al. (2017). Differential sensitivity to climate change of C and N cycling processes across soil horizons in a northern hardwood forest. *Soil Biol. Biochem.* 107, 77–84. doi: 10.1016/j.soilbio.2016.12.028

Eilers, K. G., Debenport, S., Anderson, S., and Fierer, N. (2012). Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biol. Biochem.* 50, 58–65. doi: 10.1016/j.soilbio.2012.03.011

Fierer, N., Morse, J. L., Berthrong, S. T., Bernhardt, E. S., and Jackson, R. B. (2007). Environmental controls on the landscape-scale biogeography of stream bacterial communities. *Ecology* 88, 2162–2173. doi: 10.1890/06-1746.1

Fierer, N., Schimel, J. P., and Holden, P. A. (2003). Variations in microbial community composition through two soil depth profiles. *Soil Biol. Biochem.* 35, 167–176. doi: 10.1016/S0038-0717(02)00251-1

Girvan, M. S., Bullimore, J., Pretty, J. N., Osborn, A. M., and Ball, A. S. (2003). Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Appl. Environ. Microbiol.* 69, 1800–1809. doi: 10.1128/aem.69.3.1800-1809.2003

Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., and Whiteley, A. S. (2011). The bacterial biogeography of British soils. *Environ. Microbiol.* 13, 1642–1654. doi: 10.1111/j.1462-2920.2011.02480.x

Hansel, C. M., Fendorf, S., Jardine, P. M., and Francis, C. A. (2008). Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl. Environ. Microbiol.* 74, 1620–1633. doi: 10.1128/AEM.01787-07

Hartmann, M., Lee, S., Hallam, S. J., and Mohn, W. W. (2009). Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. *Environ. Microbiol.* 11, 3045–3062. doi: 10.1111/j.1462-2920.2009.02008.x

Hawley, A. K., Brewer, H. M., Norbeck, A. D., Pasa-Tolic, L., and Hallam, S. J. (2014). Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11395–11400. doi: 10.1073/pnas.1322132111

Haynes, K. M., Preston, M. D., McLaughlin, J. W., Webster, K., and Basiliko, N. (2015). Dissimilar bacterial and fungal decomposer communities across rich to

poor fen peatlands exhibit functional redundancy. *Can. J. Soil Sci.* 95, 219–230. doi: 10.4141/cjss-2014-062

He, Z., Deng, Y., Van Nostrand, J. D., Tu, Q., Xu, M., Hemme, C. L., et al. (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J.* 4, 1167–1179. doi: 10.1038/ismej.2010.46

He, Z., Deng, Y., and Zhou, J. (2012). Development of functional gene microarrays for microbial community analysis. *Curr. Opin. Biotechnol.* 23, 49–55. doi: 10.1016/j.copbio.2011.11.001

Herfort, L., Kim, J. H., Coolen, M. J. L., Abbas, B., Schouten, S., Herndl, G. J., et al. (2009). Diversity of *Archaea* and detection of crenarchaeotal *amoA* genes in the rivers Rhine and Têt. *Aquat. Microb. Ecol.* 55, 189–201. doi: 10.3354/ame01294

Hernández-Torres, J., Castillo Villamizar, G. A., Salgar-Chaparro, S. J., Silva-Plata, B. A., Serna Daza, O. D., Martínez-Pérez, F., et al. (2015). Prokaryotic community characterization in a mesothermic and water- flooded oil reservoir in Colombia. *Geomicrobiol. J.* 33, 110–117. doi: 10.1080/01490451.2015.1039674

Hu, H. W., Wang, J. T., Li, J., Li, J. J., Ma, Y. B., Chen, D., et al. (2016). Field-based evidence for copper contamination induced changes of antibiotic resistance in agricultural soils. *Environ. Microbiol.* 18, 3896–3909. doi: 10.1111/1462-2920.13370

Hu, H.-W., Zhang, L.-M., Yuan, C.-L., and He, J.-Z. (2013). Contrasting Euryarchaeota communities between upland and paddy soils exhibited similar pH-impacted biogeographic patterns. *Soil Biol. Biochem.* 64, 18–27. doi: 10.1016/j.soilbio.2013.04.003

Hugoni, M., Domaizon, I., Taib, N., Biderre-Petit, C., Agogue, H., Galand, P. E., et al. (2015). Temporal dynamics of active Archaea in oxygen-depleted zones of two deep lakes. *Environ. Microbiol. Rep.* 7, 321–329. doi: 10.1111/1758-2229.12251

Imaya, A., Ohta, S., Tanaka, N., and Inagaki, Y. (2005). General chemical properties of brown forest soils developed from different parent materials in the submontane zone of the Kanto and Chubu Districts, Japan. *Soil Sci. Plant Nutr.* 51, 873–884. doi: 10.1111/j.1747-0765.2005.tb00122.x

Kang, S., Van Nostrand, J. D., Gough, H. L., He, Z., Hazen, T. C., Stahl, D. A., et al. (2013). Functional gene array-based analysis of microbial communities in heavy metals-contaminated lake sediments. *FEMS Microbiol. Ecol.* 86, 200–214. doi: 10.1111/1574-6941.12152

Kögel-Knabner, I., Amelung, W., Cao, Z., Fiedler, S., Frenzel, P., Jahn, R., et al. (2010). Biogeochemistry of paddy soils. *Geoderma* 157, 1–14. doi: 10.1016/j.geoderma.2010.03.009

Kuramae, E. E., Yergeau, E., Wong, L. C., Pijl, A. S., van Veen, J. A., and Kowalchuk, G. A. (2012). Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol. Ecol.* 79, 12–24. doi: 10.1111/j.1574-6941.2011.01192.x

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 75, 5111–5120. doi: 10.1128/AEM.00335-09

Lazar, C. S., Biddle, J. F., Meador, T. B., Blair, N., Hinrichs, K. U., and Teske, A. P. (2015). Environmental controls on intragroup diversity of the uncultured benthic archaea of the miscellaneous Crenarchaeotal group lineage naturally enriched in anoxic sediments of the White Oak River estuary (North Carolina, USA). *Environ. Microbiol.* 17, 2228–2238. doi: 10.1111/1462-2920.12659

Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J. M., et al. (2015). Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J.* 9, 2490–2502. doi: 10.1038/ismej.2015.59

Li, C., Yan, K., Tang, L., Jia, Z., and Li, Y. (2014). Change in deep soil microbial communities due to long-term fertilization. *Soil Biol. Biochem.* 75, 264–272. doi: 10.1016/j.soilbio.2014.04.023

Lu, S., Liu, X., Ma, Z., Liu, Q., Wu, Z., Zeng, X., et al. (2015). Vertical segregation and phylogenetic characterization of ammonia-oxidizing bacteria and archaea in the sediment of a freshwater aquaculture pond. *Front. Microbiol.* 6:1539. doi: 10.3389/fmicb.2015.01539

Lüdemann, H., Arth, I., and Liesack, W. (2000). Spatial changes in the bacterial community structure along a vertical oxygen gradient in flooded paddy soil cores. *Appl. Environ. Microbiol.* 66, 754–762. doi: 10.1128/AEM.66.2.754-762.2000

Mao, T.-T., Yin, R., and Deng, H. (2015). Effects of copper on methane emission, methanogens and methanotrophs in the rhizosphere and bulk soil of rice paddy. *Catena* 133, 233–240. doi: 10.1016/j.catena.2015.05.024

Martiny, J. B., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341

Mendes, L. W., de Lima Brossi, M. J., Kuramae, E. E., and Tsai, S. M. (2015). Land-use system shapes soil bacterial communities in Southeastern Amazon region. *Appl. Soil Ecol.* 95, 151–160. doi: 10.1016/j.apsoil.2015.06.005

Mikkonen, A., Santalahti, M., Lappi, K., Pulkkinen, A. M., Montonen, L., and Suominen, L. (2014). Bacterial and archaeal communities in long-term contaminated surface and subsurface soil evaluated through coextracted RNA and DNA. *FEMS Microbiol. Ecol.* 90, 103–114. doi: 10.1111/1574-6941.12376

Morrissey, E. M., and Franklin, R. B. (2015). Evolutionary history influences the salinity preference of bacterial taxa in wetland soils. *Front. Microbiol.* 6:1013. doi: 10.3389/fmicb.2015.01013

Nie, M., Meng, H., Li, K., Wan, J. R., Quan, Z. X., Fang, C. M., et al. (2012). Comparison of bacterial and fungal communities between natural and planted pine forests in subtropical China. *Curr. Microbiol.* 64, 34–42. doi: 10.1007/s00284-011-0031-1

Noll, M., Matthies, D., Frenzel, P., Derakshani, M., and Liesack, W. (2005). Succession of bacterial community structure and diversity in a paddy soil oxygen gradient. *Environ. Microbiol.* 7, 382–395. doi: 10.1111/j.1462-2920.2004.00700.x

Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2006). The smallest of all worlds: pollination networks. *J. Theor. Biol.* 240, 270–276. doi: 10.1016/j.jtbi.2005.09.014

Paula, F. S., Rodrigues, J. L., Zhou, J., Wu, L., Mueller, R. C., Mirza, B. S., et al. (2014). Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Mol. Ecol.* 23, 2988–2999. doi: 10.1111/mec.12786

Qin, H. L., Zhang, Z. X., Lu, J., Zhu, Y. J., Webster, R., Liu, X. L., et al. (2016). Change from paddy rice to vegetable growing changes nitrogen-cycling microbial communities and their variation with depth in the soil. *Eur. J. Soil Sci.* 67, 650–658. doi: 10.1111/ejss.12365

Reim, A., Luke, C., Krause, S., Pratscher, J., and Frenzel, P. (2012). One millimetre makes the difference: high-resolution analysis of methane-oxidizing bacteria and their specific activity at the oxic-anoxic interface in a flooded paddy soil. *ISME J.* 6, 2128–2139. doi: 10.1038/ismej.2012.57

Saavedra, S., Stouffer, D. B., Uzzi, B., and Bascompte, J. (2011). Strong contributors to network persistence are the most vulnerable to extinction. *Nature* 478, 233–235. doi: 10.1038/nature10433

Sagova-Mareckova, M., Zadorova, T., Penizek, V., Omelka, M., Tejnecky, V., Pruchova, P., et al. (2016). The structure of bacterial communities along two vertical profiles of a deep colluvial soil. *Soil Biol. Biochem.* 101, 65–73. doi: 10.1016/j.soilbio.2016.06.026

Schimel, J. P., and Schaeffer, S. M. (2012). Microbial control over carbon cycling in soil. *Front. Microbiol.* 3:348. doi: 10.3389/fmicb.2012.00348

Schneider, D., Engelhaupt, M., Allen, K., Kurniawan, S., Krashevska, V., Heinemann, M., et al. (2015). Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in Sumatra (Indonesia). *Front. Microbiol.* 6:1339. doi: 10.3389/fmicb.2015.01339

Sheng, R., Qin, H., O'Donnell, A. G., Huang, S., Wu, J., and Wei, W. (2015). Bacterial succession in paddy soils derived from different parent materials. *J. Soils Sediments* 15, 982–992. doi: 10.1007/s11368-014-1058-2

Soffer, N., Zaneveld, J., and Vega Thurber, R. (2015). Phage-bacteria network analysis and its implication for the understanding of coral disease. *Environ. Microbiol.* 17, 1203–1218. doi: 10.1111/1462-2920.12553

Steven, B., Gallegos-Graves, L. V., Belnap, J., and Kuske, C. R. (2013). Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. *FEMS Microbiol. Ecol.* 86, 101–113. doi: 10.1111/1574-6941.12143

Stone, M. M., Kan, J., and Plante, A. F. (2015). Parent material and vegetation influence bacterial community structure and nitrogen functional genes along deep tropical soil profiles at the Luquillo Critical Zone Observatory. *Soil Biol. Biochem.* 80, 273–282. doi: 10.1016/j.soilbio.2014.10.019

Su, J. Q., Ding, L. J., Xue, K., Yao, H. Y., Quensen, J., Bai, S. J., et al. (2015). Long-term balanced fertilization increases the soil microbial functional diversity in a phosphorus-limited paddy soil. *Mol. Ecol.* 24, 136–150. doi: 10.1111/mec.13010

Sun, L., Gao, J., Huang, T., Kendall, J. R., Shen, Q., and Zhang, R. (2015). Parental material and cultivation determine soil bacterial community structure and fertility. *FEMS Microbiol. Ecol.* 91, 1–10. doi: 10.1093/femsec/fiu010

Teng, Y., Su, J., Wang, J., Dai, N., Li, J., Song, L., et al. (2014). Soil microbial community response to seawater intrusion into coastal aquifer of Donghai Island, South China. *Environ. Earth Sci.* 72, 3329–3338. doi: 10.1007/s12665-014-3236-3

Thomson, B. C., Tisserant, E., Plassart, P., Uroz, S., Griffiths, R. I., Hannula, S. E., et al. (2015). Soil conditions and land use intensification effects on soil microbial communities across a range of European field sites. *Soil Biol. Biochem.* 88, 403–413. doi: 10.1016/j.soilbio.2015.06.012

Tong, H., Liu, C., Li, F., Luo, C., Chen, M., and Hu, M. (2015). The key microorganisms for anaerobic degradation of pentachlorophenol in paddy soil as revealed by stable isotope probing. *J. Hazard. Mater.* 298, 252–260. doi: 10.1016/j.jhazmat.2015.05.049

Tripathi, B. M., Kim, M., Singh, D., Lee-Cruz, L., Lai-Hoe, A., Ainuddin, A. N., et al. (2012). Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microb. Ecol.* 64, 474–484. doi: 10.1007/s00248-012-0028-8

Tripathi, B. M., Kim, M., Tateno, R., Kim, W., Wang, J., Lai-Hoe, A., et al. (2015). Soil pH and biome are both key determinants of soil archaeal community structure. *Soil Biol. Biochem.* 88, 1–8. doi: 10.1016/j.soilbio.2015.05.004

Wang, F., Liang, Y., Jiang, Y., Yang, Y., Xue, K., Xiong, J., et al. (2015). Planting increases the abundance and structure complexity of soil core functional genes relevant to carbon and nitrogen cycling. *Sci. Rep.* 5:14345. doi: 10.1038/srep14345

Wang, J., Zhang, L., Lu, Q., Raza, W., Huang, Q., and Shen, Q. (2014). Ammonia oxidizer abundance in paddy soil profile with different fertilizer regimes. *Appl. Soil Ecol.* 84, 38–44. doi: 10.1016/j.apsoil.2014.06.009

Will, C., Thurmer, A., Wollherr, A., Nacke, H., Herold, N., Schrumpf, M., et al. (2010). Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. *Appl. Environ. Microbiol.* 76, 6751–6759. doi: 10.1128/AEM.01063-10

Yan, N., and Marschner, P. (2013). Response of soil respiration and microbial biomass to changing EC in saline soils. *Soil Biol. Biochem.* 65, 322–328. doi: 10.1016/j.soilbio.2013.06.008

Yang, Y., Wu, L., Lin, Q., Yuan, M., Xu, D., Yu, H., et al. (2013). Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland. *Glob. Change Biol.* 19, 637–648. doi: 10.1111/gcb.12065

Yergeau, E., Bokhorst, S., Kang, S., Zhou, J., Greer, C. W., Aerts, R., et al. (2012). Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. *ISME J.* 6, 692–702. doi: 10.1038/ismej.2011.124

Yergeau, E., Kang, S., He, Z., Zhou, J., and Kowalchuk, G. A. (2007). Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISME J.* 1, 163–179. doi: 10.1038/ismej.2007.24

Zhang, Y., Lu, Z., Liu, S., Yang, Y., He, Z., Ren, Z., et al. (2013). Geochip-based analysis of microbial communities in alpine meadow soils in the Qinghai-Tibetan plateau. *BMC Microbiol.* 13:72. doi: 10.1186/1471-2180-13-72

Zhao, M., Sun, B., Wu, L., Gao, Q., Wang, F., Wen, C., et al. (2016). Zonal soil type determines soil microbial responses to maize cropping and fertilization. *mSystems* 1:e00075-16. doi: 10.1128/mSystems.00075-16

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bai, Wang, Deng, He, Feng and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 February 2018
doi: 10.3389/fmicb.2018.00159

Metagenomic Profiling of Soil Microbes to Mine Salt Stress Tolerance Genes

Vasim Ahmed[†], *Manoj K. Verma*[†], *Shashank Gupta, Vibha Mandhan and Nar S. Chauhan*[*]

Department of Biochemistry, Maharshi Dayanand University, Rohtak, India

Osmotolerance is one of the critical factors for successful survival and colonization of microbes in saline environments. Nonetheless, information about these osmotolerance mechanisms is still inadequate. Exploration of the saline soil microbiome for its community structure and novel genetic elements is likely to provide information on the mechanisms involved in osmoadaptation. The present study explores the saline soil microbiome for its native structure and novel genetic elements involved in osmoadaptation. 16S rRNA gene sequence analysis has indicated the dominance of halophilic/halotolerant phylotypes affiliated to *Proteobacteria*, *Actinobacteria*, *Gemmatimonadetes*, *Bacteroidetes*, *Firmicutes,* and *Acidobacteria*. A functional metagenomics approach led to the identification of osmotolerant clones SSR1, SSR4, SSR6, SSR2 harboring *BCAA_ABCtp, GSDH, STK_Pknb,* and *duf3445* genes. Furthermore, transposon mutagenesis, genetic, physiological and functional studies in close association has confirmed the role of these genes in osmotolerance. Enhancement in host osmotolerance possibly though the cytosolic accumulation of amino acids, reducing equivalents and osmolytes involving *BCAA-ABCtp*, *GSDH,* and *STKc_PknB*. Decoding of the genetic elements prevalent within these microbes can be exploited either as such for ameliorating soils or their genetically modified forms can assist crops to resist and survive in saline environment.

Keywords: metagenome, halotolerance, SSU rRNA, soil microbiome, soil ecology

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Richard Allen White III, RAW Molecular Systems (RMS) LLC, United States
Eamonn P. Culligan, Cork Institute of Technology, Ireland

***Correspondence:**
Nar S. Chauhan
nschauhan@mdurohtak.ac.in

[†]*These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *12 October 2017*
Accepted: *23 January 2018*
Published: *08 February 2018*

Citation:
Ahmed V, Verma MK, Gupta S, Mandhan V and Chauhan NS (2018) Metagenomic Profiling of Soil Microbes to Mine Salt Stress Tolerance Genes. Front. Microbiol. 9:159. doi: 10.3389/fmicb.2018.00159

INTRODUCTION

Soil is a rich and dynamic ecosystem, containing a vast number of microorganisms (van Veen et al., 1997). Geological activities like weathering of rocks, winds and poor agricultural practice are continuously increasing salt contents of the soils (Jiang et al., 2007; Canfora et al., 2014). Enhanced soil salinity modulates the microbial community structure and its physiological activity (Jiang et al., 2007; Canfora et al., 2014; Shrivastava and Kumar, 2015). The majority of microbes surviving in salt stress conditions demonstrate osmotolerance for varying duration, which may extend even to their entire lifespan (Roberts, 2005). The salt stress tolerance mechanisms are complex phenomena where pathways are coordinately linked (Culligan et al., 2012). These metabolic strengths to mitigate osmotic stress, seem to be genetically evolved through horizontal gene transfer (Koonin and Wolf, 2012; Yan et al., 2015; Gupta et al., 2017). Description of these osmotolerance mechanisms is crucial for comprehensive understanding of the biology of saline soil microbes, and exploiting them for their applications in improving soil quality and crop yields (Xiao and Roberts, 2010; Zhengbin et al., 2011; Culligan et al., 2012, 2013, 2014; Fernandes, 2014). A variety

of culture dependent studies have been carried out to decode the gene(s) involved in osmotolerance within halophilic or halotolerant microbes (Zuleta et al., 2003; Klähn et al., 2009; Naughton et al., 2009; Meena et al., 2017). These studies have deciphered the role of proteins, Na^+/H^+ pumps, compatible solutes in salt stress tolerance (Sakamoto and Murata, 2002; Roberts, 2005). However, culture independent approach provides a vast opportunity for searching salt tolerant gene (s) (Singh J. et al., 2009; Mirete et al., 2015; Kumar J. et al., 2016; Chauhan et al., 2017; Gupta et al., 2017). Only a few studies have used metagenomic approach to decode the microbial salt stress tolerance mechanisms from various environments like pond water (Kapardar et al., 2010a,b), brines and moderate-salinity rhizosphere (Mirete et al., 2015), human gut microbiome (Culligan et al., 2012, 2013, 2014). Incidentally, the number of genes/pathways identified for salt stress tolerance are far below than the number of microbes which have been identified to reside within these environments (Humbert et al., 2009; Mirete et al., 2015). Hereby, the current study was proposed to identify the genetic machinery used by microbes as survival strategies in salt stress condition using functional metagenomic approach. The current study led to the identification of a number of osmotolerant genes that could be used to develop strategies to ensure survival of microbes under saline conditions.

MATERIALS AND METHODS

Saline Soil Sample Collection and Metagenomic DNA Isolation

Saline soil samples were collected in sterile containers after carefully removing the surface layer (up to 10 cm) from Village Malab, District Nuh situated at 28.0107°N, 77.0564°E. Metagenomic DNA was extracted from 5 g of the soil sample (Supplementary Methods).

Bacterial Strains and Growth Conditions

Bacterial strains and plasmids used in the study are listed in **Table 1**. The oligonucleotides used in the study (GeNoRime, Shrimpex Biotech services Pvt. Ltd. India) are listed in Supplementary Table S1. *Escherichia coli* (DH10B) and *E. coli* (MKH13) strains were cultured in Luria-Bertani (LB) medium. Further, *E. coli* (DH10B) and *E. coli* (MKH13) strains containing *pUC19* vector were cultured in LB medium supplemented with ampicillin (100 $\mu g\ ml^{-1}$). All overnight cultures were grown in LB broth at 37°C with constant shaking at 200 rpm.

Phylogenetic Reconstruction of Saline Soil Metagenome

Saline soil metagenomic DNA was used to amplify the SSU rRNA gene (Supplementary Methods). The amplified product was used for next generation sequencing (NGS) with the aid of Roche 454 GS FLX+ platform (Morowitz et al., 2011; Gupta et al., 2017). Finally, Quantitative Insights Into Microbial Ecology (QIIME) 1.9.0 pipeline was implemented for SSU rRNA sequence data analysis (Caporaso et al., 2011). SSU rRNA gene sequence data was curated for quality, length and ambiguous bases as a quality filtering step. Each sample was pre-processed to remove sequences with length less than 200 nucleotides and more than 1,000 nucleotides and sequences with minimum average quality <25. Reads with ambiguities and barcode mismatch were discarded. Reads were assigned to operational taxonomic units (OTUs) using a closed reference OTU picking protocol using QIIME. The uclust was applied to search sequences against a subset of the Greengenes database, version 13_8 filtered at 97% sequence identity. The OTUs were classified taxonomically by using the Greengenes reference database at various taxonomic ranks (phylum, class, order, family, genus, and species).

Metagenomic Library Screening and Characterization of Salt Resistant Clones

Plasmid borne saline soil metagenomic library was prepared in *E. coli* DH10B using *pUC*19 vector (Supplementary Methods) (Chauhan et al., 2009, 2017) and manually screened for salt stress tolerant clones (Kapardar et al., 2010a,b). Salt stress resistant clones were screened by plating the soil metagenomic library (∼165,000 clones with an average insert of 1.89 Kb) on LB agar medium supplemented with ampicillin (100 $\mu g\ ml^{-1}$) and NaCl [5.8% (w/v)]. The 5.8% of NaCl (w/w) is a lethal concentration for *E. coli* DH10B cells and will allow the growth of only osmotolerant clones. RFLP analysis of salt stress tolerant clones was performed after digesting their recombinant plasmid DNA with *Eco*RI & *Hin*dIII at 37°C for 12 h. The minimum inhibitory concentration assay and growth inhibition studies were performed to analyze the salt stress tolerance property (Kapardar et al., 2010a,b). Growth inhibition assays of salt sensitive *E. coli* MKH13 clones were performed with 3% NaCl (w/v) & 3.7% KCl (w/v), while 5.8% NaCl (w/v) & 5.5 % of KCl (w/v) were used for *E. coli* DH10B clones. Graphs (created using Origin61) are presented as the average of triplicate experiments, with error bars being representative of the standard error of the mean.

Genetic and Physiological Characterization of Salt Tolerance Genes

The plasmid insert from salt resistant recombinant clones were sequenced using Sanger sequencing chemistry with primer walking approach at Eurofins Genomics India Pvt. Ltd (Bangalore, India). Sequence assembly was performed with Seq-Man sequence assembly software Lasergene package, version 5.07 (DNA Star, USA). Putative open reading frame (ORF) was predicted using an ORF finder tool at NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and checked for the database homology with Basic Local Alignment and Search Tool (BLAST) (http://www.ncbi.nlm.nih.gov/blast). Encoded protein sequences were analyzed for the presence of conserved domains (CDD) (Marchler-Bauer et al., 2014), topology prediction (HMMTOP) (Tusnády and Simon, 2001), phylogenetic analysis (MEGA7) (Kumar S. et al., 2016), and various physiological parameters (Tsirigos et al., 2015). Transposon mutagenesis of *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21* was carried out with EZ-Tn5™<Kan-2> Insertion kit (Epicenter Biotechnologies) following manufacturer's instructions. Transposon mutants of *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21* were screened for the salt

TABLE 1 | Bacterial strains and plasmids used in present study.

Strains, plasmids, and transposons	Genotype or characteristics	Source or reference
E. coli (DH10B)	F- *endA1 recA1 galE15 galK16 nupG rpsL* Δ*lacX74* Φ80*lacZ*ΔM15 *araD*139 Δ(*ara, leu*)7697 *mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ-	Lucigen corporation, Parmenter St. Middleton, USA
E. coli (MKH13)	MC4100Δ(*putPA*)101D(*proP*)2D(*proU*)	(Haardt et al., 1995)
pSSR1	*pUC19* harboring a metagenomic DNA fragment of 2939 bp	Present study
SSR1	*E. coli* (DH10B) containing *pSSR1*	Present study
pSSR4	*pUC19* harboring a metagenomic DNA fragment of 2945 bp	Present study
SSR4	*E. coli* (DH10B) containing *pSSR4*	Present study
pSSR6	*pUC19* harboring a metagenomic DNA fragment of 1456 bp	Present study
SSR6	*E. coli* (DH10B) containing *pSSR6*	Present study
pSSR21	*pUC19* harboring a metagenomic DNA fragment of 2352 bp	Present study
SSR21	*E. coli* (DH10B) containing *pSSR21*	Present study
pSSR1C1	*pUC19* recombinant plasmid harboring putative *BCAA_ABCTP* of *pSSR1* cloned at *EcoR1* and *HindIII of pUC19 MCS (Multiple cloning site)*	Present study
SSR1C1	*E. coli* (MKH13) harboring *pSSR1C1*	Present study
pSSR4C1	*pUC19* recombinant plasmid harboring putative *GSDH* of *pSSR4* cloned at *EcoR1* and *HindIII of pUC19 MCS*	Present study
SSR4C1	*E. coli* (MKH13) harboring *pSSR4C1*	Present study
pSSR21C1	*pUC19* recombinant plasmid harboring putative *duf3445* gene of *pSSR21* cloned at *EcoR1* and *HindIII of pUC19 MCS*	Present study
SSR21C1	*E. coli* (MKH13) harboring *pSSR21C1*	Present study
pUC19	Plasmid cloning vector Amp^r	Thermo Scientific
Transposon EZ $Tn5^{TM}$ <Kan-2>	$Tn5^{TM}$ Transposon Kan^r	Epicenter Biotechnologies Madison, Wisconsins, USA

stress resistant and sensitive phenotypes to identify the active osmotolerant genomic regions within the cloned DNA fragment in *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21*. Salt tolerant active loci encoding putative *BCAA-ABCtp*, *GSDH*, *STK_Pknb,* and *DUF3445* genes of *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21* were subcloned in *pUC19* vector (*E. coli* MKH13 host) using standard molecular cloning techniques. The growth studies of subclones were performed to analyze their salt stress maintenance property in the presence of salt stressors NaCl [3.0% (w/v)] and KCl [3.7% (w/v)]. All assays were performed in triplicates for calculation of standard deviation. A parametric *t*-test was used to calculate the *p*-value.

Elemental Quantification of Na^+ in Salt Tolerant Clones

Elemental Quantification of intracellular Na^+ in *E. coli* MKH13 carrying the empty vector (*pUC*19) and salt tolerant recombinant subclones (SSR1C1, SSR4C1, SSR6C1, SSR21C1) was measured with inductively coupled plasma spectroscopy-atomic emission spectroscopy (ICP-AES) analysis (Mirete et al., 2015) at SAIF, IIT Bombay, India. Results were expressed as mg of Na^+ g^{-1} dry weight of cells. A parametric *t*-test was used to calculate the *p*-value.

Data Availability

Sequence reads generate in present study has been deposited in the NCBI SRA under accession number SRS2727172.

RESULTS

Phylogenetic Reconstruction of Saline Soil Metagenome

Physico-chemical properties of saline soil showed that the pH of soil was 9.0 ± 0.025, while its electrical conductivity (EC) was 6.5 ± 0.023. Elemental analysis of soil showed the excessive presence of salts, sodium (105 ppm), potassium (155 ppm), and lithium (188 ppm), confirming its moderate saline nature. A good quality (A260/280 >1.8), high molecular weight (>23 Kb) metagenomic DNA was extracted from the saline soil sample. Saline soil metagenomic DNA was used to analyze its SSU rRNA gene sequences to decode its native microbiome structure. Clustering of SSU rRNA gene identified a total of 487 OTUs distributed across seven microbial phyla (**Figure 1**). Out of 487 OTUs, we observed 153 unique OTUs (Supplementary Table S2). The inferred phylogeny of the soil microbiome based upon 153 unique OTUs (**Figure 1**) was comparable to the taxonomic classifications against with greengenes database, with most of the diversity of the microbiome being attributed to phyla *Proteobacteria.* The phylogeny was visualized by iTOL (Letunic and Bork, 2016). The dominant microbial phyla were *Proteobacteria*, *Actinobacteria*, *Gemmatimonadetes*, B*acteroidetes*, *Firmicutes,* and *Acidobacteria* (**Figure 1**). Among these phyla, the majority of sequences were affiliated to Proteobacteria (43.7%) having a representation of *Alphaproteobacteria* (38.9%), *Betaproteobacteria* (7%),

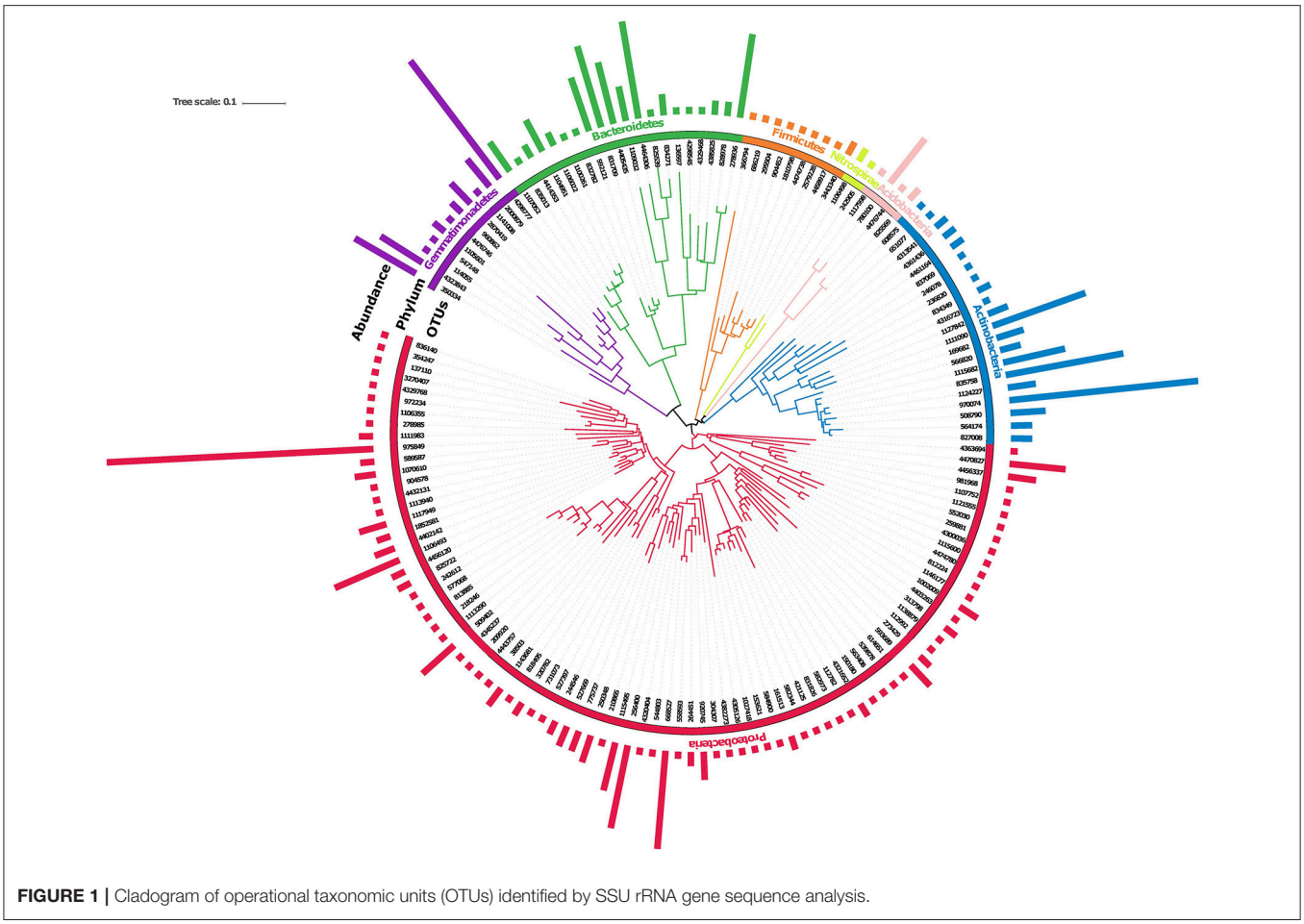


FIGURE 1 | Cladogram of operational taxonomic units (OTUs) identified by SSU rRNA gene sequence analysis.

Deltaproteobacteria (10.7%), and *Gammaproteobacteria* (43.2%); followed by Actinobacteria (21.8%) showing presence of *Acidimicrobiia* (72.47%) and *Nitriliruptoria* (18.8%); Bacteroidetes (18.1%) having a proportionate representation of *Rhodothermi* (51.9%), *Flavobacteriia* (25.96%), *Cytophagia* (19.3%), and Gemmatimonadetes (11.3%) with a percentage representation of *Gemm-2* (60.17%), *Gemm-4* (12.3%), *Gemm-1* (3.5%). Simultaneously, a minor fraction of sequences was affiliated to *Acidobacteria* (2.5%), *Firmicutes* (2.1%), and *Nitrospirae* (0.6%) microbial groups within saline soil microbiome. The taxonomic classification of saline soil microbiome confirms that a majority of microbial taxa belongs to phylum *Proteobacteria* (**Figure 1**) (Supplementary Table S2).

Screening of Salt Stress Resistant Clones from Saline Soil Microbiome

A saline soil metagenomic library was constructed with a total representation of 312 MB of cloned soil microbiome DNA. Primary screening of a saline soil metagenomic library at 5.8% NaCl (w/v) led to the identification of 24 salt stress tolerant clones. However, RFLP analysis indicated the presence of only four unique recombinant plasmids, labeled as *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21*. Minimum inhibitory concentration analysis showed almost two fold higher salt stress tolerance of SSR1, SSR4, SSR6, and SSR21 clones in comparison to the control *E. coli* (DH10B) (Supplementary Figure S1). The SSR1, SSR4, SSR6, and SSR21 also showed a statistically significant ($P = 0.0009$, $P = 0.0003$, $P = 0.0014$, $P = 0.004$) growth advantage in the presence of NaCl [4.0% (w/v)] (**Figure 2A**) and KCl [5.5% (w/v)] ($P = 0.0045$, $P = 0.0008$, $P = 0.0486$, $P = 0.0022$) as compared to *E. coli* (DH10B) strain carrying empty plasmid vector (*pUC*19) (**Figure 2B**), whereas no significant growth difference was observed between SSR1, SSR4, SSR6, SSR21, and native host *E. coli* (DH10B) carrying *pUC*19 in the presence of LB broth only (**Figure 2C**). Simultaneously *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21* successfully complemented salt stress tolerance property within salt sensitive *E. coli* (MKH13) strain and showed a statistically significant growth advantage in the presence of NaCl [3.0% (w/v)] ($P = 0.0007$, $P = 0.0002$, $P = 0.0003$, $P = 0.0001$) (**Figure 3A**) and KCl [3.7% (w/v)] ($P = 0.0003$, $P = 0.0008$, $P = 0.0003$, $P = 0.0023$) (**Figure 3B**), as compared to *E. coli* (MKH13) strain carrying empty plasmid vector (*pUC*19). At the same time, no significant growth difference was observed between *E. coli* (MKH13) strain harboring *pSSR1*, *pSSR4*, *pSSR6*, *pSSR21* as compared to *E. coli* (MKH13) strain carrying empty plasmid vector in the presence of LB broth only (**Figure 3C**).

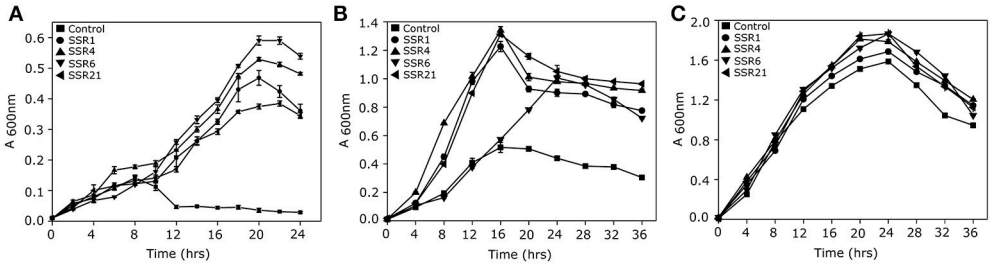


FIGURE 2 | Characterization of salt stress tolerant clones for osmotolerance property. Growth of *E. coli* (DH10B) metagenomic clones SSR1 (•), SSR4 (▲), SSR6 (▼),SSR21(◀), and *E. coli* (DH10B) host strain carrying empty plasmid vector (■) in **(A)** LB broth supplemented with 4.0% NaCl (w/v), **(B)** LB broth supplemented with 5.5% KCl and **(C)** LB broth.

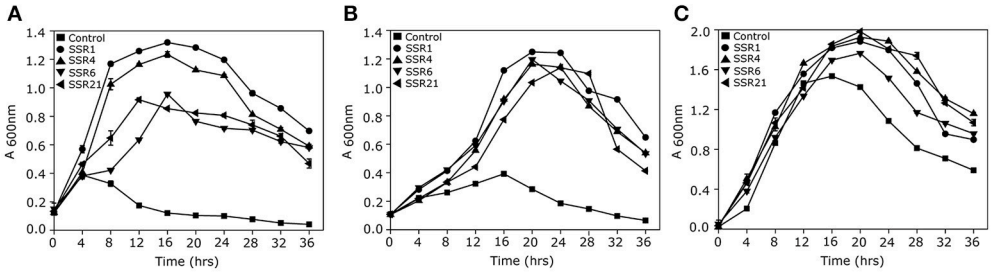


FIGURE 3 | Complementation of osmotolerance property of salt stress tolerant clones. Growth of *E. coli* (MKH13) metagenomic clones harboring *pSSR1* (•), *pSSR4*(▲), *pSSR6*(▼), *pSSR21*(◀), and *E. coli* (MKH13) host strain carrying empty plasmid vector (■) in **(A)** LB broth supplemented with 3.0% NaCl(w/v), **(B)** LB broth supplemented with 3.7% KCl,**(C)** LB broth.

Genetic and Physiological Characterization of Salt Stress Tolerant Clones

Salt Tolerant Clone SSR1

Sequence assembly of *pSSR1* resulted into a contig of 2,938 bp with a 66.87% G+C content. Cloned insert shared 76% homology with a halophilic proteobacterial lineage *Haliangium ochraceum,* indicating its plausible affiliation within proteobacterial clade. The gene prediction analysis indicated the presence of three complete and one truncated *ORF,* encoding proteins of 93, 114, 383, and 149aa respectively. Translated nucleotide sequences of these ORFs were subjected to BLASTP (maximum *e*-value cutoff of 1e-34) analysis to identify the homologous sequence in the database (**Table 2**). Transposon mutagenesis analysis confirmed functionally active locus for osmotolerance property encompassing *ORF3* (positioned between 927 and 2,078 bp) (**Figure 4**). *ORF3* encoded transmembrane protein shared homology with a transmembrane ABC transporter ATP-binding protein of *Betaproteobacteria bacterium SG8_39* (74%) and branched-chain amino acid ABC transporter substrate-binding protein of *Oceanibacterium hippocampi* (71%). Pfam database search identified the presence of a periplasmic ligand-binding domain of the ABC (ATPase Binding Cassette) type active transport systems, known to be involved in the transport of three branched chain aliphatic amino acids (leucine, isoleucine and valine) (Davidson et al., 2008). STRING analysis also predicted ORF3 as part of an interactive periplasmic binding protein dependent transport system. Further, NsitePred web server identified strong nucleotide binding sites (ATP Binding site at Gly14 and ADP binding site at Gly12) within *ORF3* encoded protein. This nucleotide binding site could be NBD, a common feature for ATP binding proteins, as predicted by its functional assignment. In consideration of physiological role and all structural features of *ORF3* encoded protein, it is a type of ABC transporter ATP-binding protein involved in salt stress maintenance possibly through energy dependent interaction with ABC membrane transporters involved in the exchange of the solutes across membrane, thus labeled as putative branched chain amino acid (BCAA) ABC transporter gene (*BCAA_ABCTP*). The putative *BCAA_ABCTP* gene was subcloned (*pSSR1C1*) to confirm its osmotolerance property. Time dependent growth curve analysis of SSR1C1 harboring *BCAA_ABCTP* showed a significant growth advantage in the presence of NaCl [3.0% (w/v)] ($P = 0.0006$) (**Figure 5A**) and KCl [3.7 % (w/v)] ($P = 0.0005$) (**Figure 5B**) as compared to salt sensitive *E. coli* mutant MKH13 carrying only the empty vector (*pUC*19), while no significant difference was observed on LB only (**Figure 5C**). The intracellular elemental analysis in the presence of ionic stressor NaCl [3.0% (w/v)] showed that SSR1C1 has effectively reduced intracellular sodium ion concentration ($P = 0.0134$) in comparison to *E. coli* mutant MKH13 (**Figure 5D**). A reduced intracellular sodium concentration within SSR1C1 could be seen as a result of its transporter property, as predicated through

TABLE 2 | Open reading frames identified in recombinant plasmids of osmotolerant clones.

	ORF details (ORF number, location, frame, product size)	Database homolog (accession no.)	Organism	Identity (%)	Coverage (%)	*e*-value	Conserved domain
pSSR1 (2938bp 66.87%)	ORF1, 13–293* +1, 93 aa	ABC transporter ATP-binding protein (LivF) (WP_056712341)	*Bosea* sp. *Leaf344*	81	94	5e-51	ATP-binding cassette transporter nucleotide-binding domain (PRK13786)
	ORF2, 305–649, +2, 114aa	Hypothetical Membrane protein (WP_037447688)	*Skermanella stibiiresistens*	58	95	2e-43	–
	ORF3, 927–2078, +2, 383aa	ABC transporter (KPK07993)	*Betaproteobacteria bacterium* SG8_39	74	93	0.0	Type I periplasmic ligand-binding domain of ABC (cd06342)
	ORF4, 2617–2168, −1, 149aa	Transcriptional regulator (OJX81021)	*Magnetospirillum* sp. 64–120	71	91	1e-76	ROS/MUCR transcriptional regulator protein (pfam05443)
pSSR4 (2945bp 63.63%)	ORF1, 21–644, +3, 207aa	Hypothetical protein AMJ56_20605 (KPK02757)	*Anaerolineae* bacterium SG8_19	64	99	1e-91	Glucose/Sorbosone dehydrogenase (GSDH) (pfam07995)
		Glucose/sorbosone dehydrogenase-like protein (WP_011340324)	*Pelobacter carbinolicus*	54	98	6e-69	
	ORF2, 2161–638, −2, 507aa	Hypothetical protein (CUR52060)	Uncultured Crenarchaeote	36	47	7e-47	mcbC-like_oxidoreductase (cd02142)
	ORF3, 2312 -2945, +2, 210aa	DUF262 domain-containing protein (WP_081445540)	*Actinomyces odontolyticus*	26	30	0.010	DUF4131 (pfam13567)
pSSR6 (1456bp 65.659%)	ORF1, (108–1145), +3, 345aa	Serine/threonine-protein kinase (WP_082990042)	*Woeseia oceani*	81	100	0.0	Catalytic domain of bacterial Serine/Threonine kinases, PknB and similar proteins (cd14014)
	ORF2, (1121–1456*), +2, 111aa	Serine/threonine phosphatase (WP_068616118)	*Woeseia oceani*	90	85	2e-59	PP2Cc Superfamily (cd00143)
pSSR21 (2352bp 69.98%)	ORF1, 913–2*, −3, 309aa	No Similarity	–	–	–	–	–
	ORF2, 912–1925, +3, 337aa	hypothetical protein A3F84_26310 (OGG46007)	*Candidatus Handelsmanbacteria* RIFCSPLOWO2_12_FULL_64_10	51	97	7e-113	DUF3445 (pfam11927)
	ORF3, 2097–2351*, −2, 84aa	Pilus assembly protein PilF (OFW28714)	*Acidobacteria* bacterium RIFCSPLOWO2_12_FULL_60_22	53	98	4e-20	TPR_10 (pfam13374)

**Indicates truncated ORF.*

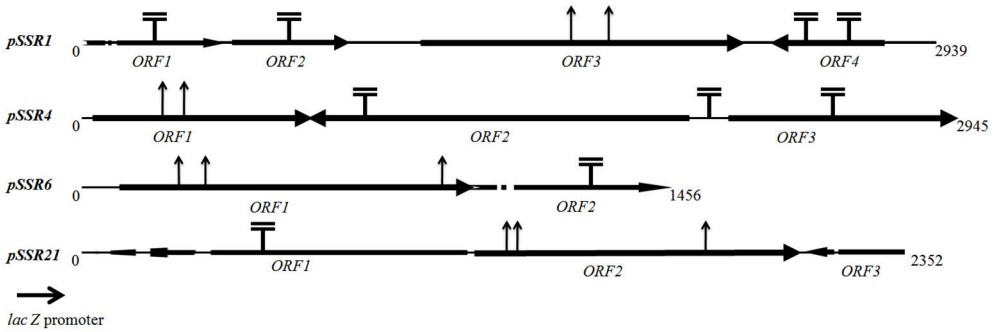


FIGURE 4 | Transposon insertion map of *pSSR1*, *pSSR4*, *pSSR6,* and *pSSR21.* T̄ indicates a transposon insertion site identified within transposon positive mutants (no effect on plasmid derived osmotolerance property) while ↑ indicate transposon insertion site identified within transposon negative mutants (loss of plasmid derived osmotolerance property).

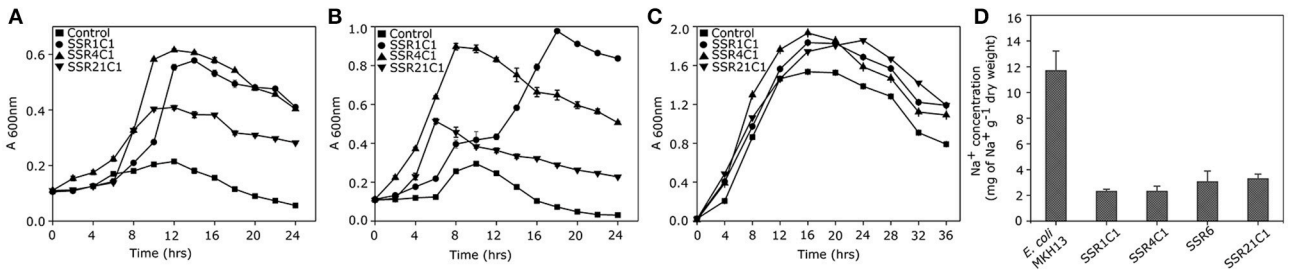


FIGURE 5 | Physiological Characterization of *BCAA_ABCTP*, *GSDH, STK_Pknb,* and *duf3445* for osmotolerance. Growth curve analysis of osmotolerant phenotype SSR1C1 (•), SSR4C1 (▲), SSR21C1 (▼), and *E. coli* (MKH13) host strain carrying empty plasmid vector (■) in **(A)** LB broth supplemented with 3% NaCl (w/v), **(B)** LB broth supplemented with 3.7 % KCl (w/v), and **(C)** LB only. Intracellular Na^+ estimation in *E. coli* (MKH13) strain harboring *pSSR1C1*, *pSSR4C1*, *pSSR6,* and *pSSR21C1* and *E. coli* (MKH13) host strain carrying empty plasmid vector **(D)**.

genetic characterization. The growth pattern of SSR1C1 was found similar to the native SSR1 that confirms that salt tolerance of salt tolerant clone SSR1 was due to SSR1C1 cloned insert encoding a branched chain amino acid (BCAA) ABC transporter protein.

Salt Tolerant Clone SSR4

The *pSSR4* harbors a G+C rich (G+C% = 63.63) insert of 2,945 bp. The cloned sequence did not share any homology at the nucleotide level in existing database sequences. A total of three *ORF*s were predicted within the cloned insert, encoding proteins of 207, 507, and 210 amino acids respectively (**Table 2**). Transposon mutagenesis analysis identified the functionally active locus within the sequence region (235 bp), encompassing *ORF1* (**Figure 4**). *ORF1* encodes a cytosolic protein homologous to hypothetical protein of *Anaerolineae* bacterium SG8_19 and Glucose/sorbosone dehydrogenase-like protein of *Pelobacter carbinolicus*. A Pfam analysis of *ORF1* encoded protein indicates it as a glucose/sorbosone dehydrogenase protein having conserved domains for the protein family GSDH. It indicates that the *ORF1* possibly encodes a glucose/sorbosone dehydrogenase involved in salt stress tolerance, possibly through the cytosolic accumulation of reducing equivalents (NADPH and GSH). Predicted G*SDH* gene was amplified from *pSSR4* and subcloned to validate its osmotolerance property. Time dependent growth curve assay of SSR4C1 showed a significant growth advantage NaCl [3.0% (w/v)] ($P = 0.0004$) (**Figure 5A**) and KCl [3.7 % (w/v)] ($P = 0.0006$) (**Figure 5B**) as compared to salt sensitive *E. coli* mutant MKH13 carrying the empty vector (*pUC*19), while no significant difference has been observed on LB only (**Figure 5C**). The growth pattern of SSR4C1 was similar to SSR4. It also confirmed that the salt tolerance property of salt tolerant clone *pSSR4* was possibly due to *GSDH* gene encoding glucose/sorbosone dehydrogenase protein.

Salt Tolerant Clone SSR6

Sequence assembly of *pSSR6* generated a G+C rich (65.66%) contig of 1,456 bp. The blastn analysis identified its low similarity to *Betaproteobacteria* GR16-43 genome sequences, indicating its affiliation from proteobacterial clade. The cloned sequence encodes only one complete *ORF* (*ORF1*), encoding a cytosolic protein of 345 amino acids with a G+C content of 65.79% (**Table 2**). Transposon mutagenesis analysis also confirmed the functionally active locus within *ORF2* (**Figure 4**). Homologs of translated *ORF2*, corresponds to putative serine/threonine protein kinase of *Woeseia oceani* and *Mycobacterium smegmatis* str. MC2 155. A pfam database search of *ORF2* encoded protein indicates it as a putative serine/threonine protein kinase having

conserved domains for the protein family STKc_PknB, i.e., the catalytic domain of bacterial Serine/Threonine kinases, PknB family. Ser/Thr protein kinase homologs were found to be involved in osmosensory signaling in microbes (Hatzios et al., 2013). These identified proteins were important for the survival and in stress responses (Donat et al., 2009). The intracellular elemental analysis in presence of the ionic stressor NaCl [3.0% (w/v)] showed that SSR6 effectively reduced intracellular sodium ion concentration ($P = 0.0200$) in comparison to *E. coli* MKH13 (**Figure 5D**). A reduced intracellular sodium concentration within SSR6C1 could be due to enhanced ion transporter activity under the influence of signals generated by putative STKc_PknB of *pSSR6* under the osmotic stress.

Salt Tolerant Clone SSR21

The *pSSR21* was found to have a G+C rich (69.98%) insert of 2352 bp. The blastn analysis identified its homology with an *Actinomycetes* strain *Allokutzneria albata*. Gene prediction has indicated the presence of three *ORFs* in pSSR21, encoding proteins of 309, 337, and 84 amino acids respectively. Among three identified *ORF*s, only *ORF*2 was complete, while other two ORFs (ORF1 and 3) were truncated. Transposon mutagenesis analysis has identified the functionally active locus within *ORF2* sequence region (1,220 and 1,330 bp). The database homologs of *ORF*2 corresponds to hypothetical protein A3F84_26310 of *Candidatus Handelsmanbacteria* bacterium (**Table 2**). The pfam analysis identified the presence of conserved domains in the protein family DUF3445, i.e., protein of unknown function (DUF3445). The G+C content of *ORF2* was found to be 69.96% and the predicted functionally active region, *ORF2* was subcloned. Time dependent growth curve assay of SSR21C1 showed a significant growth advantage NaCl [3.0% (w/v)] ($P = 0.0005$) (**Figure 5A**) and KCl [3.7% (w/v)] ($P = 0.0045$) (**Figure 5B**) as compared to salt sensitive *E. coli* mutant MKH13 carrying only the empty vector (*pUC*19), while no significant difference has been observed on LB only (**Figure 5C**). Elemental Quantification of intracellular Na^+ in *E. coli* MKH13 carrying the empty vector and salt tolerant recombinant subclones SSR21C1 clearly showed that the cloned gene insert (*duf3445*) within SSR21C1 has significantly reduced the concentration of intracellular Na^+ ion ($P = 0.0174$) (**Figure 5D**).

DISCUSSION

Metagenomics has the potential to advance our knowledge by studying the genetic components of uncultured microbes (Singh A. H. et al., 2009; Mirete et al., 2015; Chauhan et al., 2017; Yadav et al., 2017). Looking at the perspectives of metagenomics, it was used to explore soil microbiome for its composition and genetic/physiological mechanisms allowing successful adaptation of microbes in saline environments. Simultaneously, these genes could be utilized as potential candidate to develop ever demanding drought resistant transgenic crops (Zhengbin et al., 2011) or osmotolerant microbes for food processing (Fernandes, 2014) and waste water treatment applications (Xiao and Roberts, 2010). Metagenomic analysis based on SSU rRNA gene has identified dominance of *Proteobacteria*, *Actinobacteria*, *Bacteroidetes,* and *Gemmatimonadetes* in saline soil microbiome. These results are in parallel with the outcome of previous studies defining microbial community composition of saline soil environment (Zhang et al., 2003; Ma and Gong, 2013; Canfora et al., 2014; Kadam and Chuan, 2016). Canfora et al. has reported a correlation in abundance of a microbial group with respect to soil salinity. They had indicated a relative abundance of *Proteobacteria* with *Bacteroidetes* were positively and *Acidobacteria* was negatively correlated with salinity (Canfora et al., 2014). Similarly *Actinobacteria* was also been reported as another dominant microbial phylum in saline ecosystems (Kadam and Chuan, 2016). *Gemmatimonadetes* is another well-known hypersaline microbial phylum associated with biogeochemical transformations (Zhang et al., 2003). The existence of halophilic subcomponents within the identified microbial phylum, possibly making them capable to proliferate in a saline environment. These studies explain the abundance of *Proteobacteria*, *Actinobacteria*, *Gemmatimonadetes,* and *Bacteroidetes* in the studied ecosystem. Evolution of novel genetic features is a key to their successful survival in a changing environment (Gupta et al., 2017; Kumar Mondal et al., 2017). These saline microorganisms could have evolved salt stress tolerant genetic machinery to adapt and survive under salt induced osmotic stresses (Meena et al., 2017). A number of genetic elements were decoded for osmotolerance property from cultured and uncultured microbial representatives (Kapardar et al., 2010a,b; Culligan et al., 2012, 2013, 2014; Kim and Yu, 2012; Mirete et al., 2015). However, there is a disparity in a number of reported salt tolerance genes with a number of microorganisms from an ecosystem (Humbert et al., 2009; Mirete et al., 2015). This divergence could be allied with a number of factors like unculturability (Kim and Yu, 2012), lack of good quality genomic/metagenomic DNA (Kumar J. et al., 2016) and issues with foreign gene expression (Prakash and Taylor, 2012). An additional effort was made in the current study to decode osmotolerance genes prevalent in these saline soil microorganisms using functional metagenomics. It leads to the identification of four unique osmotolerant clones harboring DNA insert showing affiliation within halophile genomes. Genetic and physiological analysis has identified genes encoding putative proteins like membrane bound branched chain amino acid (BCAA) ABC transporter protein (BCAA_ABCTP), Glucose/sorbosone dehydrogenase, cytosolic STKc_PknB and DUF protein are responsible for osmotolerance property within salt stress tolerant clones. Among these, the role of the branched chain amino acid (BCAA) ABC transporter protein and glucose/sorbosone dehydrogenase in osmotolerance are well documented (Takami et al., 2002; Brosnan and Brosnan, 2006), while the scanty information is available about role for STKc_PknB and DUF3445 protein in stress maintenance (Hatzios et al., 2013).

ABC branched chain amino acid (BCAA) transporters are widely distributed in various marine microbes like *Oceanibacillus iheyensis* (Takami et al., 2002), *Salinispora, Bacillus,* and *Roseobacter* strains (Penn and Jensen, 2012). BCAA_ABCTP proteins are involved in the transport of branched chain aliphatic amino acids such as leucine, isoleucine and valine at

high salt concentration. In the presence of 2-oxoglutarate and pyridoxal-5-phosphate, these branched chain amino acids are further converted to L-glutamate by branched chain amino acid transferase (Hutson, 2001). Accumulated glutamate acts as an osmoprotectant upon hyper-osmotic shock and activates sets of genes that allow the host to achieve long-term adaptation to high osmolarity (Gralla and Vargas, 2006). They also account for a significant proportion of the genes observed in the marine metagenome. Branched chain amino acid transporters are probably an important marine adaptation because accumulated glutamate may function as a counter ion for K^+, which balances the electrical state of the cytoplasm (Penn and Jensen, 2012). Previous studies have reported a regulatory relationship between K^+ and glutamate accumulation in response to osmotic stress in enteric bacteria and haloalkaliphilic archaea *Natronococcus occultus* (Kokoeva et al., 2002). Similarly, *T. consotensis,* a halotolerant bacterium accumulates glutamate to maintain electrical equilibrium within the cell in response to high salt concentrations (Rubiano-Labrador et al., 2015). This background information explains the possible physiological role of the BCAA_ABCTP gene of *pSSR1* to increase host osmotolerance.

Glucose/Sorbosone dehydrogenase (G*SDH*) is responsible for the production of NADPH through oxidative cleavage of glucose (Oubrie et al., 1999). Under salt stress condition, NADPH acts as reducing potential for output of reduced glutathione (GSH) and involved in activity of membrane bound NADPH oxidase, which results in accumulation of hydrogen peroxide (H_2O_2) (Wang et al., 2008). H_2O_2 acts as a signal in regulating G6PDH activity and expression of this enzyme in the glutathione cycle through which the ability of GSH regeneration was increased under salt stress (Wang et al., 2008). Thus, G6PDH plays a critical role in maintaining cellular GSH levels under long-term salt stress conditions (Wang et al., 2008). It indicates that *GSDH* of *pSSR4* is involved in salt stress tolerance possibly through the cytosolic accumulation of reducing equivalents (NADPH and GSH).

While in case of *pSSR6* only one *ORF* was identified, sharing homology with Serine/Threonine kinases. This encoded protein also possesses two conserved domains for the protein family STKc_PknB, i.e. the catalytic domain of bacterial Serine/Threonine kinases, PknB and similar proteins; STKs and TOMM_kin_cyc, i.e., TOMM system kinase/cyclase fusion protein. STKs are well known for activating genetic locus concerned with osmosensing (Hatzios et al., 2013) and inducing topological changes such as DNA supercoiling (Gupta et al., 2014). These osmosensing signal and DNA supercoiling could initiate uptake of the osmolyte glycine betaine, proline (Csonka, 1989) or initiates the expression of genetic elements required to cope with osmotic stress (Higgins et al., 1988). This could be the possible mechanism by which putative STK_PknB of *pSSR6* might be extending osmotolerance to the host *E. coli*.

The *pSSR21 ORF2* encodes a protein sharing homology with a hypothetical protein of *Candidatus Handelsmanbacteria* possessing a conserved domain for DUF3445 superfamily, i.e., an uncharacterized protein family having conserved RLP sequence motif (Bateman et al., 2010). However, its physiological characterization and intracellular ion concentration analysis indicates that it extends host osmotolerance property by maintaining low intracellular ion concentration even in the presence of an ionic stressor. However, a detailed mechanism still needs to be elucidated.

CONCLUSION

In this study, the functional metagenomic approach was used to decipher salt stress tolerant genes in the saline soil microbiome. Identification of salt tolerant genes *BCAA_ABCtp*, *GSDH STK_Pknb,* and *duf3445* has enriched our understanding about the survivability and adaptability of microbes in the highly saline soil ecosystem. These salt tolerant genes can be used for crop improvement and for producing bioactive molecules under high salt conditions, which reduces the chances of contamination by other microbes.

AUTHOR CONTRIBUTIONS

NC, MV: Designed the project; MV, VA: Performed experiments and NGS sequencing; SG, VM, and VA: Performed data analyses; MV, VM, and NC: Wrote the manuscript. All authors have read and approve the manuscript.

ACKNOWLEDGMENTS

Authors would like to thank Council of Scientific and Industrial Research (CSIR) and UGC Grant Commission for fellowships under the scheme 60(0099)/11/EMRII & F. 41-1256/2012 (SR). We are thankful to Dr. Tamara Hoffmann, Philip Universitat, Marburg for providing *E. coli* (MKH13) strain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00159/full#supplementary-material

REFERENCES

Bateman, A., Coggill, P., and Finn, R. D. (2010). DUFs: families in search of function. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 66, 1148–1152. doi: 10.1107/S1744309110001685

Brosnan, J. T., and Brosnan, M. E. (2006). Branched-chain amino acids: enzyme and substrate regulation. *J. Nutr.* 136(Suppl. 1), 207S–211S. doi: 10.1093/jn/136.1.207S

Canfora, L., Bacci, G., Pinzari, F., Lo Papa, G., Dazzi, C., and Benedetti, A. (2014). Salinity and bacterial diversity: to what extent does the concentration of salt affect the bacterial community in a saline soil? *PLoS ONE* 9:e106662. doi: 10.1371/journal.pone.0106662

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Chauhan, N. S., Nain, S., and Sharma, R. (2017). Identification of arsenic resistance genes from marine sediment metagenome. *Indian J. Microbiol.* 57, 299–306. doi: 10.1007/s12088-017-0658-0

Chauhan, N. S., Ranjan, R., Purohit, H. J., Kalia, V. C., and Sharma, R. (2009). Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library: Arsenic resistance genes from sludge metagenome. *FEMS Microbiol. Ecol.* 67, 130–139. doi: 10.1111/j.1574-6941.2008.00613.x

Csonka, L. N. (1989). Physiological and genetic responses of bacteria to osmotic stress. *Microbiol. Rev.* 53, 121–147.

Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2012). Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. *ISME J.* 6, 1916–1925. doi: 10.1038/ismej.2012.38

Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2013). Functional environmental screening of a metagenomic library identifies stlA; a unique salt tolerance locus from the human gut microbiome. *PLoS ONE* 8:e82985. doi: 10.1371/journal.pone.0082985

Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2014). Metagenomic identification of a novel salt tolerance gene from the human gut microbiome which encodes a membrane protein with homology to a brp/blh-family β-carotene 15,15'-monooxygenase. *PLoS ONE* 9:e103318. doi: 10.1371/journal.pone.0103318

Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* 72, 317–364. doi: 10.1128/MMBR.00031-07

Donat, S., Streker, K., Schirmeister, T., Rakette, S., Stehle, T., Liebeke, M., et al. (2009). Transcriptome and functional analysis of the eukaryotic-type serine/threonine kinase PknB in *Staphylococcus aureus*. *J. Bacteriol.* 191, 4056–4069. doi: 10.1128/JB.00117-09

Fernandes, P. (2014). Marine enzymes and food industry: insight on existing and potential interactions. *Front. Mar. Sci.* 1:46. doi: 10.3389/fmars.2014.00046

Gralla, J. D., and Vargas, D. R. (2006). Potassium glutamate as a transcriptional inhibitor during bacterial osmoregulation. *EMBO J.* 25, 1515–1521. doi: 10.1038/sj.emboj.7601041

Gupta, M., Sajid, A., Sharma, K., Ghosh, S., Arora, G., Singh, R., et al. (2014). HupB, a nucleoid-associated protein of *Mycobacterium tuberculosis*, is modified by serine/threonine protein kinases *in vivo*. *J. Bacteriol.* 196, 2646–2657. doi: 10.1128/JB.01625-14

Gupta, S., Kumar, M., Kumar, J., Ahmad, V., Pandey, R., and Chauhan, N. S. (2017). Systemic analysis of soil microbiome deciphers anthropogenic influence on soil ecology and ecosystem functioning. *Int. J. Environ. Sci. Technol.* 14, 2229–2238. doi: 10.1007/s13762-017-1301-7

Haardt, M., Kempf, B., Faatz, E., and Bremer, E. (1995). The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of *Escherichia coli* K-12. *Mol. Gen. Genet. MGG* 246, 783–786. doi: 10.1007/BF00290728

Hatzios, S. K., Baer, C. E., Rustad, T. R., Siegrist, M. S., Pang, J. M., Ortega, C., et al. (2013). Osmosensory signaling in *Mycobacterium tuberculosis* mediated by a eukaryotic-like Ser/Thr protein kinase. *Proc. Natl. Acad. Sci. U.S.A.* 110, E5069–E5077. doi: 10.1073/pnas.1321205110

Higgins, C. F., Dorman, C. J., Stirling, D. A., Waddell, L., Booth, I. R., May, G., et al. (1988). A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* 52, 569–584.

Humbert, J.-F., Dorigo, U., Cecchi, P., Le Berre, B., Debroas, D., and Bouvy, M. (2009). Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. *Environ. Microbiol.* 11, 2339–2350. doi: 10.1111/j.1462-2920.2009.01960.x

Hutson, S. (2001). Structure and function of branched chain aminotransferases. *Prog. Nucleic Acid Res. Mol. Biol.* 70, 175–206. doi: 10.1016/S0079-6603(01)70017-7

Jiang, H., Dong, H., Yu, B., Liu, X., Li, Y., Ji, S., et al. (2007). Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. *Environ. Microbiol.* 9, 2603–2621. doi: 10.1111/j.1462-2920.2007.01377.x

Kadam, P. D., and Chuan, H. H. (2016). Erratum to: rectocutaneous fistula with transmigration of the suture: a rare delayed complication of vault fixation with the sacrospinous ligament. *Int. Urogynecol. J.* 27:505. doi: 10.1007/s00192-016-2952-5

Kapardar, R. K., Ranjan, R., Grover, A., Puri, M., and Sharma, R. (2010a). Identification and characterization of genes conferring salt tolerance to *Escherichia coli* from pond water metagenome. *Bioresour. Technol.* 101, 3917–3924. doi: 10.1016/j.biortech.2010.01.017

Kapardar, R. K., Ranjan, R., Puri, M., and Sharma, R. (2010b). Sequence analysis of a salt tolerant metagenomic clone. *Indian J. Microbiol.* 50, 212–215. doi: 10.1007/s12088-010-0041-x

Kim, M., and Yu, Z. (2012). Quantitative comparisons of select cultured and uncultured microbial populations in the rumen of cattle fed different diets. *J. Anim. Sci. Biotechnol.* 3:28. doi: 10.1186/2049-1891-3-28

Klähn, S., Marquardt, D. M., Rollwitz, I., and Hagemann, M. (2009). Expression of the ggpPS gene for glucosylglycerol biosynthesis from *Azotobacter vinelandii* improves the salt tolerance of *Arabidopsis thaliana*. *J. Exp. Bot.* 60, 1679–1689. doi: 10.1093/jxb/erp030

Kokoeva, M. V., Storch, K. F., Klein, C., and Oesterhelt, D. (2002). A novel mode of sensory transduction in archaea: binding protein-mediated chemotaxis towards osmoprotectants and amino acids. *EMBO J.* 21, 2312–2322. doi: 10.1093/emboj/21.10.2312

Koonin, E. V., and Wolf, Y. I. (2012). Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front. Cell. Infect. Microbiol.* 2:119. doi: 10.3389/fcimb.2012.00119

Kumar, J., Kumar, M., Gupta, S., Ahmed, V., Bhambi, M., Pandey, R., et al. (2016). An improved methodology to overcome key issues in human fecal metagenomic DNA extraction. *Genom. Proteom. Bioinform.* 14, 371–378. doi: 10.1016/j.gpb.2016.06.002

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kumar Mondal, A., Kumar, J., Pandey, R., Gupta, S., Kumar, M., Bansal, G., et al. (2017). Comparative genomics of host–symbiont and free-living *Oceanobacillus* species. *Genome Biol. Evol.* 9, 1175–1182. doi: 10.1093/gbe/evx076

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Ma, B., and Gong, J. (2013). A meta-analysis of the publicly available bacterial and archaeal sequence diversity in saline soils. *World J. Microbiol. Biotechnol.* 29, 2325–2334. doi: 10.1007/s11274-013-1399-9

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221

Meena, K. K., Sorty, A. M., Bitla, U. M., Choudhary, K., Gupta, P., Pareek, A., et al. (2017). Abiotic stress responses and microbe-mediated mitigation in plants: the omics strategies. *Front. Plant Sci.* 8:172. doi: 10.3389/fpls.2017.00172

Mirete, S., Mora-Ruiz, M. R., Lamprecht-Grandío, M., de Figueras, C. G., Rosselló-Móra, R., and González-Pastor, J. E. (2015). Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. *Front. Microbiol.* 6:1121. doi: 10.3389/fmicb.2015.01121

Morowitz, M. J., Denef, V. J., Costello, E. K., Thomas, B. C., Poroyko, V., Relman, D. A., et al. (2011). Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1128–1133. doi: 10.1073/pnas.1010992108

Naughton, L. M., Blumerman, S. L., Carlberg, M., and Boyd, E. F. (2009). Osmoadaptation among *Vibrio* species and unique genomic features and physiological responses of *Vibrio parahaemolyticus*. *Appl. Environ. Microbiol.* 75, 2802–2810. doi: 10.1128/AEM.01698-08

Oubrie, A., Rozeboom, H. J., Kalk, K. H., Olsthoorn, A. J., Duine, J. A., and Dijkstra, B. W. (1999). Structure and mechanism of soluble quinoprotein glucose dehydrogenase. *EMBO J.* 18, 5187–5194. doi: 10.1093/emboj/18.19.5187

Penn, K., and Jensen, P. R. (2012). Comparative genomics reveals evidence of marine adaptation in *Salinispora* species. *BMC Genomics* 13:86. doi: 10.1186/1471-2164-13-86

Prakash, T., and Taylor, T. D. (2012). Functional assignment of metagenomic data: challenges and applications. *Brief. Bioinform.* 13, 711–727. doi: 10.1093/bib/bbs033

Roberts, M. F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Syst.* 1:5. doi: 10.1186/1746-1448-1-5

Rubiano-Labrador, C., Bland, C., Miotello, G., Armengaud, J., and Baena, S. (2015). Salt stress induced changes in the exoproteome of the halotolerant bacterium *Tistlia consotensis* deciphered by proteogenomics. *PLoS ONE* 10:e0135065. doi: 10.1371/journal.pone.0135065

Sakamoto, T., and Murata, N. (2002). Regulation of the desaturation of fatty acids and its role in tolerance to cold and salt stress. *Curr. Opin. Microbiol*. 5, 208–210. doi: 10.1016/S1369-5274(02)00306-5

Shrivastava, P., and Kumar, R. (2015). Soil salinity: a serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi J. Biol. Sci*. 22, 123–131. doi: 10.1016/j.sjbs.2014.12.001

Singh, A. H., Doerks, T., Letunic, I., Raes, J., and Bork, P. (2009). Discovering functional novelty in metagenomes: examples from light-mediated processes. *J. Bacteriol*. 191, 32–41. doi: 10.1128/JB.01084-08

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., et al. (2009). Metagenomics: concept, methodology, ecological inference and recent advances. *Biotechnol. J*. 4, 480–494. doi: 10.1002/biot.200800201

Takami, H., Takaki, Y., and Uchiyama, I. (2002). Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res.* 30, 3927–3935. doi: 10.1093/nar/gkf526

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., and Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 43, W401–W407. doi: 10.1093/nar/gkv485

Tusnády, G. E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17, 849–850. doi: 10.1093/bioinformatics/17.9.849

van Veen, J. A., van Overbeek, L. S., and van Elsas, J. D. (1997). Fate and activity of microorganisms introduced into soil. *Microbiol. Mol. Biol. Rev*. 61, 121–135.

Wang, X., Ma, Y., Huang, C., Li, J., Wan, Q., and Bi, Y. (2008). Involvement of glucose-6-phosphate dehydrogenase in reduced glutathione maintenance and hydrogen peroxide signal under salt stress. *Plant Signal. Behav*. 3, 394–395. doi: 10.4161/psb.3.6.5404

Xiao, Y., and Roberts, D. J. (2010). A review of anaerobic treatment of saline wastewater. *Environ. Technol.* 31, 1025–1043. doi: 10.1080/09593331003734202

Yadav, M., Verma, M. K., and Chauhan, N. S. (2017). A review of metabolic potential of human gut microbiome in human nutrition. *Arch Microbiol.* doi: 10.1007/s00203-017-1459-x. [Epub ahead of print].

Yan, C., Zhang, D., Raygoza Garay, J. A., Mwangi, M. M., and Bai, L. (2015). Decoupling of divergent gene regulation by sequence-specific DNA binding factors. *Nucleic Acids Res*. 43, 7292–7305. doi: 10.1093/nar/gkv618

Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y., et al. (2003). *Gemmatimonas aurantiaca* gen. nov., sp. nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes phyl*. nov. *Int. J. Syst. Evol. Microbiol*. 53, 1155–1163. doi: 10.1099/ijs.0.02520-0

Zhengbin, Z., Ping, X., Hongbo, S., Mengjun, L., Zhenyan, F., and Liye, C. (2011). Advances and prospects: biotechnologically improving crop water use efficiency. *Crit. Rev. Biotechnol.* 31, 281–293. doi: 10.3109/07388551.2010.531004

Zuleta, L. F., Italiani, V. C., and Marques, M. V. (2003). Isolation and characterization of NaCl-sensitive mutants of *Caulobactercrescentus. Appl. Environ. Microbiol.* 69, 3029–3035. doi: 10.1128/AEM.69.6.3029-3035.2003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ahmed, Verma, Gupta, Mandhan and Chauhan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

MINI REVIEW
published: 26 April 2018
doi: 10.3389/fmicb.2018.00820

A Genomic Outlook on Bioremediation: The Case of Arsenic Removal

*Frédéric Plewniak[1], Simona Crognale[2], Simona Rossetti[2] and Philippe N. Bertin[1]**

[1] Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS, Université de Strasbourg, Strasbourg, France,
[2] Istituto di Ricerca sulle Acque, Consiglio Nazionale delle Ricerche, Rome, Italy

Microorganisms play a major role in biogeochemical cycles. As such they are attractive candidates for developing new or improving existing biotechnological applications, in order to deal with the accumulation and pollution of organic and inorganic compounds. Their ability to participate in bioremediation processes mainly depends on their capacity to metabolize toxic elements and catalyze reactions resulting in, for example, precipitation, biotransformation, dissolution, or sequestration. The contribution of genomics may be of prime importance to a thorough understanding of these metabolisms and the interactions of microorganisms with pollutants at the level of both single species and microbial communities. Such approaches should pave the way for the utilization of microorganisms to design new, efficient and environmentally sound remediation strategies, as exemplified by the case of arsenic contamination, which has been declared as a major risk for human health in various parts of the world.

Keywords: genomics, arsenic, bioremediation/phytoremediation, microorganism, ecosystem ecology

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council (CONICET), Argentina

Reviewed by:
Lukasz Drewniak, University of Warsaw, Poland
Ana Isabel Pelaez, Universidad de Oviedo Mieres, Spain

****Correspondence:***
Philippe N. Bertin
philippe.bertin@unistra.fr

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *02 February 2018*
Accepted: *10 April 2018*
Published: *26 April 2018*

Citation:
Plewniak F, Crognale S, Rossetti S and Bertin PN (2018) A Genomic Outlook on Bioremediation: The Case of Arsenic Removal. Front. Microbiol. 9:820. doi: 10.3389/fmicb.2018.00820

FROM GENES TO METAGENOMES

In over three billion years of evolution, microorganisms have colonized nearly all ecological niches, including the most extreme environments. Due to their multiple metabolic activities, they play a major part in biogeochemical cycles, affecting soil productivity or water quality (Madsen, 2011) and constitute an immense reservoir of genes with high potentials for biotechnology applications. For those reasons, microorganisms from the environment have aroused a strong interest since long before the microbial genomics era. A large number of enzymes and genes coding for biocatalyzers (cellulases, proteases, lipases/esterases, glycosidases, chitinases, xylanases, phosphatases) or for enzymes involved in vitamin and antibiotic biosynthesis have thus been isolated from environmental microorganisms (Colin et al., 2015; Jacques et al., 2017; Krüger et al., 2018). Many of these enzymes have been used for research, industrial or pharmaceutical applications (Madhavan et al., 2017) like, for instance, restriction enzymes and the Taq DNA polymerase that sparked a revolution in molecular biology techniques (Ishino and Ishino, 2014).

More than 20 years ago, thanks to the rise of molecular biology and the automation of DNA sequencing, microbiology embraced genomics, the ensemble of approaches which address the organization and activity of organisms within the scope of their full genome, acknowledging that no living system can be reduced to a single gene expressed at some time or another (Bertin et al., 2015). Since the very first genome sequence from a free-living organism, *Haemophilus influenzae* Rb (Fleischmann et al., 1995), the number of new microbial genome sequences published each

year has grown exponentially to reach in 2014 a total of over 30,000 publicly available sequenced bacterial genomes (Land et al., 2015).

Yet, diversity data provided by molecular methods suggest that there remains in many ecosystems a vast majority of microorganisms belonging to taxa that have not been isolated in pure culture (Rashid and Stingl, 2015) and cultivation may be extremely difficult for a majority of them. Environmental genomic approaches could nonetheless provide access directly to the genome of uncultivated organisms like '*Candidatus* Desulforudis audaxviator,' which practically represents the sole species present in a gold mine and can fix nitrogen using a cellular mechanism similar to that of Archaea (Chivian et al., 2008). Metagenomic analyses of nitrogen metabolism in anaerobic enriched cultures also led to the reconstruction of prokaryotic genomes such as *Kuenenia stuttgartiensis* (Strous et al., 2006), '*Candidatus* Nitrospira defluvii' (Lücker et al., 2010) or the archaeon '*Candidatus* Methanoperedens nitroreducens' (Haroon et al., 2013) involved in oceanic ammonium oxidation, nitrite oxidation in sewage treatment plant sludge and anaerobic oxidization of methane coupled to nitrate reduction, respectively. Similarly, the genome of an iron-oxidizer strain belonging to the *Ferrovum* genus was reconstructed from a mixed culture grown from samples collected in a mine water treatment plant (Ullrich et al., 2016).

Though molecular techniques associated with bioinformatic and genome-mining methods are invaluable tools to reveal the potential in genome data (Machado et al., 2017; Vallenet et al., 2017), cultivation remains an important challenge in microbiology, necessary for expanding our knowledge of microorganisms' physiology and for bioremediation (Overmann et al., 2017). However, microorganisms from the environment may require essential nutrients or particular growth conditions, or may be extremely slow growers or obligate symbionts. Although tackling these issues generally demands strenuous efforts to design and test many isolation media, genome characterization may highlight metabolic characteristics of the targeted organism that could be leveraged to select and cultivate a given strain (Garza and Dutilh, 2015). This strategy allowed the isolation of the first nitrifying archaeon (Schleper et al., 2005) after an analysis of the Sargasso Sea metagenome (Venter et al., 2004) had detected on the same DNA fragment an Archaea-specific ribosomal gene and a gene coding for the ammonium monooxygenase, a key enzyme in nitrification. Subsequent physiological studies showed that the nitrification function was indeed expressed (Könneke et al., 2005). Another example is provided by the isolation of *Leptospirillum ferrodiazotrophum* (Tyson et al., 2005), which a previous metagenomic study had shown to be the only strain in an acid mine tailing to be able to fix nitrogen.

Beyond approaches centered on single organisms, the developments of genomics have rendered possible a global view of microbial communities that could help a better understanding of natural remediation processes and identifying candidate species for the design of bioremediation treatment plants. In this respect, high-throughput tools such as microarrays have allowed to address ecological questions related to the structure and function of microbial communities. Developed from the genomic data present in databases, such approaches may be helpful to study the diversity and dynamics of microbial populations using nucleic acids extraction and hybridization (Zhou et al., 2015). They were successfully used, for example, to examine the responses of microbial communities after the wreck of a drilling rig in the Gulf of Mexico had released about 5 million barrels of crude oil (Beazley et al., 2012). This study suggested that the microbial community of the rhizosphere in the affected coastal salt marsh could strongly contribute to hydrocarbon natural remediation. Recently, the combination of ribosomal 16S RNA gene high-throughput sequencing with DNA-based stable isotope probing in activated sludge samples incubated with $Na_2{}^{13}CO_3$ uncovered the dynamics of ammonium-oxidizing microorganism abundance and the relative importance of archaeal and bacterial ammonium oxidation activities in a waste water treatment plant (Pan et al., 2018).

In recent years, the development of high-throughput sequencing and assembly software has allowed to determine the complete genome sequence of uncultivated microorganisms from direct sequencing of metagenomic libraries or environmental DNA from complex microbial communities. Despite a number of critical issues regarding sampling, assembly or annotation (Teeling and Glöckner, 2012; Thomas et al., 2012), more than 10,000 metagenome projects are now referenced in the Genomes Online Database (Mukherjee et al., 2016). This number is expected to increase dramatically with such massive projects as the Earth Microbiome Project whose goal is to produce a global Gene Atlas of microbial communities encompassing an estimated 500,000 genomes (Gilbert et al., 2010; Thompson et al., 2017).

Environmental genomics now permits the study of the organisms in an ecosystem as a set of elements behaving within a complex network of interactions (**Figure 1**). For example, a genome-scale study of the complex symbiosis between the termite *Macrotermes natalensis*, its domesticated fungus and several gut bacterial communities demonstrated the cooperation between microorganisms in plant biomass conversion. The results showed that the insect provides the infrastructure allowing carbohydrate decomposition thanks to the functional complementarity between the fungus and the gut microbiota (Poulsen et al., 2014). More recently, the reconstruction of 2540 genomes using metagenomic data from 15 different sediment and groundwater environments allowed to highlight the key inter-organism interactions relevant to biogeochemical cycles in an aquifer in Colorado, United States (Anantharaman et al., 2016). Applying those approaches to various biotopes may thus provide valuable insights into the functioning of ecosystems, including polluted environments whose microbial communities could constitute prospective candidates for bioremediation.

ARSENIC BIOREMEDIATION AND 'OMICS' APPROACHES: A CASE STUDY

Long-term exposure to arsenic represents a serious threat to human health worldwide (Nordstrom, 2002). Even though the

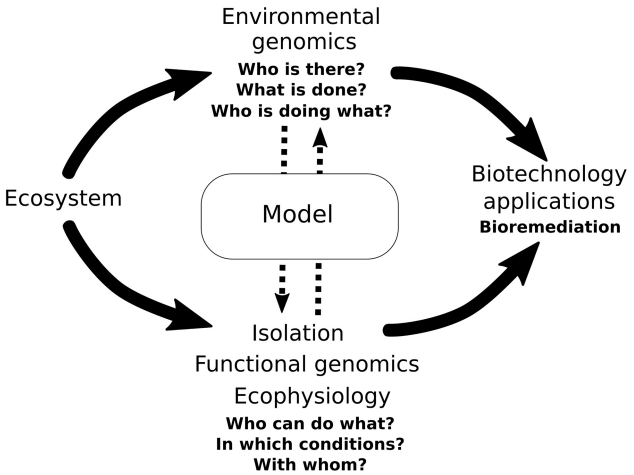


FIGURE 1 | Multi-omic studies of microorganisms from the environment provide an integrated image or model of the microbial processes involved in the ecosystem's functioning and that may be utilized for bioremediation or other biotechnology applications.

occurrence of this element in drinking water constitutes the major source of exposure, recent studies on risks of arsenic accumulation in food revealed its presence in fish and crops cultivated with arsenic-contaminated waters (WHO, 2011; Jackson et al., 2012; Molin et al., 2015; Carlin et al., 2016). Numerous physico-chemical methods are commonly used for the treatment of arsenic-rich waters: coagulation/filtration, ion exchange, enhanced lime softening, adsorption and reverse osmosis (Ng et al., 2004; Nicomel et al., 2016). Over the last years, in a search for sustainable and cost-effective methods for water treatment, arsenic remediation turned to the potentialities of biological approaches. The use of rhizosphere microorganisms was recently investigated for their capacity to enhance phytoremediation of arsenic-contaminated environments (Ma et al., 2016). In particular, several arsenic-resistant microorganisms belonging to various genera, e.g., *Bacillus*, *Achromobacter*, *Brevundimonas*, *Microbacterium*, *Ochrobactrum*, *Pseudomonas*, *Comamonas*, *Stenotrophomonas*, *Ensifer* were reported to decrease toxic effects of arsenic and enhance plant growth by acting on arsenic mobilization and accumulation in plants (Cavalca et al., 2010; Ghosh et al., 2011; Wang et al., 2011; Yang et al., 2012; Pandey et al., 2013; Mallick et al., 2014, 2018; Mesa et al., 2017). The ability of fungi to resist, solubilize, transform or uptake metal species could also be used in mycoremediation of arsenic-contaminated soil (Singh et al., 2015; Srivastava et al., 2011). The production of volatile trimethylarsine by reductive methylation from inorganic and methylated arsenic compounds was reported in several fungal strains, e.g., *Aspergillus glaucum*, *Candida humicola*, *Scopulariopsis brevicaulis*, *Gliocladium roseum*, *Penicillium gladioli,* and *Fusarium* spp. (Cullen and Reimer, 1989; Lin, 2008). Bioaugmentation could thus represent a strategy to enhance the efficiency of As removal from waters and soils by the addition of specialized bacteria or fungi, either natural or genetically engineered, able to directly remove As by volatilization (Edvantoro et al., 2004; Chen P. et al., 2017) or indirectly through the formation of biogenic Fe-Mn oxides (Bai et al., 2016). However, despite an increasing interest for mycoremediation and rhizoremediation of arsenic contamination, still very little is known about their scalability.

To date, the bioremediation of arsenic-rich environments is mainly based on the use of microorganisms able to resist or metabolize arsenic through oxidoreduction reactions (Huang, 2014). Over the last decades the ecology of arsenic has been widely studied and several arsenic-metabolizing microorganisms isolated from various ecosystems have been characterized at the genomic level (Oremland and Stolz, 2003; Páez-Espino et al., 2009; Andres and Bertin, 2016). *Herminiimonas arsenicoxydans* was the first arsenic-metabolizing bacterium to be described. This β-proteobacterium isolated from an industrial wastewater treatment plant in Germany was shown to resist to high levels of arsenic and to oxidize arsenite, As(III), into arsenate, As(V) (Muller et al., 2007). Functional genomics demonstrated that this arsenic response is biphasic: *H. arsenicoxydans* activates the resistance response based in part on the induction of efflux mechanisms before inducing the detoxification processes leading to As(III) oxidation (Cleiss-Arnold et al., 2010; Koechler et al., 2010). Additionally, electron microscopy revealed that the strain is able to sequester arsenic within an exopolysaccharide (EPS) matrix (Muller et al., 2007). *Thiomonas* sp. 3As isolated from an abandoned mine in France was also shown to produce large amounts of EPS in the presence of arsenite, making it a good candidate for the development of bioremediation strategies relying on biofilm-based bioreactors (Arsène-Ploetze et al., 2010). A strain belonging to the *Rhizobium* genus isolated from an Australian gold mine was shown to carry arsenic resistance and detoxification genes on a large plasmid, which could provide an interesting genetic tool to transfer arsenic detoxification capacity into closely related plant-associated bacteria with the perspective of phytoremediation (Andres et al., 2013). More recently, the genome of two arsenite-oxidizing strains hyper-tolerant to arsenite was fully described: *Halomonas* A3H3 isolated from multicontaminated sediments in Mediterranean Sea (Koechler et al., 2013), and *Pseudomonas xanthomarina* S11 isolated from an arsenic-contaminated former gold mine in France (Koechler et al., 2015). Overall, the identification and the exploitation of microbial metabolic potentialities for arsenic-contaminated water treatment are considered an emerging challenge as mirrored by an increasing number of recent studies (Crognale et al., 2017). Among the available bacterial-driven processes, bioprecipitation, biosynthesis of adsorbent materials, biosorption and biovolatilization, involving several microorganisms (**Table 1**), are the most interestingly described for bioremediation of arsenic-contaminated waters (Fazi et al., 2016).

In recent years, several environmental genomic studies of arsenic-contaminated ecosystems have been conducted (Huang et al., 2016) and the molecular mechanisms involved have been recently reviewed in detail (Andres and Bertin, 2016). A metagenomic study of an acid mine drainage in France yielded nearly complete reconstructions of seven microbial genomes,

TABLE 1 | Microorganisms used in As-removal processes from waters.

Microorganism	Process	Reference
Ralstonia eutropha	Bio-adsorption	Mondal et al., 2008
Rhodopseuodomonas palustris	Bio-volatilization	Liu et al., 2011
Sphingomonas desiccabilis	Bio-volatilization	Liu et al., 2011
Bacillus idriensis	Bio-volatilization	Liu et al., 2011
Cyanobacteria	Bio-volatilization	Yin et al., 2011
Klebsiella oxytoca	Bio-synthesis of adsorbent materials	Casentini et al., 2015
Mixed microbial community	Bio-precipitation	Omoregie et al., 2013
Gallionella ferruginea and *Leptothrix ochracea*	Microbial Fe-oxidation coupled to As removal	Katsoyiannis and Zouboulis, 2004
Desulfotomaculum auripigmentum	As and Fe–As sulfide precipitation driven by sulfate reducers	Newman et al., 1997
Mixed Sulfate Reducing Bacteria	As removal driven by sulfate reduction processes	Teclu et al., 2008; Serrano and Leiva, 2017
Mixed microbial community	Microbial Fe- and Mn-oxidation coupled to As removal	Thapa Chhetri et al., 2014
Mixed microbial community	As removal via co-oxidation with Fe and sorption or co-precipitation with Fe(III) (oxyhydr)oxides	Nitzsche et al., 2015
Mixed microbial community	Microbial Fe- and Mn-oxidation coupled to AsIII removal	Yang et al., 2014
Mixed microbial community	As(III) microbial oxidation coupled to FeII oxidation	Kamei-Ishikawa et al., 2017
Aliihoeflea sp. *2WW*	As(III) microbial oxidation	Corsini et al., 2014
Thiomonas arsenivorans	As(III) microbial oxidation	Wan et al., 2010; Dastidar and Wang, 2012
Rhodococcus equi	As(III) microbial oxidation	Bag et al., 2010
CAsO1 bacterial consortium	As(III) microbial oxidation	Battaglia-Brunet et al., 2002; Michel et al., 2007
Ensifer adhaerens	As(III) microbial oxidation	Ito et al., 2012
Mixed microbial community	As(III) microbial oxidation	Gude et al., 2018
Mixed microbial community	As(III) microbial oxidation	Li et al., 2016
Mixed microbial community	As(III) microbial oxidation	Sun et al., 2011
Mixed microbial community	Anoxic As(III) microbial oxidation coupled with chemolithotrophic denitrification	Sun et al., 2010

providing a better understanding of the arsenic metabolism and natural attenuation which significantly reduce arsenic concentration along the creek, thanks to arsenite oxidation followed by co-precipitation with iron and sulfur. This analysis led to the identification of the corresponding genes, in particular *aio* coding for arsenite oxidase in *Thiomonas* sp. and *rus* coding for rusticyanin in *Acidithiobacillus* sp. (Bertin et al., 2011). A comparative metagenomic study of sediments in two harbors on the Mediterranean French coast, focusing on sequence markers specific for sulfur-metabolizing bacteria uncovered a correspondence between biotic sulfate reduction and the abiotic production of highly soluble thioarsenical compounds. In combination with arsenate reduction these processes, which favor arsenic dispersion in the water column, could explain the higher mobility of arsenic observed on the most contaminated site (Plewniak et al., 2013). Recently, the assembly of 27 Micrarcheota and 12 Parvarchaeota new genomes from 12 acid mine drainage and hot spring metagenomes was reported in a study targeting Archaeal Richmond Mine Acidophilic Nanoorganisms. The analysis of these almost complete genomes suggests a possible contribution of these organisms to carbon and nitrogen cycling by organic matter degradation, as well as to iron oxidation (Chen L.-X. et al., 2017). Those studies suggest that arsenic bioremediation strategies could be based upon microbial communities with iron, sulfur, and arsenic metabolism capacities and highlight the importance of metabolisms other than those of metals in arsenic removal. In this respect, mixed microbial communities were tested for bio-precipitation capacity and arsenic removal coupled with iron and manganese oxidation in filtration systems (**Table 1**) and recently, the use of acid/metal-tolerant sulfate reducing bacteria was applied for arsenic removal from an acid mine drainage (Serrano and Leiva, 2017).

As(III) microbial oxidation can also be coupled to commonly used adsorption removal technology, without any chemicals addition nor toxic by-products (Bahar et al., 2013). The As(III)-oxidation potentialities of several As(III)-oxidizing microorganisms, such as *Aliihoeflea* sp. 2WW, *Thiomonas arsenivorans* strain b6, *Ensifer adhaerens*, *Rhodococcus equi* and other As(III)-oxidizing mixed bacterial populations as planktonic cells or associated with biofilms were successfully tested in lab-scale experiments for treating contaminated water (**Table 1**). Moreover, the anoxic As(III) microbial oxidation coupled with chemolithotrophic denitrification was successfully employed in the treatment of arsenic in bioreactors (Sun et al., 2010). To date, only one case study of full-scale treatment of arsenic contaminated groundwater using biological As(III) oxidation has been documented in the scientific literature (Katsoyiannis et al., 2008). This multi-stage treatment method was based on the biological oxidation of NH_4^+ and Mn(II) for the simultaneous As(III) oxidation and subsequent As(V) removal by coagulation. However, As removal is strongly dependent on Fe(II) and Mn(II) concentrations since the process relies on the sorption of As on iron and manganese oxides produced by autochthonous Fe(II)- and Mn(II)-oxidizing bacteria.

Although several studies demonstrated the efficacy of arsenic removal from water by microorganisms, these approaches are yet to be fully exploited for arsenic remediation, and knowledge about the diversity and distribution of functional genes controlling arsenic transformation in such processes is still quite fragmentary (Andres and Bertin, 2016; Crognale et al., 2017). The industrial application of arsenic removal from water still requires further evaluation in real situation of additional aspects such as the influence on microbial As(III) oxidation of geometric and hydraulic parameters in column systems or the requirement for carbon supply to support fast reactions. Although recent batch experiment works are addressing the question of the effects of nutrient sources and temperature in acid mine drainage (Tardy et al., 2018), there is still a want of further genomic and metagenomic studies of arsenic-contaminated ecosystems addressing not only the metabolisms of metals, arsenic and sulfurs but the full-range of microbial metabolic capacities. Such studies will be necessary for understanding the complex trophic interaction network of microorganisms in those ecosystems and for designing optimized artificial microbial communities that could be exploited in large-scale arsenic remediation systems.

CONCLUSIONS AND PERSPECTIVES

At the interface between molecular biology and ecology, environmental genomic DNA sequencing techniques allow to reach, beyond the mere description of a simple organism, the characterization of complex microbial communities including organisms recalcitrant to isolation and culture. In association with global functional approaches – metatranscriptomics, metaproteomics, metabolomics including stable-isotope probing (Fischer et al., 2016; Musat et al., 2016; Vogt et al., 2016; Zuñiga et al., 2017) – these techniques help increasing our knowledge of the functioning of ecosystems. Additionally, the sequencing depth attained by these new technologies can give access to the less represented species of an ecosystem (the rare biosphere). Allowing fast and inexpensive massive characterization of microbial communities, they could also be an asset for the continuous monitoring of microbial communities involved in bioremediation processes to avoid changes that could compromise the efficiency of the treatment (Lovley, 2003; Stenuit et al., 2008; Techtmann and Hazen, 2016). In combination with the indispensable experimentations in the laboratory and in the field, these approaches require the development of efficient reproducible sampling and extraction methods as well as of robust and new computing solutions for storing, exchanging, and analyzing the huge amounts of data they produce. Indeed, power analysis and sample size requirements estimation for high-throughput sequencing data demand computations of much higher complexity than classical statistical analyses and must be fine-tuned to the type of problem that is being addressed (Pasolli et al., 2016; Li et al., 2017). It is moreover necessary that all published studies include complementary data (meta-data) which should be collected for every genome/metagenome to permit the proper exploitation of data (Satinsky et al., 2013) as defined by the Genomic Standard Consortium (Yilmaz et al., 2011).

The public access on sites like the EBI Metagenomics (Mitchell et al., 2017) to thousands of metagenomic samples combined with big data analysis, data mining algorithms and metabolic modeling constitutes an unprecedented opportunity to study and understand how the different components of an ecosystem may function together in relation with environmental biotic and abiotic factors, largely surpassing mere inventories of biological objects. A better understanding of the concerned organisms, of their spatial and temporal distribution, of the adaptive and evolutive processes at stake and of the metabolic interactions they develop should thus provide an integrated image of the microbial communities and metabolic functions involved in the microbiological processes underlying arsenic removal from water. Using *ad hoc* predictive models, such knowledge may be expected to permit the optimal utilization of microorganisms' properties in biotechnological applications and bioremediation processes.

AUTHOR CONTRIBUTIONS

PB and FP organized the content of the entire manuscript and wrote the genomics section. SC and SR wrote the section on bioremediation.

REFERENCES

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219

Andres, J., Arsène-Ploetze, F., Barbe, V., Brochier-Armanet, C., Cleiss-Arnold, J., Coppée, J. Y., et al. (2013). Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium Rhizobium sp, NT-26. *Genome Biol. Evol.* 5, 934–953. doi: 10.1093/gbe/evt061

Andres, J., and Bertin, P. (2016). The microbial genomics of arsenic. *FEMS Microbiol. Rev.* 40, 299–322. doi: 10.1093/femsre/fuv050

Arsène-Ploetze, F., Koechler, S., Marchal, M., Coppe, J., Chandler, M., Bonnefoy, V., et al. (2010). Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet.* 6:e1000859. doi: 10.1371/journal.pgen.1000859

Bag, P., Bhattacharya, P., and Chowdhury, R. (2010). Bio-detoxification of arsenic laden groundwater through a packed bed column of a continuous flow reactor using immobilized cells. *Soil Sediment Contam.* 19, 455–466. doi: 10.1080/15320383.2010.486050

Bahar, M. M., Megharaj, M., and Naidu, R. (2013). Bioremediation of arsenic-contaminated water: recent advances and future prospects. *Water Air Soil Pollut.* 224:1722. doi: 10.1007/s11270-013-1722-y

Bai, Y., Chang, Y., Liang, J., Chen, C., and Qu, J. (2016). Treatment of groundwater containing Mn(II). Fe(II), As(III) and Sb(III) by bioaugmented quartz-sand filters. *Water Res.* 106, 126–134. doi: 10.1016/j.watres.2016.09.040

Battaglia-Brunet, F., Dictor, M., Garrido, F., Crouzet, C., and Morin, D. (2002). An arsenic(III)-oxidizing bacterial population: selection, characterization, and performance in reactors. *J. Appl. Microbiol.* 93, 656–667. doi: 10.1046/j.1365-2672.2002.01726.x

Beazley, M. J., Martinez, R. J., Rajan, S., Powell, J., Piceno, Y. M., Tom, L. M., et al. (2012). Microbial community analysis of a coastal salt marsh affected by the

Deepwater Horizon Oil spill. *PLoS One* 7:e41305. doi: 10.1371/journal.pone.0041305

Bertin, P., Michotey, V., and Normand, P. (2015). "Contributions of descriptive and functional genomics to microbial ecology," in *Environmental Microbiology: Fundamentals and Applications: Microbial Ecology*, eds J. C. Bertrand, et al. (Berlin: Springer), 831–846.

Bertin, P. N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., et al. (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J.* 5, 1735–1747. doi: 10.1038/ismej.2011.51

Carlin, D. J., Naujokas, M. F., Bradham, K. D., Cowden, J., Heacock, M., Henry, H. F., et al. (2016). Arsenic and environmental health: state of the science and future research opportunities. *Environ. Health Perspect.* 124, 890–899. doi: 10.1289/ehp.1510209

Casentini, B., Rossetti, S., Gallo, M., and Baldi, F. (2015). "Potentialities of biogenerated iron hydroxides nanoparticles in arsenic water treatment," in *Proceeding of the 6th European Bioremediation Conference*, Chania.

Cavalca, L., Zanchi, R., Corsini, A., Colombo, M., Romagnoli, C., Canzi, E., et al. (2010). Arsenic-resistant bacteria associated with roots of the wild *Cirsium arvense* (L.) plant from an arsenic polluted soil, and screening of potential plant growth-promoting characteristics. *Syst. Appl. Microbiol.* 33, 154–164. doi: 10.1016/j.syapm.2010.02.004

Chen, L.-X., Méndez-García, C., Dombrowski, N., Servín-Garcidueñas, L. E., Eloe-Fadrosh, E. A., Fang, B.-Z., et al. (2017). Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME J.* 12,756–775. doi: 10.1038/s41396-017-0002-z

Chen, P., Li, J., Wang, H.-Y., Zheng, R.-L., and Sun, G.-X. (2017). Evaluation of bioaugmentation and biostimulation on arsenic remediation in soil through biovolatilization. *Environ. Sci. Pollut. Res.* 24, 21739–21749. doi: 10.1007/s11356-017-9816-5

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science* 322, 275–278. doi: 10.1126/science.1155495

Cleiss-Arnold, J., Koechler, S., Proux, C., Fardeau, M. L., Dillies, M. A., Coppee, J. Y., et al. (2010). Temporal transcriptomic response during arsenic stress in Herminiimonas arsenicoxydans. *BMC Genomics* 11:709. doi: 10.1186/1471-2164-11-709

Colin, P. Y., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., et al. (2015). Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* 6:10008. doi: 10.1038/ncomms10008

Corsini, A., Zaccheo, P., Muyzer, G., Andreoni, V., and Cavalca, L. (2014). Arsenic transforming abilities of groundwater bacteria and the combined use of *Aliihoeflea* sp. strain 2WW and goethite in metalloid removal. *J. Hazard. Mater.* 269, 89–97. doi: 10.1016/j.jhazmat.2013.12.037

Crognale, S., Amalfitano, S., Casentini, B., Fazi, S., Petruccioli, M., and Rossetti, S. (2017). Arsenic-related microorganisms in groundwater?: a review on distribution, metabolic activities and potential use in arsenic removal processes. *Rev. Environ. Sci. Biotechnol.* 16, 647–665. doi: 10.1007/s11157-017-9448-8

Cullen, W. R., and Reimer, K. J. (1989). Arsenic-speciation in the environment. *Chem. Rev.* 89, 713–763. doi: 10.1021/cr00094a002

Dastidar, A., and Wang, Y.-T. (2012). Modeling arsenite oxidation by chemoautotrophic *Thiomonas arsenivorans* strain b6 in a packed-bed bioreactor. *Sci. Total Environ.* 432, 113–121. doi: 10.1016/j.scitotenv.2012.05.051

Edvantoro, B. B., Naidu, R., Megharaj, M., Merrington, G., and Singleton, I. (2004). Microbial formation of volatile arsenic in cattle dip site soils contaminated with arsenic and DDT. *Appl. Soil Ecol.* 25, 207–217. doi: 10.1016/j.apsoil.2003.09.006

Fazi, S., Amalfitano, S., Casentini, B., Davolos, D., Pietrangeli, B., Crognale, S., et al. (2016). Arsenic removal from naturally contaminated waters: a review of methods combining chemical and biological treatments. *Rend. Lincei* 27, 51–58. doi: 10.1007/s12210-015-0461-y

Fischer, A., Manefield, M., and Bombach, P. (2016). Application of stable isotope tools for evaluating natural and stimulated biodegradation of organic pollutants in field studies. *Curr. Opin. Biotechnol.* 41, 99–107. doi: 10.1016/j.copbio.2016.04.026

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512. doi: 10.1126/science.7542800

Garza, D. R., and Dutilh, B. E. (2015). From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cell. Mol. Life Sci.* 72, 4287–4308. doi: 10.1007/s00018-015-2004-1

Ghosh, P., Rathinasabapathi, B., and Ma, L. Q. (2011). Arsenic-resistant bacteria solubilized arsenic in the growth media and increased growth of arsenic hyperaccumulator *Pteris vittata* L. *Bioresour. Technol.* 102, 8756–8761. doi: 10.1016/j.biortech.2011.07.064

Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., et al. (2010). Meeting report: the terabase metagenomics workshop and the vision of an earth microbiome project. *Stand. Genomic Sci.* 3, 243–248. doi: 10.4056/sigs.1433550

Gude, J. C. J., Rietveld, L. C., and van Halem, D. (2018). Biological As(III) oxidation in rapid sand filters. *J. Water Process Eng.* 21, 107–115. doi: 10.1016/j.jwpe.2017.12.003

Haroon, M. F., Hu, S., Shi, Y., Imerlfort, M., Keller, J., Hugenholtz, P., et al. (2013). Anaerobic oxidation of methane coupled to nitrate redcution in a novel archaeal lineage. *Nature* 500, 567–570. doi: 10.1038/nature12375

Huang, J. (2014). Impact of microorganisms on arsenic biogeochemistry?: a review. *Water Air Soil Pollut.* 225:1848. doi: 10.1007/s11270-013-1848-y

Huang, L.-N., Kuang, J.-L., and Shu, W.-S. (2016). Microbial ecology and evolution in the acid mine drainage model system. *Trends Microbiol.* 24, 581–593. doi: 10.1016/j.tim.2016.03.004

Ishino, S., and Ishino, Y. (2014). DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field. *Front. Microbiol.* 5:465. doi: 10.3389/fmicb.2014.00465

Ito, A., Miura, J.-I., Ishikawa, N., and Umita, T. (2012). Biological oxidation of arsenite in synthetic groundwater using immobilised bacteria. *Water Res.* 46, 4825–4831. doi: 10.1016/j.watres.2012.06.013

Jackson, B. P., Taylor, V. F., Karagas, M. R., Punshon, T., and Cottingham, K. L. (2012). Arsenic, organic foods, and brown rice syrup. *Environ. Health Perspect.* 120, 623–626. doi: 10.1289/ehp.1104619

Jacques, P., Béchet, M., Bigan, M., Caly, D., Chataigné, G., Coutte, F., et al. (2017). High-throughput strategies for the discovery and engineering of enzymes for biocatalysis. *Bioprocess Biosyst. Eng.* 40, 161–180. doi: 10.1007/s00449-016-1690-x

Kamei-Ishikawa, N., Segawa, N., Yamazaki, D., Ito, A., and Umita, T. (2017). Arsenic removal from arsenic-contaminated water by biological arsenite oxidation and chemical ferrous iron oxidation using a down-flow hanging sponge reactor. *Water Supply* 17, 1249–1259. doi: 10.2166/ws.2017.025

Katsoyiannis, I. A., Zikoudi, A., and Hug, S. J. (2008). Arsenic removal from groundwaters containing iron, ammonium, manganese and phosphate: a case study from a treatment unit in northern Greece. *Desalination* 224, 330–339. doi: 10.1016/j.desal.2007.06.014

Katsoyiannis, I. A., and Zouboulis, A. I. (2004). Application of biological processes for the removal of arsenic from groundwaters. *Water Res.* 38, 17–26. doi: 10.1016/j.watres.2003.09.011

Koechler, S., Arsène-Ploetze, F., Brochier, C., Cholet, F., Jost, B., Lièvremont, D., et al. (2015). Constitutive arsenite oxidase expression detected in the arsenic hypertolerant *Pseudomonas xanthomarina* S11. *Res. Microbiol.* 166, 205–214. doi: 10.1016/j.resmic.2015.02.010

Koechler, S., Plewniak, F., Barbe, V., Battaglia-Brunet, F., Jost, B., Joulian, C., et al. (2013). Genome sequence of *Halomonas* sp. strain A3H3, isolated from arsenic-rich marine sediments. *Genome Announc.* 1:e00819-13. doi: 10.1128/genomeA.00819-13

Koechler, S., Cleiss-Arnold, J., Proux, C., Sismeiro, O., Dillies, M.-A., Goulhen-Chollet, F., et al. (2010). Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiol.* 10:53. doi: 10.1186/1471-2180-10-53

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 7058, 543–546. doi: 10.1038/nature03911

Krüger, A., Schäfers, C., Schröder, C., and Antranikian, G. (2018). Towards a sustainable biobased industry – Highlighting the impact of extremophiles. *New Biotechnol.* 40, 144–153. doi: 10.1016/j.nbt.2017.05.002

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15, 141–161. doi: 10.1007/s10142-015-0433-4

Li, C.-I., Samuels, D. C., Zhao, Y.-Y., Shyr, Y., and Guo, Y. (2017). Power and sample size calculations for high-throughput sequencing-based experiments. *Brief. Bioinform.* 5, 1752–1779. doi: 10.1093/bib/bbx061

Li, H., Zeng, X. C., He, Z., Chen, X., Guoji, E., Han, Y., et al. (2016). Long-term performance of rapid oxidation of arsenite in simulated groundwater using a population of arsenite-oxidizing microorganisms in a bioreactor. *Water Res.* 101, 393–401. doi: 10.1016/j.watres.2016.05.058

Lin, Z.-Q. (2008). "Volatilization," in *Ecological Processes. Encyclopedia of Ecology*, Vol. 5, eds S. E. Jorgensen and B. D. Fath (Oxford: Elsevier), 3700–3705. doi: 10.1016/B978-008045405-4.00302-5

Liu, S., Zhang, F., Chen, J., and Sun, G. (2011). Arsenic removal from contaminated soil via biovolatilization by genetically engineered bacteria under laboratory conditions. *J. Environ. Sci.* 23, 1544–1550. doi: 10.1016/S1001-0742(10)60570-0

Lovley, D. R. (2003). Cleaning up with genomics: applying molecular biology to bioremediation. *Nat. Rev. Microbiol.* 1, 35–44. doi: 10.1038/nrmicro731

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., et al. (2010). A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13479–13484. doi: 10.1073/pnas.1003860107

Ma, Y., Rajkumar, M., Zhang, C., and Freitas, H. (2016). Beneficial role of bacterial endophytes in heavy metal phytoremediation. *J. Environ. Manage.* 174, 14–25. doi: 10.1016/j.jenvman.2016.02.047

Machado, H., Tuttle, R. N., and Jensen, P. R. (2017). Omics-based natural product discovery and the lexicon of genone mining. *Curr. Opin. Microbiol.* 39, 136–142. doi: 10.1016/j.mib.2017.10.025

Madhavan, A., Sindhu, R., Binod, P., Sukumaran, R., and Pandey, A. (2017). Strategies for design of improved biocatalysts for industrial applications. *Bioresour. Technol.* 245, 1304–1313. doi: 10.1016/j.biortech.2017.05.031

Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* 22, 456–464. doi: 10.1016/j.copbio.2011.01.008

Mallick, I., Bhattacharyya, C., Mukherji, S., Dey, D., Sarkar, S. C., Mukhopadhyay, U. K., et al. (2018). Effective rhizoinoculation and biofilmformation by arsenic immobilizing halophilic plant growth promoting bacteria (PGPB) isolated from mangrove rhizosphere: a step towards arsenic rhizoremediation. *Sci. Total Environ.* 61, 1239–1250. doi: 10.1016/j.scitotenv.2017.07.234

Mallick, I., Tofajjen Hossain, S. K., Sinha, S., and Mukherjee, S. K. (2014). *Brevibacillus* sp. KUMAs2, a bacterial isolate for possible bioremediation of arsenic in rhizosphere. *Ecotoxicol. Environ. Saf.* 107, 236–244. doi: 10.1016/j.ecoenv.2014.06.007

Mesa, V., Navazas, A., Gonzalez-Gill, R., Gonzalez, A., Weyens, N., Lauga, B., et al. (2017). Use of endophytic and rhizosphere bacteria to improve phytoremediation of arsenic-contaminated industrial soils by Autochthonous *Betula celtiberica*. *Appl. Environ. Microbiol.* 83:e03411-16. doi: 10.1128/AEM.03411-16

Michel, C., Jean, M., Coulon, S., Dictor, M.-C., Delorme, F., Morin, D., et al. (2007). Biofilms of As(III)-oxidising bacteria: formation and activity studies for bioremediation process development. *Appl. Microbiol. Biotechnol.* 77, 457–467. doi: 10.1007/s00253-007-1169-4

Mitchell, A. L., Scheremetjew, M., Denise, H., Potter, S., Tarkowska, A., Qureshi, M., et al. (2017). EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res.* 46, D726–D735. doi: 10.1093/nar/gkx967

Molin, M., Ulven, S. M., Meltzer, H. M., and Alexander, J. (2015). Arsenic in the human food chain, biotransformation and toxicology—review focusing on seafood arsenic. *J. Trace Elem. Med. Biol.* 31, 249–259. doi: 10.1016/j.jtemb.2015.01.010

Mondal, P., Majumder, C. B., and Mohanty, B. (2008). Treatment of arsenic contaminated water in a batch reactor by using *Ralstonia eutropha* MTCC 2487 and granular activated carbon. *J. Hazard. Mater.* 153, 588–599. doi: 10.1016/j.jhazmat.2007.09.028

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., et al. (2016). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 45, D446–D456. doi: 10.1093/nar/gkw992

Muller, D., Médigue, C., Koechler, S., Barbe, V., Barakat, M., Talla, E., et al. (2007). A tale of two oxidation states?: bacterial colonization of arsenic-rich environments. *PLoS Genet.* 3:e53. doi: 10.1371/journal.pgen.0030053

Musat, N., Musat, F., Weber, P. K., and Pett-Ridge, J. (2016). Tracking microbial interactions with NanoSIMS. *Curr. Opin. Biotechnol.* 41, 114–121. doi: 10.1016/j.copbio.2016.06.007

Newman, D. K., Kennedy, E. K., Coates, J. D., Ahmann, D., Ellis, D. J., Lovley, D. R., et al. (1997). Dissimilatory arsenate and sulfate reduction in *Desulfotomaculum auripigmentum* sp. nov. *Arch. Microbiol.* 168, 380–388. doi: 10.1007/s002030050512

Ng, K. S., Ujang, Z., and Le-Clech, P. (2004). Arsenic removal technologies for drinking water treatment. *Rev. Environ. Sci. Biotechnol.* 3, 43–53. doi: 10.1023/B:RESB.0000040054.28151.84

Nicomel, N. R., Leus, K., Folens, K., Van Der Voort, P., and Du Laing, G. (2016). Technologies for arsenic removal from water: current status and future perspectives. *Int. J. Environ. Res. Public Health* 13, 1–24. doi: 10.3390/ijerph13010062

Nitzsche, K. S., Weigold, P., Lösekann-Behrens, T., Kappler, A., and Behrens, S. (2015). Microbial community composition of a household sand filter used for arsenic, iron, and manganese removal from groundwater in Vietnam. *Chemosphere* 138, 47–59. doi: 10.1016/j.chemosphere.2015.05.032

Nordstrom, D. K. (2002). Public health—worldwide occurrences of arsenic in groundwater. *Science* 296, 2143–2145. doi: 10.1126/science.1072375

Omoregie, E. O., Couture, R., Van Cappellen, P., Corkhill, C. L., Charnock, J. M., Polya, D. A., et al. (2013). Arsenic bioremediation by biogenic iron oxides and sulfides. *Appl. Environ. Microbiol.* 79, 4325–4335. doi: 10.1128/AEM.00683-13

Oremland, R. S., and Stolz, J. F. (2003). The ecology of arsenic. *Science* 300, 939–944. doi: 10.1126/science.1081903

Overmann, J., Abt, B., and Sikorski, J. (2017). Present and future of culturing bacteria. *Annu. Rev. Microbiol.* 71, 711–730. doi: 10.1146/annurev-micro-090816-093449

Páez-Espino, D., Tamames, J., de Lorenzo, V., and Canovas, D. (2009). Microbial responses to environmental arsenic. *Biometals* 22, 117–130. doi: 10.1007/s10534-008-9195-y

Pan, K.-L., Gao, J.-F., Li, H.-Y., Fan, X.-Y., Li, D.-C., and Jiang, H. (2018). Ammonia-oxidizing bacteria dominate ammonia oxidation in a full-scale wastewater treatment plant revealed by DNA-based stable isotope probing. *Bioresour. Technol.* 256, 152–159. doi: 10.1016/j.biortech.2018.02.012

Pandey, S., Ghosh, P. K., Ghosh, S., De, T. K., and Maiti, T. K. (2013). Role of heavy metal resistant *Ochrobactrum* sp. and *Bacillus* spp. strains in bioremediation of a rice cultivar and their PGPR like activities. *J. Microbiol.* 51, 11–17. doi: 10.1007/s12275-013-2330-7

Pasolli, E., Truong, D. T., Malik, F., Waldron, L., and Segata, N. (2016). Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput. Biol.* 12:e1004977. doi: 10.1371/journal.pcbi.1004977

Plewniak, F., Koechler, S., Navet, B., Dugat-Bony, E., Bouchez, O., Peyret, P., et al. (2013). Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol. Ecol.* 22, 4870–4883. doi: 10.1111/mec.12432

Poulsen, M., Hub, H., Li, C., Chen, Z., Xu, L., Otani, S., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14500–14505. doi: 10.1073/pnas.1319718111

Rashid, M., and Stingl, U. (2015). Contemporary molecular tools in microbial ecology and their application to advancing biotechnology. *Biotechnol. Adv.* 33, 1755–1773. doi: 10.1016/j.biotechadv.2015.09.005

Satinsky, B. M., Gifford, S. M., Crump, B. C., and Moran, M. A. (2013). Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods Enzymol.* 531, 237–250. doi: 10.1016/B978-0-12-407863-5.00012-5

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.* 3, 479–488. doi: 10.1038/nrmicro1159

Serrano, J., and Leiva, E. (2017). Removal of arsenic using acid/metal-tolerant sulfate reducing bacteria: a new approach for bioremediation of high-arsenic acid minewaters. *Water* 9:994. doi: 10.3390/w9120994

Singh, M., Srivastava, P. K., Verma, P. C., Kharwar, R. N., Singh, N., and Tripathi, R. D. (2015). Soil fungi for mycoremediation of arsenic pollution in agriculture soils. *J. Appl. Microbiol.* 119, 1278–1290. doi: 10.1111/jam.12948

Srivastava, P. K., Vaish, A., Dwivedi, S., Chakrabarty, D., Singh, N., and Tripathi, R. D. (2011). Biological removal of arsenic pollution by soil fungi. *Sci. Total Environ.* 409, 2430–2442. doi: 10.1016/j.scitotenv.2011.03.002

Stenuit, B., Eyers, L., Schuler, L., Agathos, S. N., and George, I. (2008). Emerging high-throughput approaches to analyze bioremediation of sites contaminated with hazardous and / or recalcitrant wastes. *Biotechnol. Adv.* 26, 561–575. doi: 10.1016/j.biotechadv.2008.07.004

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M. W., et al. (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 440, 790–794. doi: 10.1038/nature04647

Sun, W., Sierra-alvarez, R., and Field, J. A. (2011). Long term performance of an arsenite-oxidizing-chlorate-reducing microbial consortium in an upflow anaerobic sludge bed (UASB) bioreactor. *Bioresour. Technol.* 102, 5010–5016. doi: 10.1016/j.biortech.2011.01.069

Sun, W., Sierra-Alvarez, R., Hsu, I., Rowlette, P., and Field, J. A. (2010). Anoxic oxidation of arsenite linked to chemolithotrophic denitrification in continuous bioreactors. *Biotechnol. Bioeng.* 105, 909–917. doi: 10.1002/bit.22611

Tardy, V., Casiot, C., Fernandez-Rojo, L., Resongles, E., Desoeuvre, A., Joulian, C., et al. (2018). Temperature and nutrients as drivers of microbially mediated arsenic oxidation and removal from acid mine drainage. *Appl. Microbiol. Biotechnol.* 102, 2413–2424. doi: 10.1007/s00253-017-8716-4

Techtmann, S. M., and Hazen, T. C. (2016). Metagenomic applications in environmental monitoring and bioremediation. *J. Ind. Microbiol. Biotechnol.* 43, 1345–1354. doi: 10.1007/s10295-016-1809-8

Teclu, D., Tivchev, G., Liang, M., and Wallis, M. (2008). Bioremoval of arsenic species from contaminated waters by sulphate-reducing bacteria. *Water Res.* 42, 4885–4893. doi: 10.1016/j.watres.2008.09.010

Teeling, H., and Glöckner, F. O. (2012). Current opportunities and challenges in microbialmetagenome analysis - a bioinformatic perspective. *Brief. Bioinform.* 13, 728–742. doi: 10.1093/bib/bbs039

Thapa Chhetri, R., Suzuki, I., Fujita, T., Takeda, M., Koizumi, J., Fujikawa, Y., et al. (2014). Bacterial diversity in biological filtration system for the simultaneous removal of arsenic, iron and manganese from groundwater. *J. Water Environ. Technol.* 12, 135–149. doi: 10.2965/jwet.2014.135

Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2:3. doi: 10.1186/2042-5783-2-3

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi: 10.1038/nature24621

Tyson, G. W., Lo, I., Baker, B. J., Allen, E. E., Hugenholtz, P., and Banfield, J. F. (2005). Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* 71, 6319–6324. doi: 10.1128/AEM.71.10.6319-6324.2005

Ullrich, S. R., Poehlein, A., Tischler, J. S., González, C., Ossandon, F. J., Daniel, R., et al. (2016). Genome analysis of the biotechnologically relevant acidophilic iron oxidising strain JA12 indicates phylogenetic and metabolic diversity within the novel genus "Ferrovum". *PLoS One* 1:e0146832. doi: 10.1371/journal.pone.0146832

Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., et al. (2017). MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res.* 45, D517–D528. doi: 10.1093/nar/gkw1101

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi: 10.1126/science.1093857

Vogt, C., Dorer, C., Musat, F., and Richnow, H. H. (2016). Multi-element isotope fractionation concepts to characterize the biodegradation of hydrocarbons — from enzymes to the environment. *Curr. Opin. Biotechnol.* 41, 90–98. doi: 10.1016/j.copbio.2016.04.027

Wan, J., Klein, J., Simon, S., Joulian, C., Dictor, M.-C., Deluchat, V., et al. (2010). AsIII oxidation by *Thiomonas arsenivorans* in up-flow fixed-bed reactors coupled to As sequestration onto zero-valent iron-coated sand. *Water Res.* 44, 5098–5108. doi: 10.1016/j.watres.2010.08.044

Wang, Q., Xiong, D., Zhao, P., Yu, X., Tu, B., and Wang, G. (2011). Effect of applying an arsenic resistant and plant growth-promoting rhizobacterium to enhance soil arsenic phytoremediation by *Populus deltoides* LH05-17. *J. Appl. Microbiol.* 111, 1065–1074. doi: 10.1111/j.1365-2672.2011.05142.x

WHO (2011). *Evaluation of Certain Contaminants in Food. Seventy-Second Report of the Joint FAO/ WHO Expert Committee on Food Additives*. WHO Technical Report, No. 959. Rome: World Health Organization.

Yang, L., Li, X., Chu, Z., Ren, Y., and Zhang, J. (2014). Distribution and genetic diversity of the microorganisms in the biofilter for the simultaneous removal of arsenic, iron and manganese from simulated groundwater. *Bioresour. Technol.* 156, 384–388. doi: 10.1016/j.biortech.2014.01.067

Yang, Q., Tu, S., Wang, G., Liao, X., and Yan, X. (2012). Effectiveness of applying of arsenate reducing bacteria to enhance arsenic removal from polluted soils by *Pteris vittata* L. *Int. J. Phytorem.* 14, 89–99. doi: 10.1080/15226510903567471

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.* 29, 415–420. doi: 10.1038/nbt.1823

Yin, X.-X., Chen, J., Qin, J., Sun, G.-X., Rosen, B. P., and Zhu, Y.-G. (2011). Biotransformation and volatilization of arsenic by three photosynthetic cyanobacteria. *Plant Physiol.* 156, 1631–1638. doi: 10.1104/pp.111.178947

Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S. G., and Alvarez-Cohen, L. (2015). High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio* 6:e02288-14. doi: 10.1128/mBio.02288-14

Zuñiga, C., Zaramela, L., and Zengler, K. (2017). Elucidation of complexity and prediction of interactions in microbial communities. *Microb. Biotechnol.* 10, 1500–1522. doi: 10.1111/1751-7915.12855

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Plewniak, Crognale, Rossetti and Bertin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 December 2017
doi: 10.3389/fmicb.2017.02600

Proteomic Analysis of 2,4,6-Trinitrotoluene Degrading Yeast *Yarrowia lipolytica*

Irina V. Khilyas[1], Guenter Lochnit[2] and Olga N. Ilinskaya[1]*

[1] Institute of Fundamental Medicine and Biology, Kazan (Volga Region) Federal University, Kazan, Russia, [2] Protein Analytics, Institute of Biochemistry, Faculty of Medicine, Justus Liebig University Giessen, Giessen, Germany

2,4,6-trinitrotoluene (TNT) is a common component of many explosives. The overproduction and extensive usage of TNT significantly contaminates the environment. TNT accumulates in soils and aquatic ecosystems and can primarily be destroyed by microorganisms. Current work is devoted to investigation of *Yarrowia lipolytica* proteins responsible for TNT transformation through the pathway leading to protonated Meisenheimer complexes and nitrite release. Here, we identified a unique set of upregulated membrane and cytosolic proteins of *Y. lipolytica*, which biosynthesis increased during TNT transformation through TNT-monohydride-Meisenheimer complexes in the first step of TNT degradation, through TNT-dihydride-Meisenheimer complexes in the second step, and the aromatic ring denitration and degradation in the last step. We established that the production of oxidoreductases, namely, NADH flavin oxidoreductases and NAD(P)+-dependent aldehyde dehydrogenases, as well as transferases was enhanced at all stages of the TNT transformation by *Y. lipolytica*. The up-regulation of several stress response proteins (superoxide dismutase, catalase, glutathione peroxidase, and glutathione *S*-transferase) was also detected. The involvement of intracellular nitric oxide dioxygenase in NO formation during nitrite oxidation was shown. Our results present at the first time the full proteome analysis of *Y. lipolytica* yeast, destructor of TNT.

Keywords: 2,4,6-trinitrotoluene, TNT, biodegradation, *Yarrowia lipolytica*, proteomic assay, old yellow enzymes

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Rodrigo Ledesma-Amaro, Imperial College London, United Kingdom
Jing-Sheng Cheng, Tianjin University, China

****Correspondence:***
Irina V. Khilyas
irina.khilyas@gmail.com

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *06 October 2017*
Accepted: *13 December 2017*
Published: *22 December 2017*

Citation:
Khilyas IV, Lochnit G and Ilinskaya ON (2017) Proteomic Analysis of 2,4,6-Trinitrotoluene Degrading Yeast Yarrowia lipolytica. Front. Microbiol. 8:2600. doi: 10.3389/fmicb.2017.02600

INTRODUCTION

Nitroaromatic compounds play an important role in the synthesis of explosives, pharmaceutical compounds, pesticides and herbicides (Boelsterli et al., 2006). Of all nitroaromatic compounds, explosives are of particular concern to the environment (Symons and Bruce, 2006). Military products and waste containing explosives are regularly dumped on land and into water in the process of storage, dismantling, or destruction (Letzel et al., 2003). 2,4,6-trinitrotoluene (TNT) is a common component of many explosives. The overproduction and extensive usage of TNT significantly contaminates the environment. TNT accumulates in soils and aquatic ecosystems and can primarily be destroyed by microorganisms (Mulla et al., 2014). The reactive nitro groups located on an electron-deficient aromatic ring possess electron-attractive potential that can influence the stability of TNT. The TNT transformation mechanism involves a two-electron reduction of nitro groups which leads to the formation of hydroxylamino- (HADNTs) and amino-derivatives (ADNTs) by nitroreductases under aerobic conditions (Caballero et al., 2005). The

resulting HADNTs and ADNTs derivatives possess some negative properties, such as high toxicity, mutagenicity, and carcinogenicity (Spanggord et al., 1995).

The microbial degradation of TNT through the formation of TNT-Meisenheimer complexes is a unique path leading to the release of nitrogroups and conditioning the enzymatic cleavage of the aromatic ring, which is followed by its complete degradation (Ziganshin and Gerlach, 2014). *Enterobacter cloacae* PB2, *Pseudomonas fluorescence* I-C, and *Pseudomonas putida* JLR11 are the aerobic bacteria that hydrogenate the TNT aromatic ring and trigger the release of NO_2^- by the family of old yellow enzymes (OYEs) (Pak et al., 2000; Wittich et al., 2008). OYEs and their homologs can be found in prokaryotic and eukaryotic organisms (Singh, 2014). Previously, the OYE homologs were detected in higher plants (Strassner et al., 1999). Specifically, TNT was found to activate the process of nitroreductases in the root cell cytosol of soybean, which reduced the nitro group of the obtained TNT products (Adamia et al., 2006). It was established that the detoxification strategy of plants confined to the conjugation and compartmentalization of xenobiotic compounds in the vacuole or cell wall, resulting in the formation of non-extractable compounds (Sens et al., 1999). However, no direct evidence for OYEs participation in the TNT transformation by micromycetes has been provided so far. Although the OYEs homologs can be found in different species of fungi, only *Irpex lacteus* accumulates TNT-hydride-Meisenheimer complexes (Nizam et al., 2014). It is known that OYE members are involved in the oxidative stress response (Farah and Amberg, 2007). A strong correlation between the participation of OYEs homologs in the transformation of nitroaromatic compounds and detoxification of reactive oxygen species (ROS) was reported in previous studies (Fitzpatrick et al., 2003). In particular, *Bacillus subtilis* protein (YqjM) related to OYE homologs was found to reduce nitroesters and nitroaromatic compounds and to be induced by hydrogen peroxide (Fitzpatrick et al., 2003).

The study of yeast proteins that trigger the transformation of nitroaromatic compounds contributes to a better understanding of a wide range of processes attributed to biotransformation of pharmacological, agricultural and industrial toxic compounds. In this paper, we present the results of an investigation into the proteomic profiling of TNT-degrading aerobic yeast *Yarrowia lipolytica* VKPM Y-3492. In our study, we identified a unique set of upregulated membrane and cytosolic proteins of *Y. lipolytica*, which biosynthesis increased during TNT transformation through TNT-monohydride-Meisenheimer complexes in the first step of TNT degradation, through TNT-dihydride-Meisenheimer complexes in the second step, and the aromatic ring denitration and degradation in the last step. We established that the production of oxidoreductases, namely, NADH flavin oxidoreductases and NAD(P)+-dependent aldehyde dehydrogenases, as well as transferases was enhanced at all stages of the TNT transformation by *Y. lipolytica*. Simultaneously, we detected the upregulation of several stress response proteins [superoxide dismutase (SOD), catalase, glutathione peroxidase, and glutathione *S*-transferase (GST)]. Finally, we established upregulation of yeast intracellular nitric oxide dioxygenase (NOD) participating in NO release. Finally, we established the involvement of intracellular NOD in NO formation during nitrite oxidation. This finding supports the biogenic path of NO formation in addition to abiotic generation illustrated in previous studies (Ziganshin et al., 2010; Khilyas et al., 2013).

MATERIALS AND METHODS

Chemicals

TNT (purity, 99%) was purchased from ChemService (West Chester, PA, United States).

Yeast Strain and Culture Conditions

Yarrowia lipolytica was grown aerobically at 30°C for 1 day on Sabouraud agar medium containing (per liter) glucose 10 g, peptone 10 g, yeast extract 5 g, NaCl 0.25 g, and agar 20 g. Yeast cells were harvested, washed with 16 mM phosphate buffer (pH 7.0) and added into 250 mL Erlenmeyer flasks containing 50 mL of synthetic medium. The synthetic medium was composed of 28 mM glucose, 7.6 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$ and buffered with 16 mM K-Na-phosphate buffer to pH 7.0. The initial cell concentration was adjusted to an optical density (600 nm, A_{600}) of 1.0, and growth was measured using a Lambda 35 spectrophotometer (Perkin Elmer) with cell-free (filtered) culture medium as reference. TNT was added as ethanol stock solution to a final concentration 440 μM, and flasks were incubated at 30°C with shaking speed 150 rpm. TNT-free control experiments contained pure ethanol (1.33 mL absolute ethanol into 50 mL medium). Samples of yeast cells were taken during culture growth at time points corresponding maximum formation of analytically measured TNT metabolites. After centrifugation, yeast biomass was washed with 16 mM phosphate buffer twice, frozen in liquid nitrogen, lyophilized, and stored at −80°C. These samples were used for proteins analysis. All experiments were set up in two biological repetitions. To exclude proteins of the first step of TNT transformation, Boolean algorithm for binary data sets of TNT-treated and TNT-untreated yeast proteins at pH 4.6 and pH 6.6 was applied.

Analytical Methods

TNT and its biotransformation products were detected with a HPLC (Thermo Scientific™ Dionex™ UltiMate™ 3000) equipped with autoinjector, diode array detector, the column oven, and Supelcosil (C-8) column (150 by 4.6 mm; particle size, 5 μm), as described previously (Borch and Gerlach, 2004; Khilyas et al., 2013). Nitrates and nitrites were detected using ion chromatograph 761 Compact IC (Metrohm, Sweden) equipped with a Metrosep A SUPP 5 column (150/4.0; particle size, 5 μm).

Proteome Analysis

TE-buffer (EDTA, pH 8.0, Tris base) containing 1% Protease Inhibitor Mix HP (Serva, Heidelberg, Germany) was added to the yeas biomass. Yeast cells were disrupted by glass beads using a Potter-Elvehejem homogenizer at 4°C (Wheaton, Millville, NJ, United States). After centrifugation (14000 *g*

for 30 min) supernatant containing cytoplasmic proteins was collected. The pellet including membrane-associated proteins was resuspended in an ice-cold solution (8 M urea and 2 M thiourea). Delipidation of supernatant and pellet was performed by methanol-chloroform method (Wessel and Flügge, 1984). Protein quantification was carried out by 2-D Quant kit (GE Healthcare, United States). Cytosolic or membrane-associated proteins (150 μg) were subjected to gel strips (pH 7–11 NL, 13 cm) in sample buffer [4% CHAPS, 30 mM DTT, 20 mM Tris-base, 2% IPG buffer, 1% bromophenol blue (added immediately before use)] and maintained 24 h at room temperature for rehydration. Then, strips were placed in a horizontal running tray for IEF and covered with mineral oil to prevent dehydration during electrophoresis (Wenge et al., 2008). Electrophoresis was performed at 0–100 V (1 mA) for 5 h, 100–3500 V (1 mA) for 6 h and 3500 V (1 mA) for 6 h at 20°C. Strips were washed with 1% DTT for 15 min, with 4% iodoacetamide for 15 min and three times with 50 mM Tris buffer (pH 6.8). Equilibrated strips were transferred on 12.5% polyacrylamide SDS gel (18 cm × 20 cm) using tweezers. Electrophoresis conditions per one gel were 15 mA (600 V, 50 W) for 15 min and 110 mA (600 V, 50 W) for 6 h. Densitometric analyses were done with the PDQuest software (Bio-Rad)after Coomassie-staining.

Tryptic in-Gel Digestion of Proteins

Coomassie-stained protein spots (n = 404) were excised with the ExQuest Spot Cutter (BioRad) and transferred into 96-wells plates. Digestion of protein was done using liquid handling system (Microstarlet, Hamiltonrobotics, Martinsried, Germany). Gel plugs were washed 10 min with 150 μl 50% acetonitrile (ACN), dehydrated 2 min with 150 μl 100% ACN, swelled 5 min with 150 μl 50 mM NH_4HCO_3, dehydrated 2 min with 150 μl 100% ACN. The proteins were digested for 120 min at 45°C in 25 mM NH_4HCO_3 containing 0.5 mg trypsin (sequencing grade, Promega, Mannheim, Germany). Peptides were stabilized by 10 μL 0.1% TFA and stored at 21°C until use (Wenge et al., 2008).

MALDI-TOF MS

MALDI-TOF MS was performed using an UltraflexI TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a nitrogen laser and a LIFT-MS/MS facility. The instrument was operated in the positive-ion reflectron mode using 2,5-dihydroxybenzoic acid (Sigma) in 50% ACN/1% phosphoric acid as matrix. Acquired sum spectra consist of 200–400 single spectra. For data processing and instrument control the Compass 1.1 software package consisting of FlexControl 2.4, FlexAnalysis 2.4, and ProteinScape 3.0 was used.

Database Search

Proteins were identified by MASCOT peptide mass-fingerprint search[1] using the MSDB database. The search was restricted to *Yarrowia* with a mass tolerance of 100 ppm and carbamidomethylation of cysteine as global modification and oxidation of methionine as variable modification.

[1]http://www.matrixscience.com

RESULTS

Three Steps of TNT Transformation by *Y. lipolytica*

After 3 h of cultivation in TNT-free medium, yeast cells were grown to 1.7 units (cell density of 600 nm), whereas in TNT medium cells reached the density of 1.3 units after 5 h of cultivation only (Supplementary Figure S1A). In view of the fact that the biosynthesis for a spectrum of synthesized proteins depends on pH, we sampled the yeast biomass at different time points (3 and 5 h), but at equal levels of pH = 6.5. HPLC data showed that the TNT transformation through aromatic ring reduction led to a maximum accumulation of TNT-monohydride complexes (H^-—TNT) after 5 h of yeast cultivation. The maximum concentrations 3-H^--TNT and 1-H^--TNT were 289 and 17 μM, accordingly (Supplementary Figure S1B). At the same time, the TNT transformation via nitro group reduction resulted in the accumulation of 26 μM of 2-hydroxylamino-4,6-dinitrotoluene (2-HADNT), 46 μM of 4-hydroxylamino-2,6-dinitrotoluene (4-HADNT) and 43 μM of nitrites (Supplementary Figure S1C). Thus, the yeast cells grown for 3 h without TNT and the cells grown for 5 h with TNT were used for protein analysis. We relate these time point to the first step of TNT transformation, which leads to the accumulation of TNT-monohydride complexes.

After 12 h of incubation, the pH of culture medium decreased to 4.6. Under these conditions, the TNT-monohydride complexes transformed into TNT-dihydride complexes (3,5-2H^--TNT·H^+), and the concentration of 2-HADNT and 4-HADNT increased to 49 and 96 μM, respectively. The appearance of 2,4-DNT (11 μM) during 3,5-2H^--TNT·H^+ destruction was associated with a release of 133 μM NO_2^-, which was partially oxidized to NO_3^- and accompanied by NO formation as it was shown previously (Ziganshin et al., 2010; Khilyas et al., 2013). The time-point of 12 h was chosen for the second step of protein analysis.

The continued cultivation of *Y. lipolytica* for 24 h resulted in a pH shift to a strong acidic range (pH 3.1-3.6) and the intensive destruction of TNT-mono- and dihydride complexes. However, HADNTs, 4-ADNT, and 2,4-DNT, as well as NO_3^-. As a product of NO_2^- oxidation could be still found in growth medium (Supplementary Figures S1B,C). This time point (24 h) was chosen for the last step of protein analysis.

Twenty Proteins of *Y. lipolytica* Are Up-Regulated at Maximal TNT-Monohydride Complexes Formation

Next, the membrane and cytosolic proteins were analyzed to determine the quantitative and qualitative changes in the protein profiling of *Y. lipolytica* based on the maximum accumulation of TNT-monohydride complexes after 5 h of cultivation. Approximately, 1300–2000 protein spots were detected on each gel: eight gels correspond to TNT-free cultivation (four for membrane proteins and four for cytosolic proteins) and eight gels were obtained based on the cultivation with TNT. For twenty

revealed protein spots, the difference ($p < 0.05$) between TNT-treated and TNT-untreated yeast was more than fivefold. All these proteins were up-regulated in the presence of TNT. The influence of pH alteration in medium was excluded (Supplementary Table S1).

Figure 1A shows the intracellular distribution of 20 up-regulated proteins in the compartments of yeasts and their role in the cellular processes. A significant number of the up-regulated proteins could be assigned to cytosolic (45% of the total proteins) and mitochondrial (35% of the total proteins) compartments of yeast cells; 15% of all up-regulated proteins were related to plasma membrane proteins. Five percent of other proteins were associated with proteasomes, endoplasmic reticulum, and microsomes, respectively. About 5% of proteins could not be identified. More detailed information is presented in Supplementary Table S1. The cytosolic and membrane proteins from *Y. lipolytica* were involved in the metabolic process (40%), redox process (23%), electron-proton transport (7%), ATP synthesis (7%), biosynthetic process (4%), catabolic process (4%), fatty acid metabolic process (3%), RNA splicing (3%), propanoate metabolism (3%), and glutathione metabolism (3%). The functions performed by 3% of proteins remained unknown (**Figure 1B**).

In general, TNT primarily induced the biosynthesis of oxidoreductases and transferases of *Y. lipolytica*. Three membrane isoforms of NADH flavin oxidoreductases/NADH oxidases demonstrated a higher up-regulation (by 16–60 times) (**Figure 1C**). Five different isoforms of retinal dehydrogenases 2, belonging to the NAD(P)+-dependent aldehyde dehydrogenase superfamily (ALDH-SF), also increased by eight times. GST and disulfide isomerase/thioredoxin showed a 15- and 10-fold increase, respectively (**Figure 1C**). Thus, the first step of TNT transformation was associated with an increased level of yeast proteins catalyzing metabolic redox reactions.

Proteins of *Y. lipolytica* Are Up-Regulated at Maximal TNT-Dihydride Complexes Formation

The second step of TNT transformation was characterized by the accumulation of TNT-dihydride complexes. In total, 102 up-regulated proteins in cytosolic and membrane fractions were revealed, that is, about five times higher compared with the first step of transformation (Supplementary Table S2). In addition, the number of mitochondrial proteins grew from 35 to 48% compared with the first step. At the same time, the number of cytosolic proteins insignificantly reduced (41.2%) (**Figures 1A**, **2A**). Others proteins were localized in peroxisomes (4.9%), microsomes and cell wall (3.9% each), plasma membrane, endoplasmic reticulum and vacuoles (2% of each), proteasomes and nucleus (1% of each), while the localization of 4.9% of the proteins remained unknown (**Figure 2A**).

Figure 2B illustrates the functional classification of up-regulated intracellular *Y. lipolytica* proteins which can be attributed to the accumulation of TNT-dihydride complexes at pH 4.6. The identified proteins participated in metabolic (29%) and redox (17%) processes, and the tricarboxylic acid (TCA) cycle (10%), as well as in biosynthetic (9%) and catabolic (6%) processes (**Figure 2B** and Supplementary Table S2). In the second step, we also obtained the insignificant amount of proteins that were not up-regulated during the first stage of TNT transformation. These proteins participated in protein biogenesis (5%), biogenesis of RNA (5%), ATP synthesis (2%), glycolysis (2%), gluconeogenesis (2%), glyoxylate cycle (2%), propanoate metabolism (1%), electron-proton transport (1%), and other processes (8%) (**Figure 2B** and Supplementary Table S2).

Interestingly, the level of NADH flavin oxidoreductases/NADH oxidases was much lower than in the first step of TNT transformation (corresponding to a sixfold increase only) (**Figure 2C**). Six isoforms of retinal dehydrogenase 2 were up-regulated 45 times. At this stage only, two isoforms of acetaldehyde dehydrogenases were up-regulated. The isoform 2 of NADH flavin oxidoreductase was not detected in the TNT-free proteome at both stages of TNT transformation, whereas the NADPH:quinone reductase was for the first time detected at the second stage (**Figure 2C**). The GST and disulfide isomerase/thioredoxin were up-regulated, although the level of up-regulation was lower compared with that achieved in the first step of TNT transformation (by 4.5 times only) (**Figure 2C**). Additionally, the level of glutathione transferase protein showed a 33-fold increase (**Figure 2C**). Finally, we established an eightfold up-regulation of thioredoxin reductase (**Figure 2C**).

After 12 h of TNT transformation, we observed an increase in the level of proteins involved in the regulation of RNA biogenesis (transcription and translation elongation) and protein biogenesis (protein folding, modification, and proteolysis) (Supplementary Table S2). Several TNT-induced heat shock proteins (HSP), including HSP 60 (GroEL), HSP78 and piso0_004415 (ATPases associated with a wide variety of cellular activities), ATP-dependent molecular chaperone HSC82, and chaperone protein (DnaK) were identified (Supplementary Table S2). Furthermore, saccharopepsin and carboxypeptidase C, which have a specific location in vacuoles and perform multiple beneficial functions inside the yeast cells, such as proteolysis, were identified (Supplementary Table S2).

The amount of unknown and hypothetical proteins activated at the second stage of TNT-transformation was significantly increased compared with the stage of accumulating the TNT-monohydride complexes (Supplementary Tables S1, S2). Thus, the second step of TNT transformation was characterized by the sustained increase in yeast proteins participating in metabolic redox reactions, proteolysis, and stress-response.

Proteins of *Y. lipolytica* Are Up-Regulated at Stage of TNT-Hydride Complexes Destruction

The third step of TNT transformation was characterized by the destruction of TNT-dihydride complexes. Specifically, we identified 31 up-regulated proteins in the cytosolic and membrane fractions. **Figure 3A** shows the intracellular distribution of up-regulated proteins. A significant number of the up-regulated proteins were localized in the cytosolic (41.9%) and mitochondrial (29%) compartments of yeast cells

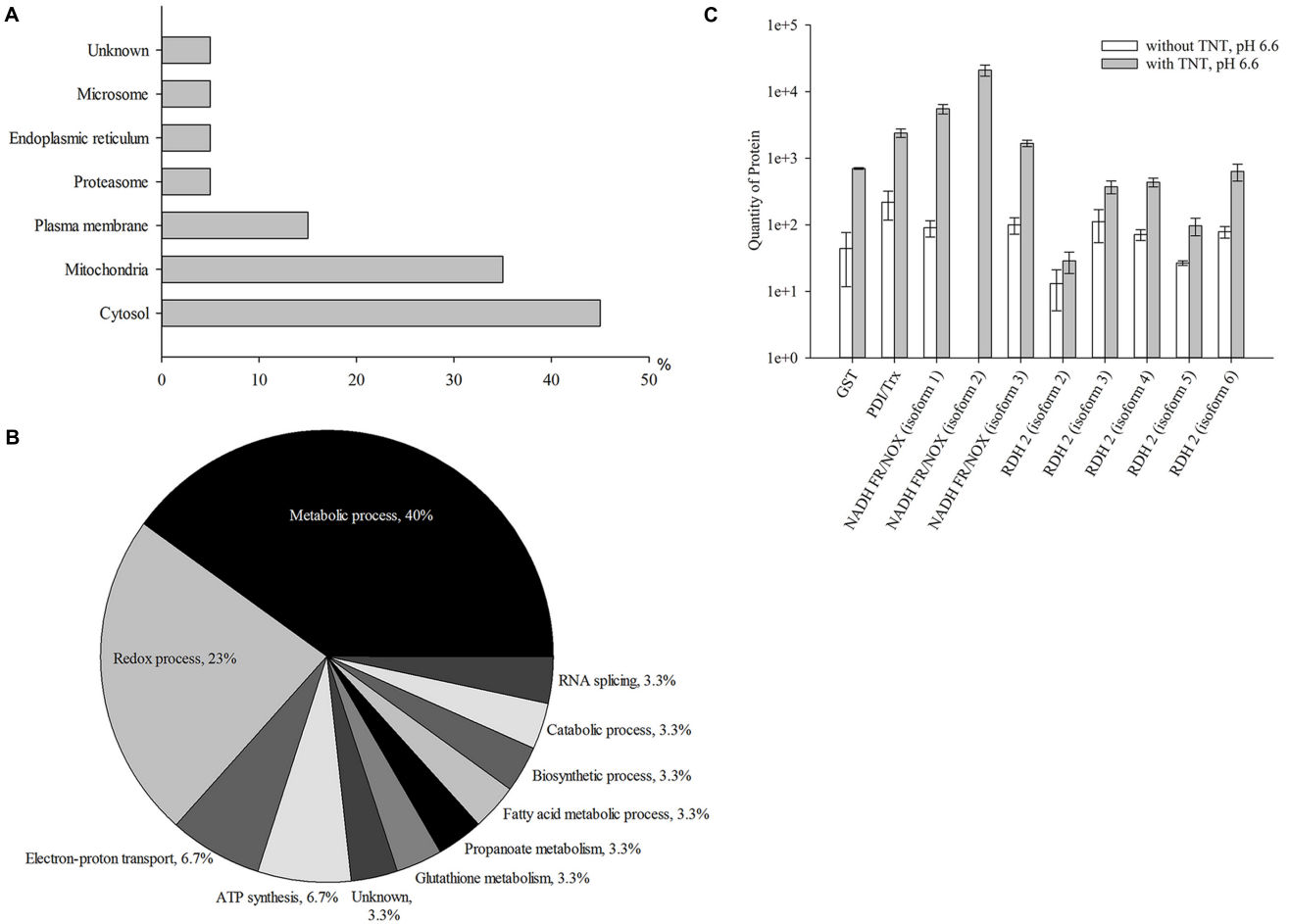


FIGURE 1 | Up-regulated proteins at the first stage of TNT transformation by *Yarrowia lipolytica*: their intracellular localization **(A)**; participation in metabolic processes **(B)** and the quantity of major proteins maximal up-regulated by TNT **(C)**. **(A,B)** For 100% all identified proteins at the first stage of TNT transformation were taken. **(C)** Error bars represent the standard deviation of duplicate biological experiments with two technical repetitions. GST, glutathione *S*-transferase; PDI/Trx, protein disulfide isomerase/thioredoxin; NADH FR/NOX, NADH flavin oxidoreductases/NADH oxidases; RDH, retinal dehydrogenase 2.

(**Figure 3A**). Other proteins were localized in microsomes (12.9%), plasma membrane (9.7%) and vacuoles (9.7%), peroxisome (6.5%), the perinuclear region (3.2%), lipid droplet (3.2%), eisosomes (3.2%) and glyoxysomes (3.2%) (**Figure 3A**).

Figure 3B illustrates the functional classification of up-regulated intracellular *Y. lipolytica* proteins that can be attributed to the destruction of TNT-dihydride complexes at pH 3.6. The identified proteins participated in metabolic (11%) and redox (29%) processes, TCA cycle, ATP synthesis, ionic transfer and proteolysis (8.6% of each), vesicle trafficking (5.7%) and cell wall biogenesis (5.7%) (**Figure 3B** and Supplementary Table S3).

The last stage of TNT transformation by *Y. lipolytica* led to the inhibition of metabolic processes and the activation of redox processes (**Figure 3B**). The average amount of proteins participating in energetic processes insignificantly decreased compared with the first and the second stages of TNT transformation, although the part of cellular processes occurring in plasma membrane, microsomes and vacuoles was higher (**Figures 1A**, **2A**, **3A**).

Figure 3C illustrates the induction of six retinal dehydrogenases 2, NADPH quinone reductase, and two isoforms of the NAD(P)+-dependent aldehyde dehydrogenase. The isoform 2 of retinal dehydrogenase 2 particularly increased in the third step of TNT transformation.

Furthermore, an increase in the level of sphingolipid long chain base-responsive protein located in eisosome nearly to plasma membrane was detected (Supplementary Table S3). In addition, we observed the overexpression of the potassium channel subunit, which was involved in secretory and endocytic vesicular trafficking pathways. Among HSPs, only HSP 60, or GroEL was found to be up-regulated (Supplementary Table S3).

The yeast cells that adapted to the TNT-induced stress reached an optical density similar to the one that could be observed for the TNT-free system at the end of the third stage of TNT transformation (Supplementary Figure S1A). Yeast cells performed the destruction of TNT-dihydride complexes by NADH flavin oxidoreductases/NADH oxidases, whereas a

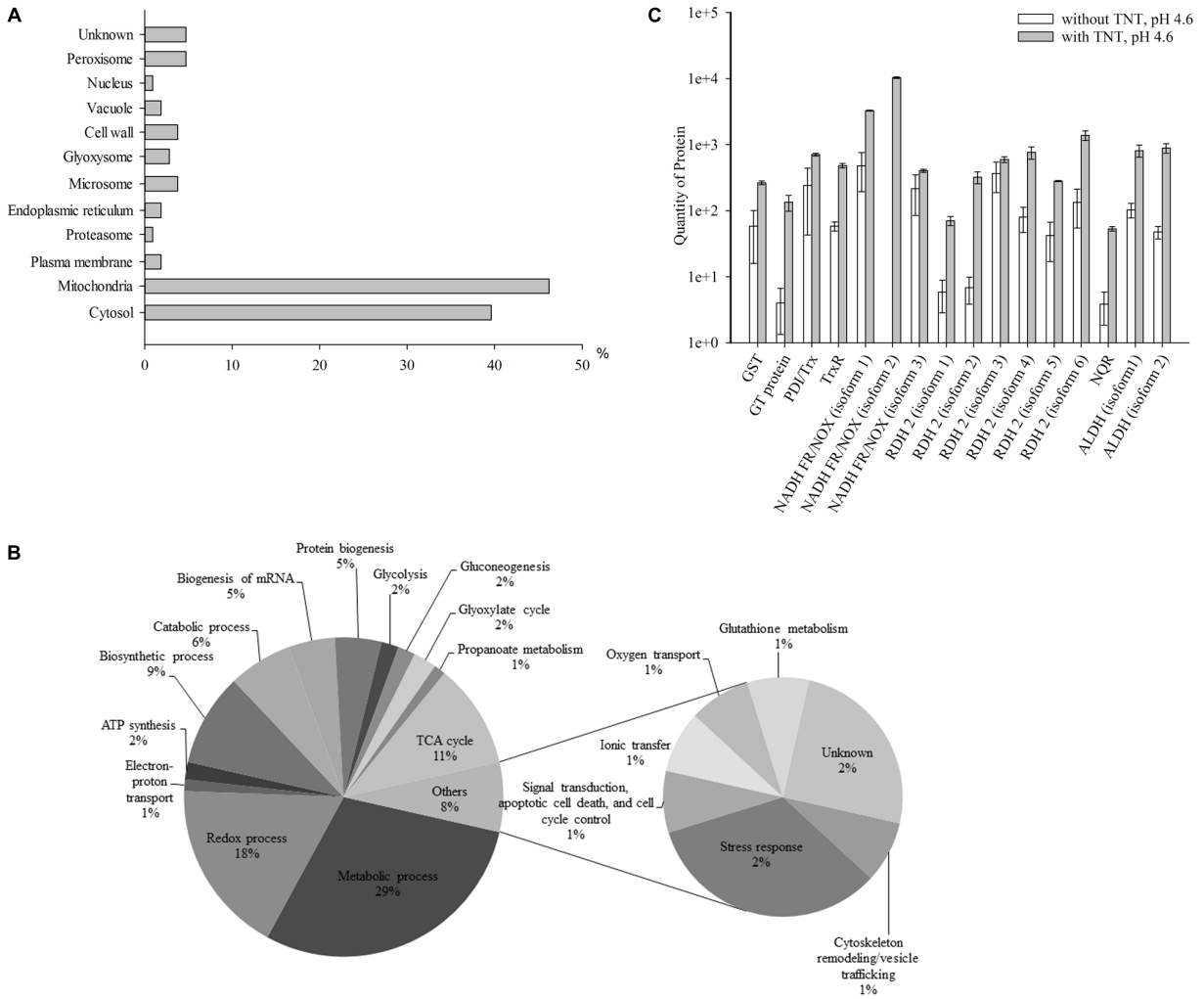


FIGURE 2 | Up-regulated proteins at the second stage of TNT transformation by *Y. lipolytica*: their intracellular localization **(A)**; participation in metabolic processes **(B)** and the quantity of major proteins maximal up-regulated by TNT transformation products. **(A,B)** For 100% all identified proteins at the second stage of TNT transformation were taken. **(C)** Error bars represent the standard deviation of duplicate biological experiments with two technical repetitions. GST, glutathione *S*-transferase; GT, glutathione transferase protein; PDI/Trx, protein disulfide isomerase/thioredoxin; TrxR, thioredoxin reductase; NADH FR/NOX, NADH flavin oxidoreductases/NADH oxidases; RDH, retinal dehydrogenase 2; NQR, NADPH:quinone oxidoreductase; ALDH, aldehyde dehydrogenases.

decrease in HADNTs and ADNTs was entailed by proteins of the NAD(P)+-dependent aldehyde dehydrogenase superfamily (ALDH-SF) (**Figures 1C**, **2C**, **3C**). Three isoforms of NADH flavin oxidoreductases/NADH oxidases, as well as disulfide isomerase/thioredoxin, thioredoxin reductase, and peroxiredoxin were found to be up-regulated at all stages of TNT transformation (**Figures 1C**, **2C**, **3C**). Generally, the predominant part of the up-regulated enzymes belonged to oxidoreductases. This fact points to the role of such enzymes in TNT transformation.

Identification of Proteins Involved in Detoxification Processes of TNT Biotransformation

A significant part of the up-regulated enzymes besides NADH flavin oxidoreductases/NADH oxidases was related to transferases (**Figure 4**). An increase in the level of these enzymes reflects the activation of the second phase of xenobiotics detoxification. Active in the first phase of TNT transformation, oxidoreductases led to the formation of TNT metabolites, namely, HADNTs and ADNTs, which could not be fully utilized by *Y. lipolytica* (Ziganshin et al., 2007). Therefore, the conjugation of these compounds with cellular substrates such as glutathione was necessary to excrete the toxic substances. Indeed, we could observe a high level of transferases during the entire cycle of TNT transformation (**Figure 4**).

The TNT transformation process through the aromatic ring reduction and reduction of nitro groups by *Y. lipolytica* was closely associated with the generation of ROS. Catalase and SOD could be found at the up-regulated level during all stages of TNT transformation (**Figure 4**). It is important to note that the high level of catalase and SOD was observed in TNT-untreated

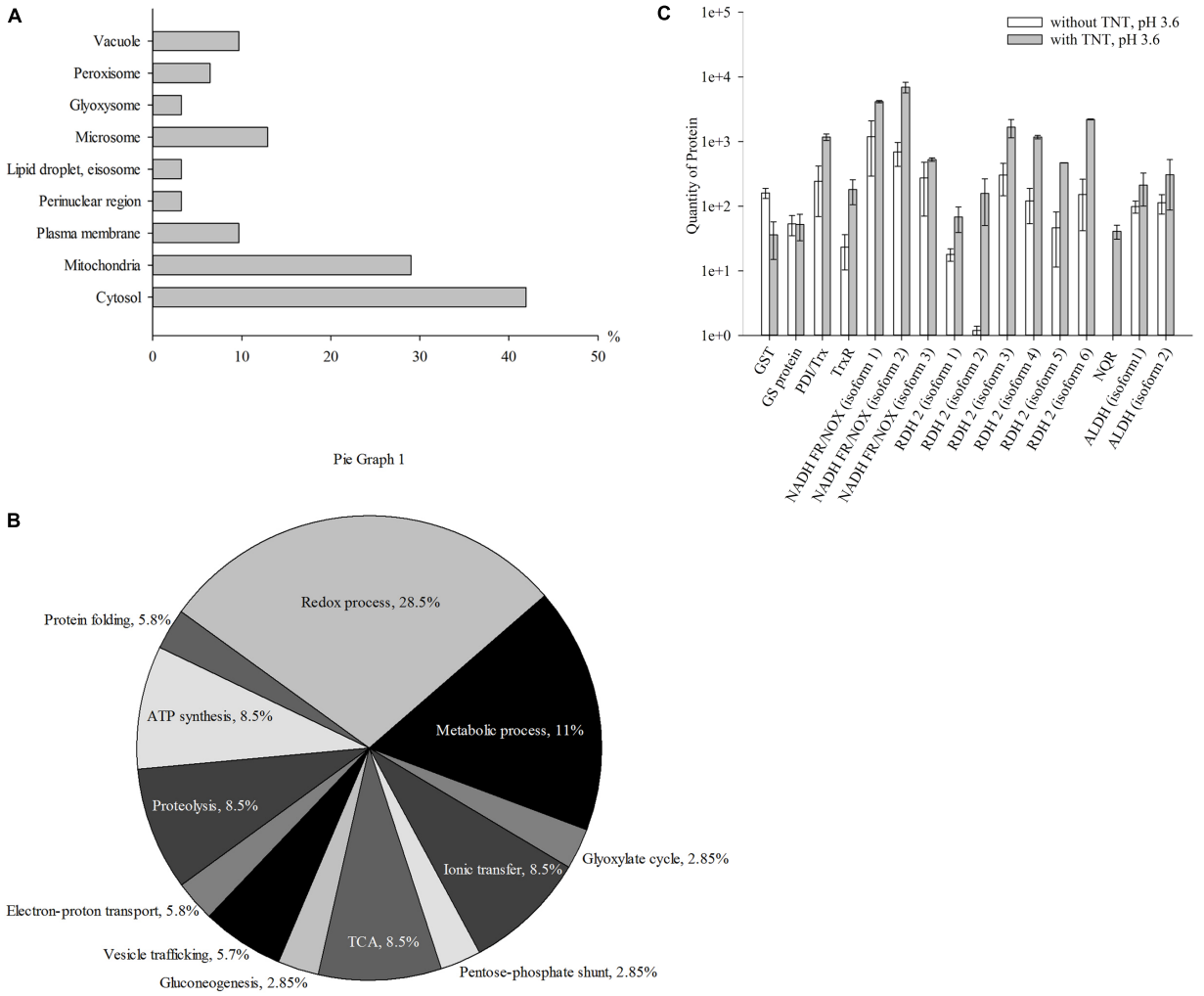


FIGURE 3 | Up-regulated proteins at the third stage of TNT transformation by *Y. lipolytica*: their intracellular localization **(A)**; participation in metabolic processes **(B)** and the quantity of major proteins maximal up-regulated by TNT transformation products. **(A,B)** For 100% all identified proteins at the third stage of TNT transformation were taken. **(C)** Error bars represent the standard deviation of duplicate biological experiments with two technical repetitions. GST, glutathione *S*-transferase; GT, glutathione transferase protein; PDI/Trx, protein disulfide isomerase/thioredoxin; NADH FR/NOX, NADH flavin oxidoreductases/NADH oxidases; RDH, retinal dehydrogenase 2; ALDH, aldehyde dehydrogenases.

cells, which is characteristic of yeast cell growth under aerobic conditions.

Previously, we demonstrated the extracellular generation of nitric oxide (NO) during TNT transformation by *Y. lipolytica* using ESR spectroscopy (Khilyas et al., 2013). ESR spectra confirmed that NO_2^- was converted into NO and NO_3^- following a decrease in pH below 4.5, which coincides with the biotransformation of $3H^-$-TNT to $3,5\text{-}2H^-$-TNT·H^+ (Khilyas et al., 2013). In this study, we observed a unique induction of intracellular NOD of *Y. lipolytica* at the second and third stage of TNT transformation (**Figure 4**). NOD catalyzed the reaction of superoxide anion detoxification with the production of less toxic NO_3^- and was induced simultaneously to NO generation. However, minor amounts of NOD were detected in the TNT-untreated cells when pH of the medium reduced to 3.6 (**Figure 4**). The cellular generation of NO in the absence of TNT could be associated with NO-synthase activity which catalyzed the reaction between L-arginine, molecular oxygen and NAD(P)H (Xia et al., 2000; Ignarro et al., 2002).

DISCUSSION

TNT is a nitroaromatic agent which is highly adapted for military needs as it can spread and persist in soils, surface, and groundwater (Rylott et al., 2011). Members of the protein families involved in TNT biotransformation are quite known and well-characterized. Microorganisms use several metabolic pathways of TNT transformation. Aerobic bacteria initiate HADNTs and ADNTs formation by nitroreductases and PETN reductase, although PETN reductase, XenB reductase

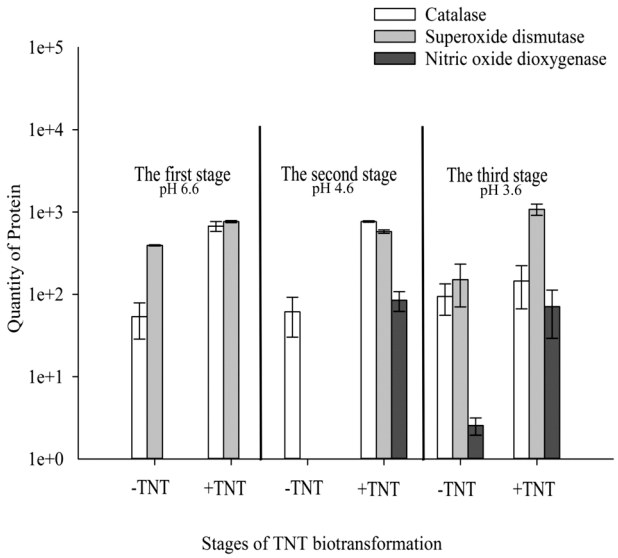


FIGURE 4 | Quantitative changes of up-regulated intracellular proteins of *Y. lipolytica* involved in detoxification of reactive oxygen and nitrogen species during all stages of TNT biotransformation. Error bars represent the standard deviation of duplicate biological experiments with two technical repetitions.

and members of the OYE family participate as well in the production of a Meisenheimer complex with following TNT denitration (French et al., 1998; Pak et al., 2000; Oh et al., 2001; Khan et al., 2002; Caballero et al., 2005; Iman et al., 2017). We confirm by proteomic analysis the upregulation of several proteins in *Y. lipolytica* treated by TNT. Three steps of TNT transformation by yeasts led to upregulation of NADH flavin oxidoreductases/NADH oxidases isoforms. Protein-disulfide isomerase was detected not only in yeast TNT-treated proteome, but also in *Stenotrophomonas maltophilia* (Lee et al., 2009). It could be noted, that the part of expressed proteins has not any relation to degradation enzymes. The pool of heat-shock proteins of *Y. lipolytica* (see Supplementary Material), *Stenotrophomonas* sp. (Ho et al., 2004), *Pseudomonas* sp. (Lee et al., 2008) and alginate-producing enzymes of *Pseudomonas* sp. growing on the reach medium with TNT (Cho et al., 2009) has been detected. However, no comprehensive metaproteome analysis of non-conventional yeasts based on an in-depth investigation of key enzymes which are responsible for the formation of TNT metabolite has been conducted to date.

Pathways of nitroaromatic compound biotransformation through enzymatic oxidative and reductive reactions are widely distributed among aerobic and anaerobic bacteria, fungi, plants, animals, and humans (Williams and Bruce, 2002; Stenuit and Agathos, 2010; Claus, 2014). The oxygenolytic mechanisms of TNT transformation could not be observed in living organisms due the chemical structure of such molecules (Esteve-Nunez et al., 2001). Hence, the microorganisms used different reduction pathways of TNT biotransformation preferentially leading to hydroxylamino and amino metabolites (Williams et al., 2004). Several strains carried out the reduction of aromatic ring with the formation of hydride Meisenheimer complexes, whereas some groups harbored both pathways (Smets et al., 2007).

Earlier, it was shown that *Y. lipolytica* formed TNT hydride-Meisenheimer complexes during the first 6 h of cultivation (Zaripov et al., 2004; Ziganshin et al., 2007). This points to the fact that the most energetic equivalents are used in reductive reactions of yeast in the beginning of its growth. Thus, the enzymes induced by TNT in the initial step of transformation participate in energy-dependent processes. Furthermore, we identified the upregulation of NADH flavin oxidoreductases/NADH oxidases and GST (**Figures 1–3**). Oxidoreductases localized in the plasma membrane of yeast cells are the pioneering proteins in TNT transformation via TNT monohydride-Meisenheimer complexes (**Figures 1–3**, **5** and Supplementary Table S1). Different molecular masses of NADH flavin oxidoreductases/NADH oxidases indicate that the enzyme exists in several isoforms (Supplementary Table S1). The 45% similarity of NADH flavin oxidoreductases/NADH oxidases with PETN reductase of *E. cloacae* PB2 suggests a reaction mechanism via TNT-hydride complexes formation by *Y. lipolytica* occurs by the transfer of hydride from the enzyme to the aromatic ring (French et al., 1998; Khan et al., 2002).

Further, the spectrum of up-regulated oxidoreductases expanded during the formation of dihydride-Meisenheimer complexes. Similarly to retinal dehydrogenase 2, acetaldehyde dehydrogenases participated in $2e^-$ reductions processes. NADPH quinone reductase catalyzing the two-electron reduction of quinones and nitroaromatic compounds is up-regulated at the stage of maximum accumulation of TNT-dihydride complexes (**Figures 2C**, **5**). In addition, the second stage of TNT transformation was characterized by the up-regulation of glutathione transferase protein (**Figures 2C**, **5**).

The complete destruction of TNT-dihydride complexes occurred in the stationary phase of *Y. lipolytica* growth, which corresponds to the third stage of TNT transformation, and hydroxylamino- and amino-dinitrotoluenes became predominant metabolites (Ziganshin et al., 2010; Khilyas et al., 2013). All enzymes, which were up-regulated at the second stage of TNT transformation, retained a high level of biosynthesis at the third stage. Therefore, it could be concluded that the up-regulated enzymes identified in this study participate in the formation of all types of TNT intermediates.

A broad range of key metabolic enzymes are up-regulated during TNT transformation (**Figures 1–3**). It is known that toxic effects of TNT are partially attributed to reactive oxygen radicals (ROS) generated through the formation of nitro anion radicals (Spain, 1995; Kumagai et al., 2004). Furthermore, the interaction of NO and superoxide anion $O_2^{\cdot -}$ generates peroxynitrite with a more disruptive power (Kumagai et al., 2004). This reaction is catalyzed by neuronal nitric oxide synthase. Peroxinitrite is a stronger oxidant than both NO and $O_2^{\cdot -}$ and it might increase the oxidative stress through its binding with biomolecules (Beckman et al., 1990; Beckman and Koppenol, 1996; Vasquez-Vivar et al., 1997).

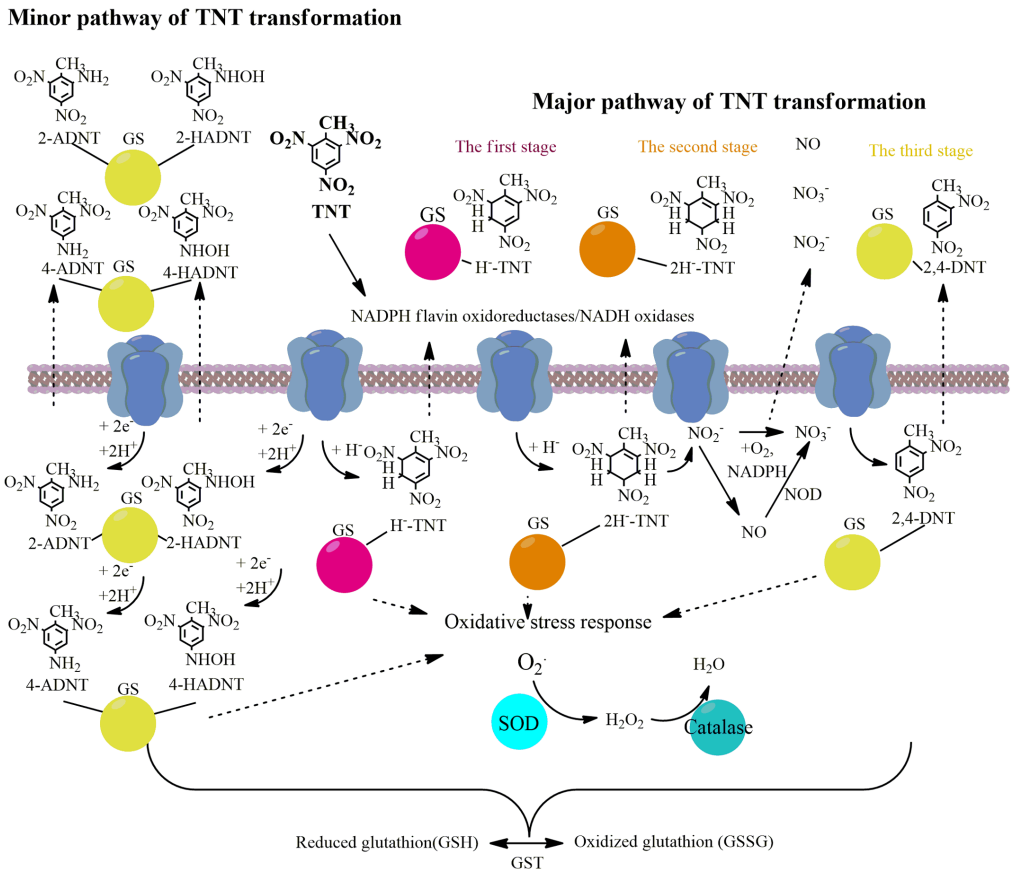


FIGURE 5 | Schema illustrating the main pathways and key upregulated enzymes of TNT transformation by *Y. lipolytica*.

Reactive oxygen species and reactive nitrogen species (RNS) could be neutralized by three cellular mechanisms: (a) by enzymes as SOD, catalase, glutathione peroxidase, and NO dioxigenase (NOD); (b) by transition metals such as Fe II, Cu II, Mn II; and (c) by antioxidant-scavengers such as ascorbic acid, cysteine, and SH groups of plasma proteins (Ignarro et al., 2002; Yao et al., 2004). In this study, we established the defense properties of enzymatic yeast during TNT transformation. We have shown that SOD and catalase were up-regulated at all stages of TNT biotransformation. Recently, it was established that TNT is an inductor of antioxidant system in human, mouse and rat hepatoma cell lines (Naumenko et al., 2016). Of these two, the upregulated level of SOD was higher (**Figure 4**). Although the main part of nitrate occurred in yeast growth medium through the abiotic nitrite oxidation at acidic pH, the up-regulation of NOD at the second stage of TNT transformation shows that this enzyme could also participate in nitrate formation from NO and $O_2^{\cdot -}$. Furthermore, we observed the activation of GST, disulfide isomerase and thioredoxin, peroxiredoxine, and thioredoxin reductases which participated in TNT stress response and oxidative stress response. The primary metabolic function of GST is the conjugation of electrophilic xenobiotic compounds followed by the efflux of the GST-xenobiotic complex from the cell (Barreto et al., 2006). An increase in the activity of disulfide isomerase located in endoplasmic reticulum and the subsequent microsome formation containing GST were found to trigger the lipid-soluble xenobiotics metabolism inside the yeast cells (Barreto et al., 2006). Thus, our data supports the concept of enzymatic neutralization of TNT-induced stress.

CONCLUSION

The results of the present study show that membrane-bound oxidoreductases are the pioneering proteins that can trigger the TNT transformation via TNT monohydride-Meisenheimer complexes. In particular, NADH flavin oxidoreductases/NADH oxidases related to OYEs, and some transferases can be up-regulated at all stages of TNT transformation. The upregulation of several stress response proteins (SOD, catalase, glutathione peroxidase, and GST) was also detected. Finally, the involvement of intracellular NOD in NO formation during nitrites oxidation was established. These findings support the biogenic method of NO formation in addition to the abiotic formation pathway explored in previous studies.

AUTHOR CONTRIBUTIONS

IK, GL, and OI planned and performed experiments, analyzed data, contributed reagents, and wrote the paper.

FUNDING

This work was performed within the Program of Competitive Growth of Kazan Federal University and supported by DAAD program "Yevgeny Zavoisky." IK thanks RFBR (grant no. 16-34-60200) for partial support of this work.

ACKNOWLEDGMENT

The authors thank the Interdisciplinary Center for Collective Use (ID RFMEFI59414X0003) sponsored by Ministry of Education and Science of the Russian Federation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02600/full#supplementary-material

REFERENCES

Adamia, G., Ghoghoberidze, M., Graves, D., Khatisashvili, G., Kvesitadze, G., Lomidze, E., et al. (2006). Absorption, distribution, and transformation of TNT in higher plants. *Ecotoxicol. Environ. Saf.* 64, 136–145. doi: 10.1016/j.ecoenv.2005.05.001

Barreto, L., Garcera, A., Jansson, K., Sunnerhagen, P., and Herrero, E. (2006). A peroxisomal glutathione transferase of *Saccharomyces cerevisiae* is functionally related to sulfur amino acid metabolism. *Eukaryot. Cell* 5, 1748–1759. doi: 10.1128/ec.00216-06

Beckman, J. S., Beckman, T. W., Chen, J., Marshall, P. A., and Freeman, B. A. (1990). Apparent hydroxyl radical production by peroxynitrite: implications for endothelial injury from nitric oxide and superoxide. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1620–1624. doi: 10.1073/pnas.87.4.1620

Beckman, J. S., and Koppenol, W. H. (1996). Nitric oxide, superoxide, and peroxynitrite: the good, the bad, and ugly. *Am. J. Physiol.* 271(5 Pt 1), C1424–C1437.

Boelsterli, U. A., Ho, H. K., Zhou, S., and Leow, K. Y. (2006). Bioactivation and hepatotoxicity of nitroaromatic drugs. *Curr. Drug Metab.* 7, 715–727. doi: 10.2174/138920006778520606

Borch, T., and Gerlach, R. (2004). Use of reversed-phase high-performance liquid chromatography–diode array detection for complete separation of 2,4,6-trinitrotoluene metabolites and EPA Method 8330 explosives: influence of temperature and an ion-pair reagent. *J. Chromatogr. A* 1022, 83–94. doi: 10.1016/j.chroma.2003.09.067

Caballero, A., Lazaro, J. J., Ramos, J. L., and Esteve-Nunez, A. (2005). PnrA, a new nitroreductase-family enzyme in the TNT-degrading strain *Pseudomonas putida* JLR11. *Environ. Microbiol.* 7, 1211–1219. doi: 10.1111/j.1462-2920.2005.00801.x

Cho, Y. S., Lee, B. U., Kahng, H. Y., and Oh, K. H. (2009). Comparative analysis of 2,4,6-trinitrotoluene (TNT)-induced cellular responses and proteomes in *Pseudomonas* sp. HK-6 in two types of media. *J. Microbiol.* 47, 220–224. doi: 10.1007/s12275-008-0108-0

Claus, H. (2014). "Microbial degradation of 2,4,6-Trinitrotoluene In Vitro and in natural environments," in *Biological Remediation of Explosive Residues*, ed. S. N. Singh (Cham: Springer International Publishing), 15–38.

Esteve-Nunez, A., Caballero, A., and Ramos, J. L. (2001). Biological degradation of 2,4,6-trinitrotoluene. *Microbiol. Mol. Biol. Rev.* 65, 335–352. doi: 10.1128/mmbr.65.3.335-352.2001

Farah, M. E., and Amberg, D. C. (2007). Conserved actin cysteine residues are oxidative stress sensors that can regulate cell death in yeast. *Mol. Biol. Cell* 18, 1359–1365. doi: 10.1091/mbc.E06-08-0718

Fitzpatrick, T. B., Amrhein, N., and Macheroux, P. (2003). Characterization of YqjM, an Old Yellow Enzyme homolog from Bacillus subtilis involved in the oxidative stress response. *J. Biol. Chem.* 278, 19891–19897. doi: 10.1074/jbc.M211778200

French, C. E., Nicklin, S., and Bruce, N. C. (1998). Aerobic degradation of 2,4,6-trinitrotoluene by *Enterobacter cloacae* PB2 and by pentaerythritol tetranitrate reductase. *Appl. Environ. Microbiol.* 64, 2864–2868.

Ho, E. M., Chang, H. W., Kim, S. I., Kahng, H. Y., and Oh, K. H. (2004). Analysis of TNT (2,4,6-trinitrotoluene)-inducible cellular responses and stress shock proteome in *Stenotrophomonas* sp. OK-5. *Curr. Microbiol.* 49, 346–352. doi: 10.1007/s00284-004-4322-7

Ignarro, L. J., Napoli, C., and Loscalzo, J. (2002). Nitric oxide donors and cardiovascular agents modulating the bioactivity of nitric oxide: an overview. *Circ. Res.* 90, 21–28. doi: 10.1161/hh0102.102330

Iman, M., Sobati, T., Panahi, Y., and Mobasheri, M. (2017). Systems biology approach to bioremediation of nitroaromatics: constraint-based analysis of 2,4,6-trinitrotoluene biotransformation by *Escherichia coli*. *Molecules* 22:E1242. doi: 10.3390/molecules22081242

Khan, H., Harris, R. J., Barna, T., Craig, D. H., Bruce, N. C., Munro, A. W., et al. (2002). Kinetic and structural basis of reactivity of pentaerythritol tetranitrate reductase with NADPH, 2-cyclohexenone, nitroesters, and nitroaromatic explosives. *J. Biol. Chem.* 277, 21906–21912. doi: 10.1074/jbc.M200637200

Khilyas, I. V., Ziganshin, A. M., Pannier, A. J., and Gerlach, R. (2013). Effect of ferrihydrite on 2,4,6-trinitrotoluene biotransformation by an aerobic yeast. *Biodegradation* 24, 631–644. doi: 10.1007/s10532-012-9611-4

Kumagai, Y., Kikushima, M., Nakai, Y., Shimojo, N., and Kunimoto, M. (2004). Neuronal nitric oxide synthase (NNOS) catalyzes one-electron reduction of 2,4,6-trinitrotoluene, resulting in decreased nitric oxide production and increased nNOS gene expression: implication for oxidative stress. *Free Radic. Biol. Med.* 37, 350–357. doi: 10.1016/j.freeradbiomed.2004.04.023

Lee, B. U., Cho, Y. S., Park, S. C., and Oh, K. H. (2009). Enhanced degradation of TNT by genome-shuffled *Stenotrophomonas* maltophilia OK-5. *Curr. Microbiol.* 59, 346–351. doi: 10.1007/s00284-009-9443-6

Lee, B. U., Park, S. C., Cho, Y. S., and Oh, K. H. (2008). Exopolymer biosynthesis and proteomic changes of *Pseudomonas* sp. HK-6 under stress of TNT (2,4,6-trinitrotoluene). *Curr. Microbiol.* 57, 477–483. doi: 10.1007/s00284-008-9272-z

Letzel, S., Göen, T., Bader, M., Angerer, J., and Kraus, T. (2003). Exposure to nitroaromatic explosives and health effects during disposal of military waste. *Occup. Environ. Med.* 60, 483–488. doi: 10.1136/oem.60.7.483

Mulla, S. I., Talwar, M. P., and Ninnekar, H. Z. (2014). "Bioremediation of 2, 4, 6-trinitrotoluene explosive residues," in *Biological Remediation of Explosive Residues*, ed. S. N. Singh (Cham: Springer International Publishing), 201–233.

Naumenko, E. A., Ahlemeyer, B., and Baumgart-Vogt, E. (2016). Species-specific differences in peroxisome proliferation, catalase, and SOD2 upregulation as well as toxicity in human, mouse, and rat hepatoma cells induced by the explosive and environmental pollutant 2,4,6-trinitrotoluene. *Environ. Toxicol.* 32, 989–1006. doi: 10.1002/tox.22299

Nizam, S., Verma, S., Borah, N. N., Gazara, R. K., and Verma, P. K. (2014). Comprehensive genome-wide analysis reveals different classes of enigmatic old yellow enzyme in fungi. *Sci. Rep.* 4:4013. doi: 10.1038/srep04013

Oh, B. T., Sarath, G., and Shea, P. J. (2001). TNT nitroreductase from a *Pseudomonas aeruginosa* strain isolated from TNT-contaminated soil. *Soil Biol. Biochem.* 33, 875–881. doi: 10.1016/S0038-0717(00)00233-9

Pak, J. W., Knoke, K. L., Noguera, D. R., Fox, B. G., and Chambliss, G. H. (2000). Transformation of 2,4,6-trinitrotoluene by purified xenobiotic reductase B from *Pseudomonas fluorescens* I-C. *Appl. Environ. Microbiol.* 66, 4742–4750. doi: 10.1128/AEM.66.11.4742-4750.2000

Rylott, E. L., Lorenz, A., and Bruce, N. C. (2011). Biodegradation and biotransformation of explosives. *Curr. Opin. Biotechnol.* 22, 434–440. doi: 10.1016/j.copbio.2010.10.014

Sens, C., Scheidemann, P., and Werner, D. (1999). The distribution of 14 C-TNT in different biochemical compartments of the monocotyledonous *Triticum aestivum*. *Environ. Pollut.* 104, 113–119. doi: 10.1016/S0269-7491(98)00142-0

Singh, S. N. (2014). *Biological Remediation of Explosive Residues*. Cham: Springer International Publishing.

Smets, B. F., Yin, H., and Esteve-Nunez, A. (2007). TNT biotransformation: when chemistry confronts mineralization. *Appl. Microbiol. Biotechnol.* 76, 267–277. doi: 10.1007/s00253-007-1008-7

Spain, J. C. (1995). Biodegradation of nitroaromatic compounds. *Annu. Rev. Microbiol.* 49, 523–555. doi: 10.1146/annurev.mi.49.100195.002515

Spanggord, R. J., Stewart, K. R., and Riccio, E. S. (1995). Mutagenicity of tetranitroazoxytoluenes: a preliminary screening in *Salmonella typhimurium* strains TA100 and TA100NR. *Mutat. Res.* 335, 207–211. doi: 10.1016/0165-1161(95)00022-4

Stenuit, B. A., and Agathos, S. N. (2010). Microbial 2,4,6-trinitrotoluene degradation: could we learn from (bio)chemistry for bioremediation and vice versa? *Appl. Microbiol. Biotechnol.* 88, 1043–1064. doi: 10.1007/s00253-010-2830-x

Strassner, J., Furholz, A., Macheroux, P., Amrhein, N., and Schaller, A. (1999). A homolog of old yellow enzyme in tomato. Spectral properties and substrate specificity of the recombinant protein. *J. Biol. Chem.* 274, 35067–35073. doi: 10.1074/jbc.274.49.35067

Symons, Z. C., and Bruce, N. C. (2006). Bacterial pathways for degradation of nitroaromatics. *Nat. Prod. Rep.* 23, 845–850. doi: 10.1039/b502796a

Vasquez-Vivar, J., Denicola, A., Radi, R., and Augusto, O. (1997). Peroxynitrite-mediated decarboxylation of pyruvate to both carbon dioxide and carbon dioxide radical anion. *Chem. Res. Toxicol.* 10, 786–794. doi: 10.1021/tx970031g

Wenge, B., Bönisch, H., Grabitzki, J., Lochnit, G., Schmitz, B., and Ahrend, M. H. (2008). Separation of membrane proteins by two-dimensional electrophoresis using cationic rehydrated strips. *Electrophoresis* 29, 1511–1517. doi: 10.1002/elps.200700546

Wessel, D., and Flügge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 138, 141–143. doi: 10.1016/0003-2697(84)90782-6

Williams, R. E., and Bruce, N. C. (2002). 'New uses for an old enzyme'–the old yellow enzyme family of flavoenzymes. *Microbiology* 148(Pt 6), 1607–1614. doi: 10.1099/00221287-148-6-1607

Williams, R. E., Rathbone, D. A., Scrutton, N. S., and Bruce, N. C. (2004). Biotransformation of explosives by the old yellow enzyme family of flavoproteins. *Appl. Environ. Microbiol.* 70, 3566–3574. doi: 10.1128/AEM.70.6.3566-3574.2004

Wittich, R. M., Haidour, A., Van Dillewijn, P., and Ramos, J. L. (2008). OYE flavoprotein reductases initiate the condensation of TNT-derived intermediates to secondary diarylamines and nitrite. *Environ. Sci. Technol.* 42, 734–739. doi: 10.1021/es071449w

Xia, Y., Cardounel, A. J., Vanin, A. F., and Zweier, J. L. (2000). Electron paramagnetic resonance spectroscopy with N-methyl-D-glucamine dithiocarbamate iron complexes distinguishes nitric oxide and nitroxyl anion in a redox-dependent manner: applications in identifying nitrogen monoxide products from nitric oxide synthase. *Free Radic. Biol. Med.* 29, 793–797. doi: 10.1016/S0891-5849(00)00427-5

Yao, D., Vlessidis, A. G., and Evmiridis, N. P. (2004). Determination of nitric oxide in biological samples. *Microchim. Acta* 147, 1–20. doi: 10.1007/s00604-004-0212-8

Zaripov, S. A., Naumov, A. V., Suvorova, E. S., Garusov, A. V., and Naumova, R. P. (2004). [Initial stages of 2,4,6-trinitrotoluene transformation by microorganisms]. *Mikrobiologiia* 73, 472–478. doi: 10.1023/B:MICI.0000036983.04480.19

Ziganshin, A. M., and Gerlach, R. (2014). "Pathways of 2, 4, 6-trinitrotoluene transformation by aerobic yeasts," in *Biological Remediation of Explosive Residues*, ed. S. N. Singh (Cham: Springer International Publishing), 301–311.

Ziganshin, A. M., Gerlach, R., Borch, T., Naumov, A. V., and Naumova, R. P. (2007). Production of eight different hydride complexes and nitrite release from 2,4,6-trinitrotoluene by *Yarrowia lipolytica*. *Appl. Environ. Microbiol.* 73, 7898–7905. doi: 10.1128/aem.01296-07

Ziganshin, A. M., Naumova, R. P., Pannier, A. J., and Gerlach, R. (2010). Influence of pH on 2,4,6-trinitrotoluene degradation by *Yarrowia lipolytica*. *Chemosphere* 79, 426–433. doi: 10.1016/j.chemosphere.2010.01.051

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Khilyas, Lochnit and Ilinskaya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 February 2018
doi: 10.3389/fmicb.2018.00232

Metagenomic Analysis of a Biphenyl-Degrading Soil Bacterial Consortium Reveals the Metabolic Roles of Specific Populations

*Daniel Garrido-Sanz, Javier Manzano, Marta Martín, Miguel Redondo-Nieto and Rafael Rivilla**

Departamento de Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

Polychlorinated biphenyls (PCBs) are widespread persistent pollutants that cause several adverse health effects. Aerobic bioremediation of PCBs involves the activity of either one bacterial species or a microbial consortium. Using multiple species will enhance the range of PCB congeners co-metabolized since different PCB-degrading microorganisms exhibit different substrate specificity. We have isolated a bacterial consortium by successive enrichment culture using biphenyl (analog of PCBs) as the sole carbon and energy source. This consortium is able to grow on biphenyl, benzoate, and protocatechuate. Whole-community DNA extracted from the consortium was used to analyze biodiversity by Illumina sequencing of a 16S rRNA gene amplicon library and to determine the metagenome by whole-genome shotgun Illumina sequencing. Biodiversity analysis shows that the consortium consists of 24 operational taxonomic units (≥97% identity). The consortium is dominated by strains belonging to the genus *Pseudomonas*, but also contains betaproteobacteria and *Rhodococcus* strains. whole-genome shotgun (WGS) analysis resulted in contigs containing 78.3 Mbp of sequenced DNA, representing around 65% of the expected DNA in the consortium. Bioinformatic analysis of this metagenome has identified the genes encoding the enzymes implicated in three pathways for the conversion of biphenyl to benzoate and five pathways from benzoate to tricarboxylic acid (TCA) cycle intermediates, allowing us to model the whole biodegradation network. By genus assignment of coding sequences, we have also been able to determine that the three biphenyl to benzoate pathways are carried out by *Rhodococcus* strains. In turn, strains belonging to *Pseudomonas* and *Bordetella* are the main responsible of three of the benzoate to TCA pathways while the benzoate conversion into TCA cycle intermediates via benzoyl-CoA and the catechol meta-cleavage pathways are carried out by beta proteobacteria belonging to genera such as *Achromobacter* and *Variovorax*. We have isolated a *Rhodococcus* strain WAY2 from the consortium which contains the genes encoding the three biphenyl to benzoate pathways indicating that this strain is responsible for all the biphenyl to benzoate transformations. The presented results show that metagenomic analysis of consortia allows the identification of bacteria active in biodegradation processes and the assignment of specific reactions and pathways to specific bacterial groups.

Keywords: biphenyl, PCBs, bacterial consortium, metagenomics, *Rhodococcus*

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council (CONICET), Argentina

Reviewed by:
Marc Viñas, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Spain
Sonja Kristine Fagervold, Université Pierre et Marie Curie (UPMC), France

****Correspondence:***
Rafael Rivilla
rafael.rivilla@uam.es

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *28 September 2017*
Accepted: *30 January 2018*
Published: *15 February 2018*

Citation:
Garrido-Sanz D, Manzano J, Martín M, Redondo-Nieto M and Rivilla R (2018) Metagenomic Analysis of a Biphenyl-Degrading Soil Bacterial Consortium Reveals the Metabolic Roles of Specific Populations. Front. Microbiol. 9:232. doi: 10.3389/fmicb.2018.00232

INTRODUCTION

Biphenyl has been widely used as a mineralizable polychlorinated biphenyls (PCBs) analog in biodegradation studies (Leigh et al., 2006; Uhlik et al., 2009; Leewis et al., 2016; Vergani et al., 2017a). PCBs are a family of man-made persistent organic chemicals that consist of a biphenyl skeleton where 1–10 hydrogen atoms are substituted by chlorine giving rise to up to 209 congeners. PCBs have been widely manufactured because of their chemical and physical properties (National Research Council, 1979) and a significant amount of PCBs has been released into the environment (Pieper, 2005; Sharma et al., 2014). The relative volatility of PCBs contributes to their spread throughout the globe (Gomes et al., 2013) where they bioaccumulate and biomagnify in the food web (Turrio-Baldassarri et al., 2007). PCBs have been shown to pose a broad range of exposure-related health effects in humans (Ross, 2004; Quinete et al., 2014) and are categorized as carcinogens (Mayes et al., 1998; Lauby-Secretan et al., 2013). Because of their chemical stability, poor water solubility, and toxicity, PCBs are considered recalcitrant toxics.

Bacteria can co-metabolize PCBs anaerobically and aerobically. Anaerobic cometabolism consists of reductive dehalogenation, a process in which highly chlorinated PCBs act as electron acceptors and reduce their chlorination (Quensen et al., 1988; Fennell et al., 2004). Thus, the biphenyl skeleton is not degraded through this pathway. Aerobic biodegradation on the contrary is better suited for low chlorinated congeners (Pieper, 2005; Furukawa and Fujihara, 2008; Pieper and Seeger, 2008) and biphenyl can be aerobically mineralized either by a single microorganism or by a consortium (Hernandez-Sanchez et al., 2013). Aerobic bioremediation of PCBs has been one of the main approaches to alleviate their persistence (Harkness et al., 1993; Pieper, 2005; Sharma et al., 2017) and usually occurs through its cometabolism by enzymes of the biphenyl upper degradation pathway, encoded by the *bphABCDEFG* gene cluster (Furukawa and Fujihara, 2008), although gene clusters for ethylbenzene (*etb*) and naphthalene (*nar*) degradation have also been shown to contribute to biphenyl and aerobic degradation of PCBs (Kimura et al., 2006; Iwasaki et al., 2007), resulting in the formation of (chloro)benzoic acid using biphenyl as carbon and energy source (Pieper, 2005; Pieper and Seeger, 2008). The specificity toward different PCB congeners depends mainly of the particular BphA enzyme (Gibson and Parales, 2000), some of which have been shown to produce the dechlorination of certain chlorinated biphenyls (Haddock et al., 1995; Seeger et al., 2001). The genes from the biphenyl upper degradative pathway have been extensively studied in *Paraburkholderia xenovorans* LB400, *Pseudomonas pseudoalcaligenes* KF707, and *Rhodococcus jostii* RHA1 due to the wide range of PCB congeners that they are able to metabolize (Seeger et al., 1995; Seto et al., 1995; Mondello et al., 1997; Furukawa and Fujihara, 2008). Aerobic degradation of PCBs usually occurs via cometabolism as their chlorinated derivatives might be channeled into dead-end pathways (Brenner et al., 1994) and it has been shown that some chlorinated intermediates are toxic to bacteria (Dai et al., 2002; Camara et al., 2004). After formation of (chloro)benzoic acid, it can be further funneled through catechol, protocatechuate, or the *box* pathways, ending up into tricarboxylic acid (TCA) cycle intermediates (Harwood and Parales, 1996; Gescher et al., 2002), known as the lower biphenyl degradation pathways.

Strategies for bioremediation of PCBs have been mainly focused on single microorganisms, either natural or modified (Haluska et al., 1995; Abbey et al., 2003; Sierra et al., 2003; Villacieros et al., 2005; Saavedra et al., 2010), which combined with biostimulation and bioaugmentation have resulted in enhanced degradation capabilities of a wide range of congeners (Singer et al., 2000; Fava et al., 2003; Ohtsubo et al., 2004; Field and Sierra-Alvarez, 2008). On the other hand, plant–microorganism interaction also plays a major role in degradation of PCBs (Leigh et al., 2006; Gerhardt et al., 2009; Vergani et al., 2017b). The use of PCB-degrading strains together with others that are capable of degrading their metabolic products (i.e., chlorinated benzoic acids) has also shown to extend the degradation rate of PCBs and results in complete mineralization of certain chlorobiphenyls (Fava et al., 1994; Hernandez-Sanchez et al., 2013).

In this study, we report the isolation and characterization of a soil bacterial consortium that is able to grow aerobically with the PCBs analog biphenyl as the sole carbon and energy source. In order to characterize this consortium, we have followed a metagenomic approach. Previous work using stable isotope probing (SIP) has shown to be useful in order to identify the bacterial populations implicated in biphenyl and benzoate degradation in soil microcosms (Leewis et al., 2016). However, the complexity of the bacterial community and the abundance of cross-feeders limit the study. Here, we show that reducing the community complexity to a lower number of bacterial populations by means of enrichment cultures, the metagenomic analysis allows not only to identify the populations playing a role in biphenyl and benzoate degradation but also to assign specific reactions and pathways to specific populations and therefore elucidating the trophic relationships occurring within the consortium to a higher detail.

MATERIALS AND METHODS

Isolation of the Biphenyl-Degrading Consortium and Growth Conditions

For the isolation of the biphenyl-degrading consortium, 2 g of rhizospheric soil collected near a petrol station (Tres Cantos, Madrid, Spain) was added to 500 ml of sterile liquid minimal salt medium (MM) (Brazil et al., 1995), supplemented with 1 ml/l of phosphate-buffered mineral medium salts (PAS) (Bedard et al., 1986) and 0.005% of yeast extract. One gram per liter of biphenyl crystals was added as the sole carbon and energy source. The culture was grown at 28°C with shaking (135 rpm) and maintained within a 9-day subculture. After five subcultures, when the culture was unable to grow without biphenyl as the sole carbon and energy source, 20 ml of the culture was centrifuged at 4,248 × *g*. The pellet was then resuspended in 0.75 ml of MM+PAS and mixed with 0.25 ml of glycerol (80%) and deep-frozen at −80°C. The isolated consortium was routinely grown

on MM+PAS with 1 g/l of biphenyl as the sole carbon and energy source at 28°C with shaking. For solid media, 1.5% agar (w/v) was added to the media and the biphenyl crystals were placed on the Petri dish lid.

The culture growth assessment on different organic compounds was performed as above but benzoic acid, protocatechuate, benzoate, 2-chlorobenzoic acid, 3-chlorobenzoic acid, or 4-chlorobenzoic acid (1 g/l) were added as the sole carbon and energy source.

DNA Extraction, Sequencing, Processing of Reads, and Assembly

DNA extraction from the biphenyl-degrading consortium at exponential growth (OD_{600} = 0.6) was carried out using the Realpure Genomic DNA Extraction Kit (Durviz, Spain). The 16S rRNA gene and the complete metagenome were sequenced by means of amplification of the V3–V4 16S rRNA region (primers 16SV3-V4-CS1; 5′-ACA CTG ACG ACA TGG TTC TAC ACC TAC GGG NGG CWG CAG-3′ and 16SV3-V4-CS2; 5′-TAC GGT AGC AGA GAC TTG GTC TGA CTA CHV GGG TAT CTA ATC C-3′) prior to libraries preparation and by whole-genome shotgun sequencing, respectively. The sequencing was carried out by Parque Científico de Madrid (Spain) using Illumina MiSeq paired 300-bp reads. Reads from the 16S rRNA gene and the whole metagenome were filtered and trimmed using Trimmomatic v0.36 (Bolger et al., 2014) software. Those with less than 50 nts in the case of the 16S rRNA gene or 100 nts in the case of the whole metagenome were removed. Reads from whole-metagenome sequencing were assembled using SPAdes v.10.1 software (Bankevich et al., 2012), metaSPAdes option, and default settings. Assembly quality was assessed using QUAST v4.4 (Gurevich et al., 2013). The resulting contigs were annotated using RAST (Aziz et al., 2008).

Reconstruction of Nearly Complete Genomes from Metagenome Shotgun Sequencing

Trimmed pair-reads from the whole-metagenome shotgun sequencing (as described above) were mapped against all available and closed NCBI genomes of *Achromobacter*, *Bordetella*, *Cupriavidus*, *Microbacterium*, *Pseudomonas*, *Rhodococcus*, and *Stenotrophomonas* using bowtie2 v 2.3.3.1 software (Langmead and Salzberg, 2012) with an expected range of inter-mate distances between 373 and 506 nts, consecutive seed extension attempts of 20, number of mismatches allowed in a seed alignment of 0, and length of the seed substrings to align of 20. For each genus, mapping reads and those without matching alignments across all genera examined were merged, processed, and retrieved with samtools v1.6 software (Li et al., 2009) for further assembly with SPAdes. Chimeric and misassigned contigs were checked by comparing assemblies of each genus against the same databases used for reads mapping using BLAST v.2.2.28+ software (Camacho et al., 2009). Contigs without positive hits within the expected genus were removed along with those with matching hits belonging to different genera. Contigs of *Cupriavidus*, *Microbacterium*, and *Rhodococcus* assemblies were also removed as genomic sizes were too small for a complete or nearly complete genome. In the case of *Pseudomonas*, contigs were also classified as belonging to *P. pseudoalcaligenes* or *P. putida* based on best blast hits.

Diversity Analysis of the 16S rRNA Gene and Coding DNA Sequences (CDSs)

Data analysis of the 16S rRNA gene diversity was assessed with QIIME v1.9.0 (Caporaso et al., 2010) and UPARSE v9 (Edgar, 2013) following the 16S profiling data analysis pipeline specified in the Brazilian Microbiome Project[1]. Briefly, filtered and trimmed forward and reverse reads were assembled using the fastq-join algorithm[2] and further length-filtered by a minimum of 430 nts, representing more than 99% of total reads. Singletons were also removed. These sequences were imported into UPARSE to identify operational taxonomic units (OTUs) at a 97% sequence identity. Chimeras were removed using SILVA v123 database (Quast et al., 2013) as reference, which was also used for genus assignation. QIIME was also used to perform alpha rarefaction analysis. Convergence of observed OTUs rarefaction curve was determined using R (R Core Team, 2013) and the R package iNEXT (Hsieh et al., 2016) with a bootstrapping of 1,000 and a confidence interval of 5%.

To assess the diversity of coding DNA sequences (CDSs), after whole-metagenome assembly and annotation (see above), CDSs were blasted against the NCBI nt database (on April 2017) using blastn from BLAST v2.2.28+ software (Camacho et al., 2009). For each query, the first hit was kept and further filtered by a minimum of 75% sequence identity and 50% coverage. Genus assignation of the CDSs was based on the subject entry.

Identification of CDSs Involved in Biphenyl Metabolism and Phylogenetic Analysis

Aminoacid sequences for biphenyl 2,3-dioxygenase (BphA1), BenA, benzoate-CoA ligase (BclA), CatA, CatE, PobA, protocatechuate 4,5-dioxygenase alpha subunit (LigA), and protocatechuate 3,4-dioxygenase alpha subunit (PcaG) enzymes (Supplementary File 1) were downloaded from the NCBI and used to build blast databases using makeblastdb from BLAST. These databases were used as queries for orthologs identification within the whole-metagenome proteome. Results were filtered by 75% sequence identity, 50% coverage, and 1e−10 expected value and further blasted against the nr NCBI database (on April 2017) to validate their annotation. After orthologs identification, clusters of CDSs were searched within the whole-metagenome contigs and represented using own Perl scripts. Contigs carrying *bph* CDSs were also compared with those reported on reference sequences of *Rhodococcus* strains HA99 (AB272986.1), RHA1 (AB120955.1), and SAO101 (AB110633.1) to reconstruct the gene clusters using Clustal Omega (Sievers et al., 2011). Synteny representation was based on GenBank annotations and represented as described above.

[1] http://www.brmicrobiome.org/

[2] https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqJoin.md

Phylogenetic Analysis

BphA1, NarA1, and EtbA1 protein sequences from the metagenome annotation of the biphenyl-degrading consortium were aligned using Clustal Omega (Sievers et al., 2011) against 15 well-known BphA1 and closely related NarA1 and EtbA1 protein sequences. Results were imported into MEGA v7 (Kumar et al., 2016) to build the phylogenetic tree using maximum-likelihood with Tamura–Nei model, 1,000 bootstrap replicates, and represented with MEGA. BenA protein sequence of *Pseudomonas putida* PRS200 was used as an outgroup.

Rhodococcus Isolation and Genetic Analysis

Rhodococcus sp. WAY2 was isolated by plating washed ($NaCl_2$ 0.85%) and diluted biphenyl-degrading consortium culture on MM+PAS solid medium with biphenyl (1 g/l) as the sole carbon and energy source. After 12 days of incubation at 28°C, colonies were replated under the same conditions as above. This process was repeated twice. Finally, a single colony was grown on liquid MM+PAS media supplemented with 1 g/l of biphenyl. The culture was centrifuged at 4,248 × *g* prior to DNA extraction using the Realpure Genomic DNA Extraction Kit (Durviz, Spain). 16S rRNA gene was amplified using the universal primer pairs 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492R (5′-CTA CGR RTA CCT TGT TAC GAC-3′) (Weisburg et al., 1991). Amplicons were cloned into pGEM®-T Easy Vector System I (Promega) and transformed into *E. coli* DH5α. Plasmid DNA was extracted using the kit Wizard® Plus SV Minipreps DNA Purification System (Promega). Inserts were sequenced by means of Sanger sequencing using the universal primers T7 and SP6.

The three *bph* gene clusters identified in the whole metagenome of the biphenyl-degrading consortium were screened by PCR on the genome of the isolated *Rhodococcus* sp. WAY2 using the own-designed primers BphClus1F (5′-CGC CTC ATC ACG AAT GTG ACC G-3′), BphClus1R (5′-GCG TCC TCA TGC GTA CAG GTG TCC-3′), BphClus2F (5′-CGA CTG CTC GGA CTG GAG GG-3′), BphClus2R (5′-CCC ATC GAG TTA CCG ACT ATG TGC G-3′), BphClus3F (5′-GCC CGA CCA AGC AGT ACA AAG TG-3′), and BphClus3R (5′-GTC CAG TCG GAC TTC ACG TCG-3′). Primers were designed on the genomic sequence of these clusters. Melting temperature, absence of dimerization and hairpin formation, and lack of secondary priming sites were assessed with OligoAnalyzer 3.1[3]. PCR was carried out in a total volume of 25 µl containing 2.5 µl of 10× PCR buffer $MgCl_2$ free, 1 µl $MgCl_2$ 50 mM, 0.5 µl dNTP mix 10 mM (2.5 µM each), 1 µl of each primer at 10 µM, 1 µl of Taq DNA polymerase 1 U/µl (Biotools), and 1 µl of DNA template 30–50 ng/µl. The cycling conditions consisted in a first denaturation step at 95°C for 5 min followed by 32 cycles of amplification (45 s denaturation at 95°C, 45 s of primer annealing at 58°C, and an elongation step at 72°C for 1.5 min) followed by a final elongation step at 72°C for 7 min. PCR products were electrophoretically separated in 0.8% (w/v) agarose gels and post-dyed with GelRed.

[3]https://eu.idtdna.com/calc/analyzer

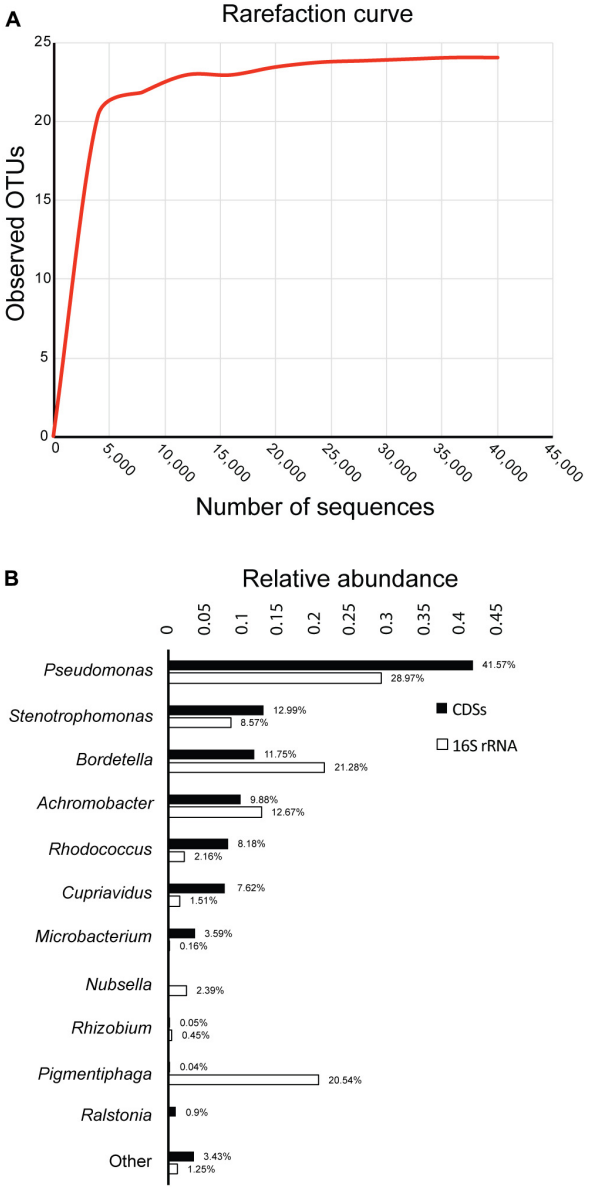


FIGURE 1 | Diversity and composition of the biphenyl-degrading consortium. **(A)** Rarefaction curve of observed OTUs (≥97% sequence identity) over the number of 16S rRNA sequences and **(B)** relative abundance of genus based on 16S rRNA and CDSs taxonomic assignment. Only taxa with a minimum relative abundance of 0.15% for 16S rRNA and 0.9% for CDSs is represented.

Sequence Deposition

Raw reads of the 16S rRNA gene amplicons and whole-metagenome shotgun sequencing of the biphenyl-degrading consortium were deposited to the NCBI Sequence Read Archive under the accession numbers SRR6076973 and SRR6076972, respectively. Assemblies of *Achromobacter* sp., *Bordetella* sp., *P. pseudoalcaligenes*, *Pseudomonas* sp., and *Stenotrophomonas* sp. reconstructed from the metagenome were deposited to GenBank under the accession numbers PKCB00000000, PKCD00000000, PKCC00000000, PKCE00000000, and

PKCF00000000, respectively. The 16S rRNA gene sequence of the isolated *Rhodococcus* sp. WAY2 was submitted to GenBank and it is available under the accession number MF996860. The 16S rRNA gene sequence of the 24 identified OTUs is shown in Supplementary File 2.

RESULTS AND DISCUSSION

Metagenomic Sequencing and Bacterial Diversity

After sequencing the 16S rRNA genes of bacteria in the biphenyl-degrading consortium, a total of 44,644 sequences were obtained and assigned to 24 OTUs (≥97% sequence identity). The rarefaction curve shows a clear and early saturation of observed OTUs, as shown in **Figure 1A**, which indicates that a full community coverage was achieved before 40,000 sequences and the presence of other taxa is unlikely. Furthermore, statistical analysis of the rarefaction curve (Supplementary File 3) showed that doubling the sampling would not increase the number of detected OTUs. On the other hand, the whole-genome shotgun sequencing of the metagenome resulted in 78.4 Mpb distributed in 45,046 contigs (Supplementary File 4). After annotation, 66,967 coding DNA sequences (CDSs) were obtained, from which 47,689 (71.2%) were assigned to the genus level, showing a high concordance with the identified OTUs. The relative abundance of the 16S rRNA and the CDSs (**Figure 1B**) shows that the biphenyl-degrading consortium is clearly dominated by *Pseudomonas* (28.97% 16S rRNA and 41.57% CDSs). Other genera that are present in the consortium are *Bordetella* (21.28% 16S rRNA and 11.75% CDSs), *Achromobacter* (12.67% 16S rRNA and 9.88% CDSs), *Stenotrophomonas* (8.57% 16S rRNA and 12.99% CDSs), *Rhodococcus* (2.18% 16S rRNA and 8.17% CDSs), and *Cupriavidus* (1.51% 16S rRNA and 7.62% CDSs). This distribution is detailed in Supplementary File 5. The main difference between the 16S rRNA and CDSs relative genus abundance lies in *Pigmentiphaga*, which is relatively abundant in the 16S rRNA analysis (20.54%) but is almost absent on CDSs representation (0.04%). This is probably due to lack of sequenced *Pigmentiphaga* genomes in the NCBI database, which makes CDSs assignation to this genus impossible and explains the higher relative abundance of the remaining genera in the CDSs diversity analysis. However, some genera, such as *Bordetella* and *Achromobacter*, have a lower relative CDSs representation than in the 16S rRNA. This could be explained by an incomplete metagenome, given that around 120 Mpb metagenome size was expected (considering an average bacterial genome size of 5 Mpb) to achieve a full genomic representation of the 24 OTUs identified in the biphenyl-degrading consortium. Furthermore, the presence of only 16 16S rRNA genes annotated in the metagenome is congruent with an incomplete one. However, it is important to indicate that the seven most represented genera represent more than 95% of the bacterial community and 96% of the identified CDS (**Figure 1B**), indicating a high coverage of the metagenome. This level of coverage would be impossible to achieve analyzing directly a soil sample or microcosm.

On the other hand, we have been able to reconstruct five nearly complete genomes from the whole-metagenome sequence, which correspond with the most abundant OTUs identified in the consortium (**Table 1**). These include two genomes classified as *P. pseudoalcaligenes* and *Pseudomonas* sp., *Achromobacter* sp., *Bordetella* sp., and *Stenotrophomonas* sp. Their genomic sizes and %GC content are congruent with their closest relative genome.

Identification of Biphenyl Upper Degradative Pathway Gene Clusters

In order to identify the metabolic pathways involved in the biphenyl biodegradation that are present in the whole metagenome of the biphenyl-degrading consortium, alpha subunits of the BphA1 were used as query to search for orthologous sequences. Three different BphA1 were identified (**Table 2**), which are present in three different contigs and are classified as belonging to the *Rhodococcus* genus by sequence identity (Supplementary File 6). BphA1 encodes the α subunit of biphenyl dioxygenases, and are responsible for the enzyme specificity (Gibson and Parales, 2000). As shown in **Figure 2D**, BphA1 proteins can be classified into three families. Typical BphA1 have been identified and characterized in many bacterial

TABLE 1 | Genomic statistics of the five nearly complete genomes reconstructed from the whole-metagenome sequence of the biphenyl-degrading consortium.

Assembly (Accs. No.)	OTU No.[a]	Closest relative genome[b] (Accs. No.)	Contigs	Largest contig	Total length	GC%	N50
Achromobacter sp. (PKCB00000000)	4	*A. xylosoxidans* DPB_1 (MTLI00000000.1)	2,603	31,468	7,352,073	65.7	3,641
Bordetella sp. (PKCD00000000)	2	*B. petrii* DSM 12804 (NC_010170.1)	2,637	33,811	5,644,602	65.3	3,914
Pseudomonas pseudoalcaligenes (PKCC00000000)	3	*P. pseudoalcaligenes* KF707 (NZ_AP014862.1)	1,404	73,619	5,455,424	66.2	8,550
Pseudomonas sp. (PKCE00000000)	12	*P. putida* KF715 (NZ_AP015029.1)	7,053	12,148	6,703,495	63.4	1,079
Stenotrophomonas sp. (PKCF00000000)	5	*S. maltophilia* ISMMS3 (NZ_CP011010.1)	411	38,486	4,489,164	66.8	12,006

[a] *See Supplementary File 2.* [b] *According to contigs size and best blast hits.*

TABLE 2 | Summary of the number and genus affiliation of the main CDSs for enzymes involved in the biphenyl and metabolic derivatives degradation identified in the biphenyl-degrading consortium.

Gene	Protein/genus assignation	Number of CDSs
bphA[1]	Biphenyl 2,3-dioxigenase (EC 1.14.12.18)	3
	Rhodococcus	3
benA[1]	Benzoate 1,2-dioxigenase (EC 1.14.12.10)	10
	Pseudomonas	5
	Bordetella	4
	Rhodococcus	1
catA	Catechol 1,2-dioxygenase (EC 1.13.11.1)	13
	Pseudomonas	5
	Rhodococcus	4
	Bordetella	2
	Achromobacter	1
	Variovorax	1
catE	Catechol 2,3-dioxygenase (EC 1.13.11.2)	5
	Variovorax	2
	Cupriavidus	2
	Uncultured/unclassified	1
pobA	4-Hydroxybenzoate 3-monooxygenase (EC 1.14.13.2)	10
	Pseudomonas	4
	Bordetella	2
	Uncultured/unclassified	1
	Achromobacter	1
	Ralstonia	1
	Rhodococcus	1
pcaG[1]	Protocatechuate 3,4-dioxygenase (EC 1.13.11.13)	9
	Pseudomonas	4
	Achromobacter	2
	Bordetella	1
	Cupriavidus	1
	Ralstonia	1
ligA[1]	Protocatechuate 4,5-dioxygenase (EC 1.13.11.8)	4
	Uncultured/unclassified	2
	Pseudomonas	1
	Bordetella	1
boxA[1]	Benzoyl-CoA oxygenase (EC 1.14.13.208)	4
	Achromobacter	3
	Variovorax	1

[1] *Only alpha subunits of multimeric enzymes are considered.*

strains, including *P. xenovorans* LB400 (Seeger et al., 1995), *P. pseudoalcaligenes* KF707 (Taira et al., 1992), and *R. jostii* RHA1 (Seto et al., 1995). None of the BphA1 CDS identified here belongs to this family. A second family of atypical BphA1 was identified in several strains of the genus *Rhodococcus*, including strains HA99 and R04 (Taguchi et al., 2007; Yang et al., 2007). One of the CDS identified here is identical to these atypical BphA1. The other family is formed by proteins with proved BphA1 activity, but formerly identified as NarA1 or EtbA1. These proteins have also been identified within the genus *Rhodococcus* (Kimura et al., 2006; Iwasaki et al., 2007) and two of the BphA1 CDSs identified here are identical to CDSs in *Rhodococcus opacus* SAO101 and *R. jostii* RHA1, respectively. On the other hand, the comparison between these CDSs and the ones previously reported in other *Rhodococcus* strains sequences allowed us to reconstruct the *bph* gene clusters from the whole-metagenome contigs, as shown in **Figure 2**. The first cluster (**Figure 2A**) was reconstructed from four different metagenome contigs and shows high sequence identity with the *bph* gene clusters reported in *Rhodococcus* sp. HA99 (Taguchi et al., 2007). This cluster is composed by *bphBCA1A2A3A4* and *bphD*, which are responsible for biphenyl and PCBs degradation into (chloro)benzoate and 2-hydroxypenta-2,4-dienoate (Taguchi et al., 2007). The second gene cluster (**Figure 2B**) was reconstructed from three different metagenome contigs and presents high sequence identity with *bph* and *etb* gene clusters which have been reported to be involved in both, biphenyl and PCBs degradation in *R. jostii* RHA1 (Iwasaki et al., 2006, 2007). This cluster is composed by *etbA1A2C* and *bphDE2F2*. The third gene cluster is present in a single metagenome contig (**Figure 2C**) and shows high sequence identity with *nar* gene clusters previously described in the plasmid pWK301 of *R. opacus* SAO101 (Kimura et al., 2006). This gene cluster is composed by *narA1A2BC* and two transcriptional regulators *narR1R2* and it has been reported to be involved in the degradation of a wide range of substrates, including biphenyl and PCBs (Kimura and Urushigawa, 2001; Kitagawa et al., 2004; Kimura et al., 2006). These results strongly suggest that *Rhodococcus* is the only genus responsible for initiating the biphenyl degradation in the consortium and that initial degradation can proceed through three distinct pathways. To our knowledge, multiple pathways have only been found in *R. jostii* RHA1, where a *bph* and an *etb* pathways have been described (Iwasaki et al., 2006, 2007).

To further study if the *bph*, *etb*, and *nar* gene clusters identified in the metagenome belong to one or multiple *Rhodococcus* strains that might be present in the biphenyl-degrading consortium, we isolated a *Rhodococcus* strain (*R.* sp. WAY2) from the consortium and tested for the presence of these three gene clusters by means of PCR. The results revealed that the three clusters are present in a single *Rhodococcus* strain WAY2, which 16S rRNA showed a high sequence identity (>99%) with *R. jostii* RHA1. This might suggest that the *etb* gene cluster is present in the chromosome of the isolated WAY2 strain as it is in the case of RHA1, while *bph* and *nar* gene clusters could be present in plasmids, as reported in strains HA99 and SAO101, respectively (Kimura et al., 2006; Taguchi et al., 2007).

Identification of Biphenyl Lower Degradative Pathway Genes

Biphenyl is metabolized to benzoate and 2-hydroxypenta-2,4-dienoate by either the *bph*, *etb*, or *nar* gene clusters. Benzoate can be then further mineralized by three different aerobic pathways: catechol, protocatechuate, or benzoyl-coA ligation (Harwood and Parales, 1996; Rather et al., 2010; Fuchs et al., 2011). All the CDSs for enzymes of these aerobic benzoate degradation pathways were screened and found in the metagenome of the biphenyl-degrading consortium and are summarized in **Table 2** (for details see Supplementary File 6). The benzoate degradative pathway via catechol formation

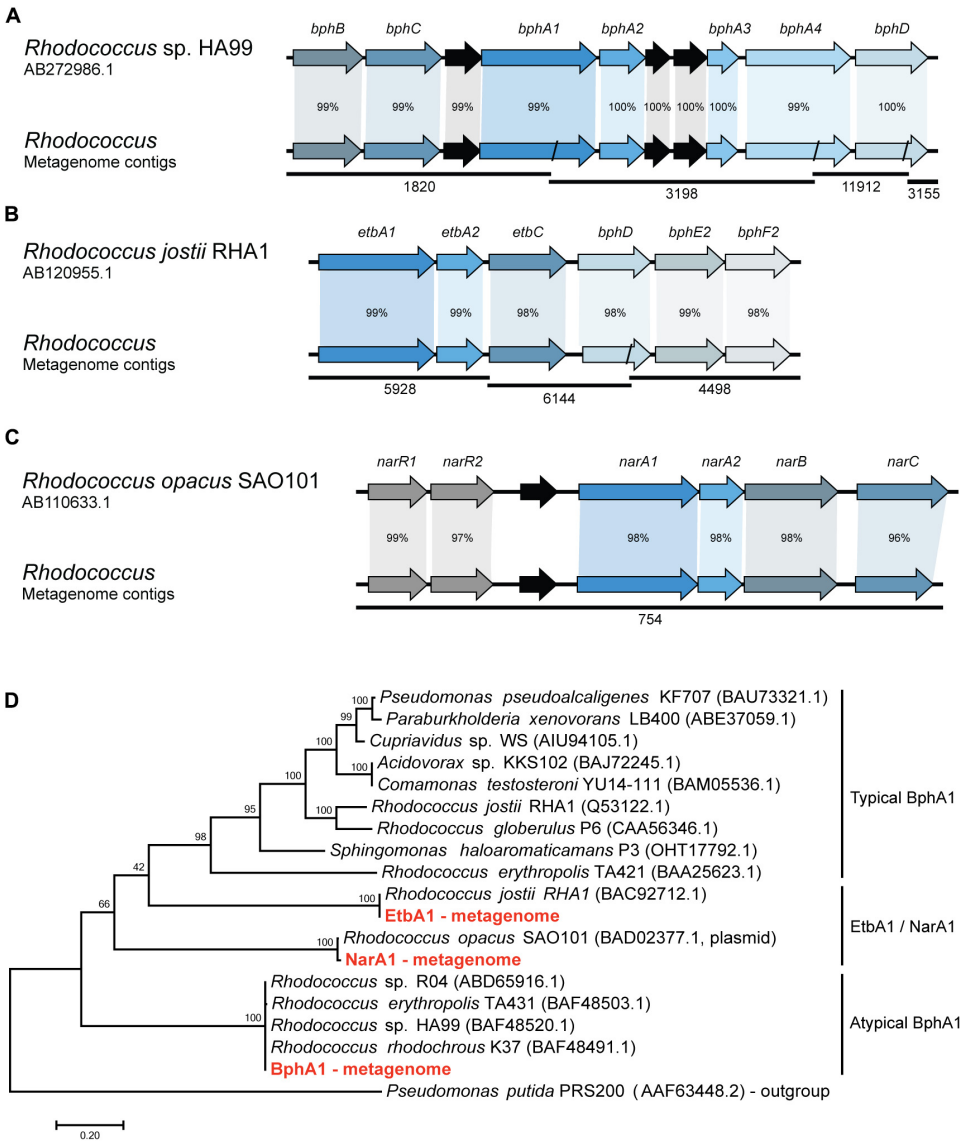


FIGURE 2 | Synteny and sequence identity of gene clusters involved in biphenyl degradation compared with reference sequences. **(A)** Biphenyl degradative gene cluster, **(B)** ethylbenzene degradative gene cluster, and **(C)** naphthalene degradative gene cluster. *Rhodococcus* sp. HA99, *R. jostii* RHA1, and *R. opacus* SAO101 sequences are shown as reference. Contigs from the metagenome are represented as black lines and their ID number is shown below. Black arrows represent hypothetical genes. Percentage according to nucleotide sequence identity of the CDSs. **(D)** Phylogenetic tree showing the relation of the isolated BphA1 protein sequences with previously characterized proteins. A BenA protein sequence from *Pseudomonas putida* was used as an outgroup.

is first initiated by BenABCD to form catechol. The coding sequence for benzoate 1,2-dioxygenase alpha subunit (BenA) was found 10 times in different contigs and was mainly assigned to *Pseudomonas* (five) and *Bordetella* (four). The remaining one was assigned to *Rhodococcus* (**Table 2**). After catechol formation, it can be further mineralized by ortho or meta cleavage, in which catechol 1,2-dioxygenase (CatA) or catechol 2,3-dioxygenase (CatE) is, respectively, involved. The coding sequence of CatA was found 13 times in the metagenome and was mainly assigned to *Pseudomonas* (five) and *Rhodococcus* (four). The remaining ones were assigned to *Bordetella* (two), *Achromobacter* (one), and *Variovorax* (one) (**Table 2**). On the other hand, the coding sequence for CatE was found five times in the metagenome and was assigned to *Variovorax* (two), *Cupriavidus* (two), and the remaining two could not be assigned (**Table 2**). These results suggest that the degradation of benzoate via catechol is mainly supported by *Pseudomonas*, *Bordetella*, and *Rhodococcus*, while other genera such as *Achromobacter*, *Variovorax*, and *Cupriavidus* have a smaller involvement in this pathway. Regarding the presence of this pathway in *Rhodococcus*, the isolated strain *R.* sp. WAY2 was unable to grow on benzoate as the sole carbon and energy source, suggesting that another *Rhodococcus* strain, different than the one harboring the *bph*,

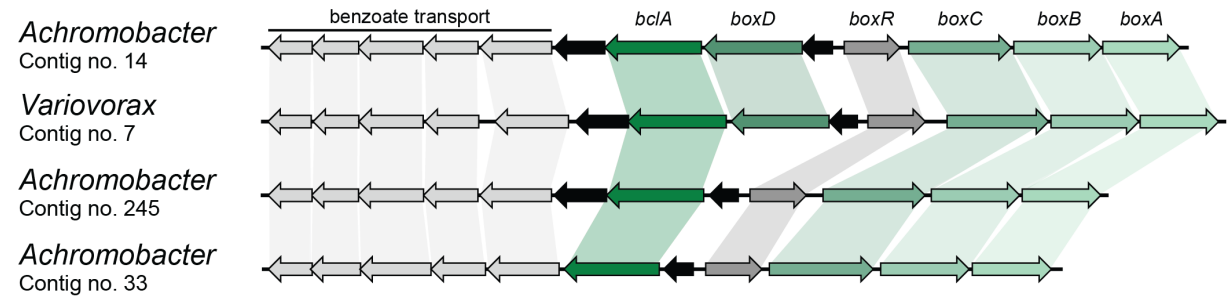


FIGURE 3 | *Box* gene clusters identified in the metagenome of the biphenyl-degrading consortium. Black arrows represent genes with no involvement in benzoate degradation. Genus assignation of the clusters based on sequence identity of CDSs.

TABLE 3 | Summary of the pathways assigned to the main genus present in the biphenyl-degrading consortium.

			Metabolic pathway identified							
Genus	**Relative abundance (%)**		**Biphenyl to benzoate**	**Benzoate to catechol**	**Catechol ortho cleavage**	**Catechol meta cleavage**	**Benzoate to protocatechuate**	**Protocatechuate ortho cleavage**	**Protocatechuate meta cleavage**	**Benzoate to benzoyl-CoA**
	16S rRNA	**CDSs**								
Pseudomonas	28.97	41.57		+	+		+	+	+	
Bordetella	21.28	11.75		+	+		+	+	+	
Pigmentiphaga	20.54	0.04								
Achromobacter	12.67	9.88			+		+	+		+
Stenotrophomonas	8.57	12.99								
Rhodococcus	2.16	8.18	+	+	+		+			
Cupriavidus	1.51	7.62				+	+	+		

etb, and *nar* gene clusters, is present in the biphenyl-degrading consortium.

Benzoate can also be metabolized via protocatechuate formation, in which a benzoate 4-monooxygenase (CYP450) and a 4-hydroxybenzoate 3-monooxygenase (PobA) are involved (Fuchs et al., 2011). The coding sequence of PobA was found 10 times in different contigs in the metagenome and was assigned to *Pseudomonas* (four), *Bordetella* (two), *Achromobacter* (one), *Ralstonia* (one), *Rhodococcus* (one), and the remaining one could not be assigned to any genus (**Table 2**). After protocatechuate formation, it can also be mineralized via ortho and meta cleavage, in which protocatechuate 3,4-dioxygenase (PcaGH) and protocatechuate 4,5-dioxygenase (LigAB) are, respectively, involved. The coding sequence for PcaG was found nine times in the metagenome and was assigned to *Pseudomonas* (four), *Achromobacter* (two), *Bordetella* (one), *Cupriavidus* (one), and *Ralstonia* (one) (**Table 2**). On the other hand, the coding sequence of LigA was found four times in the metagenome and was assigned to *Pseudomonas* (one) and *Bordetella* (one). The remaining ones could not be assigned to any genus (**Table 2**). These results suggest that the degradation of benzoate via protocatechuate formation is also dominated by *Pseudomonas* and *Bordetella*, harboring both, the ortho and meta protocatechuate cleavage pathways, while *Achromobacter*, *Ralstonia*, and *Cupriavidus* only have the coding sequences for protocatechuate formation and/or its ortho-cleavage pathway.

Finally, benzoate can also be mineralized by a novel pathway in which acetyl-CoA is first ligated to benzoate by a BclA and further epoxidated by benzoyl-CoA 2,3-epoxidase (BoxAB) (Rather et al., 2010). The coding sequence for BoxA was found four times in different contigs in the metagenome and was assigned to *Achromobacter* (three) and *Variovorax* (one) (**Table 2**). Contigs carrying the BoxA-coding sequence were also found to contain the remaining genes for the *box* cluster (*boxABCD* and *bclA*), along with the transcriptional regulator *boxR* and several coding sequences involved in benzoate transport, as shown in **Figure 3**. However, two of these contigs assigned to *Achromobacter* lack the *boxD* gene, which might result in dead-end production of 3,4-didehydroadipyl-CoA semialdehyde and formate, although they could be source of carbon and energy through alternative pathways.

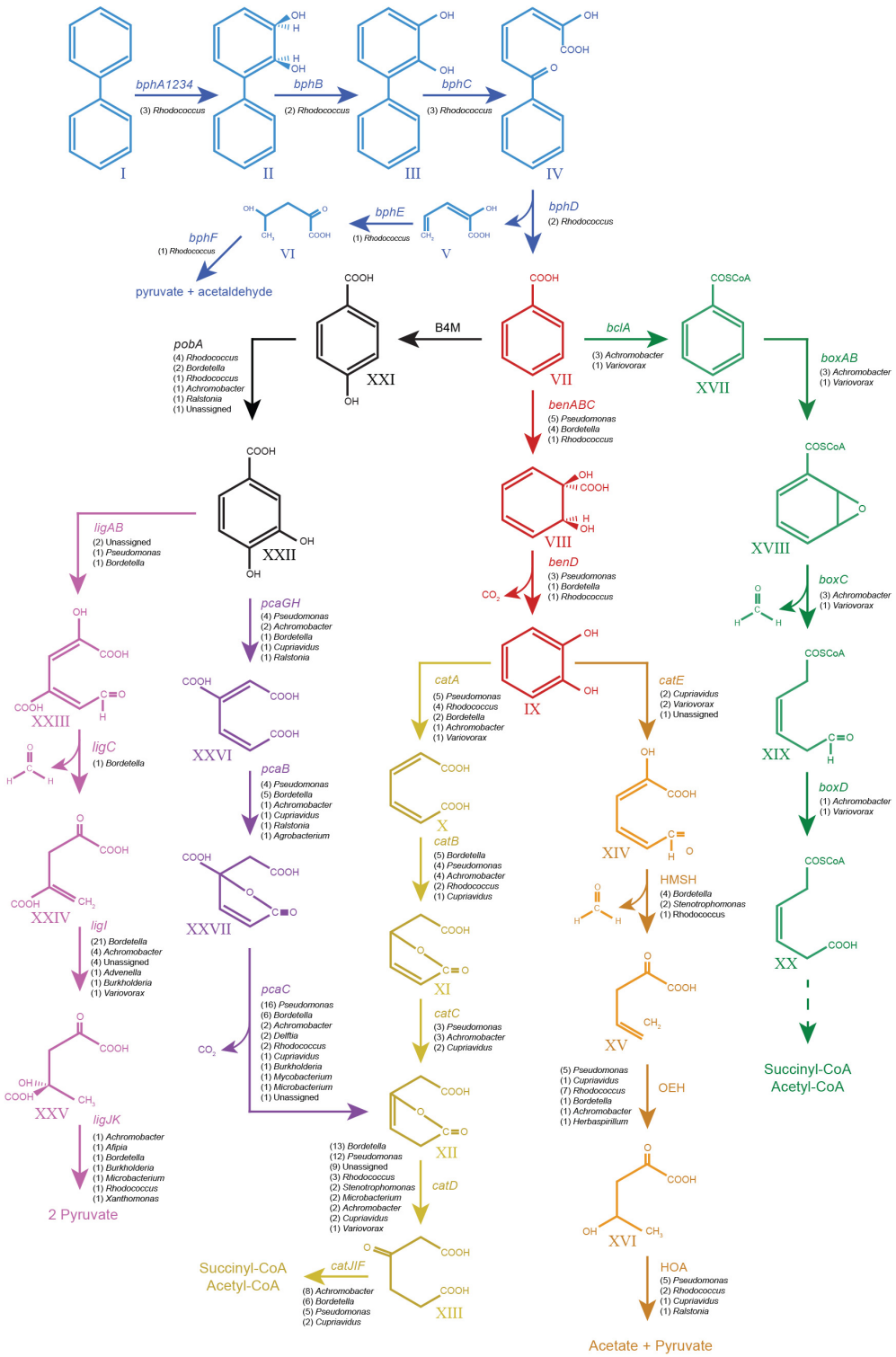


FIGURE 4 | Pathways from biphenyl degradation identified in the metagenome of the biphenyl-degrading consortium. Blue, biphenyl degradation; red, benzoate degradation via catechol; black, benzoate degradation via protocatechuate; violet, protocatechuate degradation via meta cleavage; purple, protocatechuate degradation via ortho cleavage; yellow, catechol degradation via ortho cleavage; orange, catechol degradation via meta cleavage; green, benzoate degradation via benzoyl-CoA formation. All the genes shown in the graph have been found in the metagenome of the biphenyl-degrading consortium. Their number of CDSs and genus assignation are specified under the gene names. Compounds: I, biphenyl; II, 2,3-dihydroxy-4-phenylhexa-4,6-diene; III, 2,3-dihydroxybiphenyl; IV, 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate; V, 2-hydroxypenta-2,4-dienoate; VI, 4-hydroxy-2-oxopenta; VII, benzoate; VIII, 2-hydro-1,2-dihydroxybenzoate;

(Continued)

FIGURE 4 | Continued
IX, catechol; X, *cis*,*cis*-muconate; XI, mucolactone; XII, 3-oxooadipate enol-lactone; XIII, 3-oxoadipate; XIV, 2-hydroxy-muconate-6-semialdehyde; XV, 2-oxo-penta-4-enoate; XVI, 4-hydroxy-2-oxovalerate; XVII, benzoyl-CoA; XVIII, 2,3-epoxy-benzoyl-CoA; XIX, 3,4-dehydroadipyl-CoA semialdehyde; XX, 3,4-dehydroadipyl-CoA; XXI, hydroxybenzoate; XXII, protocatechuate; XXIII, 2-hydroxy-4-carboxymuconic semialdehyde; XXIV, 2-keto-4-carboxypenta-enoate; XXV, 4-hydroxy-4-carboxy-2-ketovalerate; XXVI, 3-carboxy-*cis*,*cis*-muconate; and XXVII, 4-carbxymucolactone. Genes: *bphA1A2A3A4*, biphenyl 2,3-dioxygenase; *bphB*, *cis*-2,3-dihydrobiphenyl-2,3-diol dehydrogenase; *bphC*, biphenyl-2,3-diol 1,2-dioxygenase; *bphD*, 2,6-dioxo-6-phenylhexa-3-enoate hydrolase; *bphE*, 2-hydroxypenta-2,4-dienoate hydratase; *bphF*, 4-hydroxy-2-oxovalerate aldolase; *benABC*, benzoate 1,2-dioxygenase; *benD*, 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase; *catA*, catechol 1,2-dioxygenase; *catB*, muconate cycloisomerase; *catC*, muconolactone delta-isomerase; *catD*, 3-oxoadipate enol-lactonase; *catIJ*, 3-oxoadipate CoA-transferase; *catF*, 3-oxoadipyl-CoA thiolase; *catE*, catechol 2,3-dioxygenase; 2HM H, 2-hydroxymuconate semialdehyde hydrolase; 2OE H, 2-oxopent-4-enoate hydratase, 4HO A, 4-hydroxy-2-oxovalerate aldolase; B4M, benzoate 4-monooxygenase; *pobA*, 4-hydroxybenzoate 3-monooxygenase; *ligAB*, protocatechuate 4,5-dioxygenase; *ligC*, 2-hydroxy-4-carboxymuconate semialdehyde hemiacetal dehydrogenase; *ligI*, 2-pyrone-4,6-dicarboxylate lactonase; *ligJ*, 4-oxalomesaconate hydratase; *ligK*, 4-hydroxy-4-methy-2-oxoglutarate aldolase; *pcaGH*, protocatechuate 3,4-dioxygenase; *pcaB*, 3-carboxy-*cis*,*cis*-muconate cycloisomerase; *pcaC*, 4-carboxymuconolactone decarboxylase; *blcA*, benzoate CoA-ligase; *boxAB*, benzoyl-CoA 2,3-epoxidase; *boxC*, 2,3-epoxybenzoyl-CoA dihydrolase; and *boxD*, 3,4-dehydroadipyl-CoA semialdehyde dehydrogenase (NADP(+)).

Population Roles in the Biphenyl-Degrading Consortium

The catabolic pathways for biphenyl and its metabolic derivatives found in the metagenome of the biphenyl-degrading consortium and the genus affiliation of the coding sequences for these pathways (**Table 2** and Supplementary File 6) provide a complete understanding of the different roles of the main bacterial populations that are present in the consortium with regard of their relative abundance. It is interesting to note that the seven most represented genera in the consortium have been identified as the source of 90% of the CDSs identified in the biphenyl/benzoate degradation pathways and that these genera harbor all the enzymatic activities in the degradation pathways. These results reflect a high degree of functional redundancy, as the same reactions seem to be carried out by different taxa. These results are summarized in **Table 3** and the metabolic pathways reconstructed for the biphenyl-degrading consortium is represented in **Figure 4**. *Rhodococcus* is the genus responsible for initiating the biphenyl degradation into benzoate as the three BphA1 that have been found in the metagenome have been only assigned to this genus. Furthermore, the presence of complete gene clusters for *bph*, *etb*, and *nar* in a single *Rhodococcus* strain, and the previous reports of the involvement of these clusters in both biphenyl and PCBs degradation (Kimura et al., 2006; Iwasaki et al., 2007; Taguchi et al., 2007), makes this strain suited for bioremediation of PCBs. However, although the consortium was not able to grow in any of the chlorobenzoates tested (2-, 3-, or 4-chlorobenzoic acid) as the sole carbon and energy source (**Table 4**), cometabolism of chlorobenzoates as well as PCB congeners should be further analyzed. After formation of benzoate as the product of biphenyl degradation, the remaining bacterial populations can thrive, either by using benzoate, catechol, or protocatechuate. Our results show that protocatechuate and catechol degradative pathways in the consortium are rather abundant (**Table 2**), and are dominated by *Pseudomonas* and *Bordetella*, harboring genes for both, ortho and meta cleavage of protocatechuate and ortho cleavage of catechol. The relative high abundance of this genus in the consortium can be explained by the different alternative pathways for benzoate and its metabolic derivates degradation. Other genera such as *Achromobacter* and *Cupriavidus* are likely using catechol and/or protocatechuate to grow (**Table 3**). In addition, the consortium was able to grow on benzoate and protocatechuate as the sole carbon and energy source (**Table 4**), which is in agreement with the results presented here. On the other hand, the benzoate degradative pathway via acetyl-CoA ligation was mainly assigned to *Achromobacter*, which explains its presence in the consortium although it could also use protocatechuate and catechol via ortho cleavage (**Table 3**).

TABLE 4 | Consortium growth on different organic compounds as the sole carbon and energy source.

Substrate	Growth
Biphenyl	+
Benzoic acid	+
Protocatechuic acid	+
2-Chlorobenzoate	–
3-Chlorobenzoate	–
4-Chlorobenzoate	–

Interestingly, two of the most abundant genera within the consortium, *Pigmentiphaga* and *Stenotrophomonas* (20.54 and 8.57% 16S rRNA relative abundance, respectively) do not have any of the coding sequences for enzymes screened in the metagenome (**Table 2** and Supplementary File 6). In the case of *Pigmentiphaga*, it is clear that the lack of sequenced genomes available on the NCBI database (on April 2017) prevented the affiliation of CDSs to this genus. However, it is unclear if any of the coding sequences for enzymes of these pathways that could not been assigned to any genus (**Table 2**) might belong to *Pigmentiphaga* or if other metabolic abilities are involved. Regarding *Stenotrophomonas*, it is a common member of biphenyl, PCBs, and other aromatics-degrading communities (Leigh et al., 2007; Uhlik et al., 2013; Wald et al., 2015) and exhibits high metabolic versatility (Hauben et al., 1999). Its presence in the biphenyl-degrading consortium might be explained by cross-feeding on secondary metabolites produced by the rest of the consortium members, as it has been previously suggested (Wald et al., 2015). These results show that the metagenomic analysis of this consortium allows the determination of the biodegradation network involved in biphenyl degradation, being able to determine the specific role of different bacterial populations in the biodegradation process. The combination of these data with transcriptomic/proteomic

and metabolomic approaches could result in robust models of biodegradation processes, explaining the metabolic fluxes. This approach is also a proof of concept of the possibility of generating rationally designed inoculants for environmental restoration. Consortia, as this described here, can be thoroughly characterized and could be used as an inoculant, as a source of novel bioremediation strains or as a background for bioaugmentation with previously isolated strains.

The results presented here show that metagenomic analysis is a powerful tool for the functional characterization of consortia designed for bioremediation of complex contaminants. The analysis of consortia rather than soil microcosms has obvious advantages. First of all, while a typical soil microcosm usually contains thousands of genotypes, a consortium such as the one shown here contains less than a hundred genotypes, and therefore the depth of sequencing is much higher. Furthermore, while most of the genotypes detected in the consortium play a role in the biodegradation process, as shown here, most of the populations in a microcosm are irrelevant for the process. Furthermore, metagenomic analysis has proven to be advantageous over SIP in analyzing the biodegrading populations. While SIP was able to identify the bacterial populations involved in biphenyl and benzoate degradation in a soil microcosm and to determine that biphenyl and benzoate were mostly degraded by different populations (Leewis et al., 2016), here we have been able to determine not only the biodegrading populations, but also to assign specific functions and reactions to specific populations, identifying all the biodegradation pathways and therefore providing a deeper insight in the biodegradation process.

AUTHOR CONTRIBUTIONS

DG-S and JM performed the experiments and bioinformatic analysis. MM, MR-N, and RR designed the study and supervised the work. DG-S and RR drafted the manuscript.

FUNDING

This research was funded by grant BIO2015-64480-R from MINECO/FEDER EU. DG-S was granted by FPU fellowship program (FPU14/03965) from Ministerio de Educación, Cultura y Deporte, Spain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00232/full#supplementary-material

REFERENCES

Abbey, A. M., Beaudette, L. A., Lee, H., and Trevors, J. T. (2003). Polychlorinated biphenyl (PCB) degradation and persistence of a gfp-marked *Ralstonia eutropha* H850 in PCB-contaminated soil. *Appl. Microbiol. Biotechnol.* 63, 222–230. doi: 10.1007/s00253-003-1380-x

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bedard, D. L., Unterman, R., Bopp, L. H., Brennan, M. J., Haberl, M. L., and Johnson, C. (1986). Rapid assay for screening and characterizing microorganisms for the ability to degrade polychlorinated biphenyls. *Appl. Environ. Microbiol.* 51, 761–768.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brazil, G. M., Kenefick, L., Callanan, M., Haro, A., De Lorenzo, V., Dowling, D. N., et al. (1995). Construction of a rhizosphere pseudomonad with potential to degrade polychlorinated biphenyls and detection of bph gene expression in the rhizosphere. *Appl. Environ. Microbiol.* 61, 1946–1952.

Brenner, V., Arensdorf, J. J., and Focht, D. D. (1994). Genetic construction of PCB degraders. *Biodegradation* 5, 359–377. doi: 10.1007/BF00696470

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Camara, B., Herrera, C., Gonzalez, M., Couve, E., Hofer, B., and Seeger, M. (2004). From PCBs to highly toxic metabolites by the biphenyl pathway. *Environ. Microbiol.* 6, 842–850. doi: 10.1111/j.1462-2920.2004.00630.x

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Dai, S., Vaillancourt, F. H., Maaroufi, H., Drouin, N. M., Neau, D. B., Snieckus, V., et al. (2002). Identification and analysis of a bottleneck in PCB biodegradation. *Nat. Struct. Biol.* 9, 934–939. doi: 10.1038/nsb866

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Fava, F., Bertin, L., Fedi, S., and Zannoni, D. (2003). Methyl-beta-cyclodextrin-enhanced solubilization and aerobic biodegradation of polychlorinated biphenyls in two aged-contaminated soils. *Biotechnol. Bioeng.* 81, 381–390. doi: 10.1002/bit.10579

Fava, F., Di Gioia, D., Cinti, S., Marchetti, L., and Quattroni, G. (1994). Degradation and dechlorination of low-chlorinated biphenyls by a three-membered bacterial co-culture. *Appl. Microbiol. Biotechnol.* 41, 117–123. doi: 10.1007/BF00166092

Fennell, D. E., Nijenhuis, I., Wilson, S. F., Zinder, S. H., and Haggblom, M. M. (2004). Dehalococcoides ethenogenes strain 195 reductively dechlorinates diverse chlorinated aromatic pollutants. *Environ. Sci. Technol.* 38, 2075–2081. doi: 10.1021/es034989b

Field, J. A., and Sierra-Alvarez, R. (2008). Microbial transformation and degradation of polychlorinated biphenyls. *Environ. Pollut.* 155, 1–12. doi: 10.1016/j.envpol.2007.10.016

Fuchs, G., Boll, M., and Heider, J. (2011). Microbial degradation of aromatic compounds - from one strategy to four. *Nat. Rev. Microbiol.* 9, 803–816. doi: 10.1038/nrmicro2652

Furukawa, K., and Fujihara, H. (2008). Microbial degradation of polychlorinated biphenyls: biochemical and molecular features. *J. Biosci. Bioeng.* 105, 433–449. doi: 10.1263/jbb.105.433

Gerhardt, K. E., Huang, X.-D., Glick, B. R., and Greenberg, B. M. (2009). Phytoremediation and rhizoremediation of organic soil contaminants: potential and challenges. *Plant Sci.* 176, 20–30. doi: 10.1016/j.plantsci.2008.09.014

Gescher, J., Zaar, A., Mohamed, M., Schagger, H., and Fuchs, G. (2002). Genes coding for a new pathway of aerobic benzoate metabolism in *Azoarcus evansii*. *J. Bacteriol.* 184, 6301–6315. doi: 10.1128/JB.184.22.6301-6315.2002

Gibson, D. T., and Parales, R. E. (2000). Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr. Opin. Biotechnol.* 11, 236–243. doi: 10.1016/S0958-1669(00)00090-2

Gomes, H. I., Dias-Ferreira, C., and Ribeiro, A. B. (2013). Overview of in situ and ex situ remediation technologies for PCB-contaminated soils and sediments and obstacles for full-scale application. *Sci. Total Environ.* 445–446, 237–260. doi: 10.1016/j.scitotenv.2012.11.098

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Haddock, J. D., Horton, J. R., and Gibson, D. T. (1995). Dihydroxylation and dechlorination of chlorinated biphenyls by purified biphenyl 2,3-dioxygenase from *Pseudomonas* sp. strain LB400. *J. Bacteriol.* 177, 20–26. doi: 10.1128/jb.177.1.20-26.1995

Haluska, L., Barancikova, G., Balaz, S., Dercova, K., Vrana, B., Paz-Weisshaar, M., et al. (1995). Degradation of PCB in different soils by inoculated *Alcaligenes xylosoxidans*. *Sci. Total Environ.* 175, 275–285. doi: 10.1016/0048-9697(95)04927-4

Harkness, M. R., Mcdermott, J. B., Abramowicz, D. A., Salvo, J. J., Flanagan, W. P., Stephens, M. L., et al. (1993). In situ stimulation of aerobic PCB biodegradation in Hudson River sediments. *Science* 259, 503–507. doi: 10.1126/science.8424172

Harwood, C. S., and Parales, R. E. (1996). The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* 50, 553–590. doi: 10.1146/annurev.micro.50.1.553

Hauben, L., Vauterin, L., Moore, E. R., Hoste, B., and Swings, J. (1999). Genomic diversity of the genus *Stenotrophomonas*. *Int. J. Syst. Bacteriol.* 49(Pt 4), 1749–1760. doi: 10.1099/00207713-49-4-1749

Hernandez-Sanchez, V., Lang, E., and Wittich, R. M. (2013). The three-species consortium of genetically improved strains *Cupriavidus necator* RW112, *Burkholderia xenovorans* RW118, and *Pseudomonas pseudoalcaligenes* RW120 Grows with Technical Polychlorobiphenyl, Aroclor 1242. *Front. Microbiol.* 4:90. doi: 10.3389/fmicb.2013.00090

Hsieh, T. C., Ma, K. H., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* 7, 1451–1456. doi: 10.1111/2041-210X.12613

Iwasaki, T., Miyauchi, K., Masai, E., and Fukuda, M. (2006). Multiple-subunit genes of the aromatic-ring-hydroxylating dioxygenase play an active role in biphenyl and polychlorinated biphenyl degradation in *Rhodococcus* sp. *strain RHA*1. *Appl. Environ. Microbiol.* 72, 5396–5402. doi: 10.1128/AEM.00298-06

Iwasaki, T., Takeda, H., Miyauchi, K., Yamada, T., Masai, E., and Fukuda, M. (2007). Characterization of two biphenyl dioxygenases for biphenyl/PCB degradation in A PCB degrader, *Rhodococcus* sp. strain RHA1. *Biosci. Biotechnol. Biochem.* 71, 993–1002. doi: 10.1271/bbb.60663

Kimura, N., Kitagawa, W., Mori, T., Nakashima, N., Tamura, T., and Kamagata, Y. (2006). Genetic and biochemical characterization of the dioxygenase involved in lateral dioxygenation of dibenzofuran from *Rhodococcus opacus* strain SAO101. *Appl. Microbiol. Biotechnol.* 73, 474–484. doi: 10.1007/s00253-006-0481-8

Kimura, N., and Urushigawa, Y. (2001). Metabolism of dibenzo-p-dioxin and chlorinated dibenzo-p-dioxin by a gram-positive bacterium, *Rhodococcus opacus* SAO101. *J. Biosci. Bioeng.* 92, 138–143. doi: 10.1016/S1389-1723(01)80214-0

Kitagawa, W., Kimura, N., and Kamagata, Y. (2004). A novel p-nitrophenol degradation gene cluster from a gram-positive bacterium, *Rhodococcus opacus* SAO101. *J. Bacteriol.* 186, 4894–4902. doi: 10.1128/JB.186.15.4894-4902.2004

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lauby-Secretan, B., Loomis, D., Grosse, Y., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., et al. (2013). Carcinogenicity of polychlorinated biphenyls and polybrominated biphenyls. *Lancet Oncol.* 14, 287–288. doi: 10.1016/S1470-2045(13)70104-9

Leewis, M. C., Uhlik, O., and Leigh, M. B. (2016). Synergistic processing of Biphenyl and Benzoate: carbon flow through the bacterial community in polychlorinated-biphenyl-contaminated soil. *Sci. Rep.* 6:22145. doi: 10.1038/srep22145

Leigh, M. B., Pellizari, V. H., Uhlik, O., Sutka, R., Rodrigues, J., Ostrom, N. E., et al. (2007). Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *ISME J.* 1, 134–148. doi: 10.1038/ismej.2007.26

Leigh, M. B., Prouzova, P., Mackova, M., Macek, T., Nagle, D. P., and Fletcher, J. S. (2006). Polychlorinated biphenyl (PCB)-degrading bacteria associated with trees in a PCB-contaminated site. *Appl. Environ. Microbiol.* 72, 2331–2342. doi: 10.1128/AEM.72.4.2331-2342.2006

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Mayes, B. A., Mcconnell, E. E., Neal, B. H., Brunner, M. J., Hamilton, S. B., Sullivan, T. M., et al. (1998). Comparative carcinogenicity in Sprague-Dawley rats of the polychlorinated biphenyl mixtures Aroclors 1016, 1242, 1254, and 1260. *Toxicol. Sci.* 41, 62–76.

Mondello, F. J., Turcich, M. P., Lobos, J. H., and Erickson, B. D. (1997). Identification and modification of biphenyl dioxygenase sequences that determine the specificity of polychlorinated biphenyl degradation. *Appl. Environ. Microbiol.* 63, 3096–3103.

National Research Council (1979). *Polychlorinated Biphenyls*. Washington, DC: The National Academies Press.

Ohtsubo, Y., Kudo, T., Tsuda, M., and Nagata, Y. (2004). Strategies for bioremediation of polychlorinated biphenyls. *Appl. Microbiol. Biotechnol.* 65, 250–258. doi: 10.1007/s00253-004-1654-y

Pieper, D. H. (2005). Aerobic degradation of polychlorinated biphenyls. *Appl. Microbiol. Biotechnol.* 67, 170–191. doi: 10.1007/s00253-004-1810-4

Pieper, D. H., and Seeger, M. (2008). Bacterial metabolism of polychlorinated biphenyls. *J. Mol. Microbiol. Biotechnol.* 15, 121–138. doi: 10.1159/000121325

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Quensen, J. F. III, Tiedje, J. M., and Boyd, S. A. (1988). Reductive dechlorination of polychlorinated biphenyls by anaerobic microorganisms from sediments. *Science* 242, 752–754. doi: 10.1126/science.242.4879.752

Quinete, N., Schettgen, T., Bertram, J., and Kraus, T. (2014). Occurrence and distribution of PCB metabolites in blood and their potential health effects in humans: a review. *Environ. Sci. Pollut. Res. Int.* 21, 11951–11972. doi: 10.1007/s11356-014-3136-9

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rather, L. J., Knapp, B., Haehnel, W., and Fuchs, G. (2010). Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *J. Biol. Chem.* 285, 20615–20624. doi: 10.1074/jbc.M110.124156

Ross, G. (2004). The public health implications of polychlorinated biphenyls (PCBs) in the environment. *Ecotoxicol. Environ. Saf.* 59, 275–291. doi: 10.1016/j.ecoenv.2004.06.003

Saavedra, J. M., Acevedo, F., Gonzalez, M., and Seeger, M. (2010). Mineralization of PCBs by the genetically modified strain *Cupriavidus necator* JMS34 and its application for bioremediation of PCBs in soil. *Appl. Microbiol. Biotechnol.* 87, 1543–1554. doi: 10.1007/s00253-010-2575-6

Seeger, M., Camara, B., and Hofer, B. (2001). Dehalogenation, denitration, dehydroxylation, and angular attack on substituted biphenyls and related compounds by a biphenyl dioxygenase. *J. Bacteriol.* 183, 3548–3555. doi: 10.1128/JB.183.12.3548-3555.2001

Seeger, M., Timmis, K. N., and Hofer, B. (1995). Conversion of chlorobiphenyls into phenylhexadienoates and benzoates by the enzymes of the upper pathway for polychlorobiphenyl degradation encoded by the bph locus of *Pseudomonas* sp. strain LB400. *Appl. Environ. Microbiol.* 61, 2654–2658.

Seto, M., Kimbara, K., Shimura, M., Hatta, T., Fukuda, M., and Yano, K. (1995). A novel transformation of polychlorinated biphenyls by Rhodococcus sp. Strain RHA1. *Appl. Environ. Microbiol.* 61, 3353–3358.

Sharma, J. K., Gautam, R. K., Misra, R. R., Kashyap, S. M., Singh, S. K., and Juwarkar, A. A. (2014). Degradation of Di- through hepta-chlorobiphenyls in clophen oil using microorganisms isolated from long term PCBs contaminated soil. *Indian J. Microbiol.* 54, 337–342. doi: 10.1007/s12088-014-0459-7

Sharma, J. K., Gautam, R. K., Nanekar, S. V., Weber, R., Singh, B. K., Singh, S. K., et al. (2017). Advances and perspective in bioremediation of polychlorinated

biphenyl-contaminated soils. *Environ. Sci. Pollut. Res. Int.* doi: 10.1007/s11356-017-8995-4 [Epub ahead of print].

Sierra, I., Valera, J. L., Marina, M. L., and Laborda, F. (2003). Study of the biodegradation process of polychlorinated biphenyls in liquid medium and soil by a new isolated aerobic bacterium (Janibacter sp.). *Chemosphere* 53, 609–618. doi: 10.1016/S0045-6535(03)00418-1

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Singer, A. C., Gilbert, E. S., Luepromchai, E., and Crowley, D. E. (2000). Bioremediation of polychlorinated biphenyl-contaminated soil using carvone and surfactant-grown bacteria. *Appl. Microbiol. Biotechnol.* 54, 838–843. doi: 10.1007/s002530000472

Taguchi, K., Motoyama, M., Iida, T., and Kudo, T. (2007). Polychlorinated biphenyl/biphenyl degrading gene clusters in *Rhodococcus* sp. K37, HA99, and TA431 are different from well-known *bph* gene clusters of Rhodococci. *Biosci. Biotechnol. Biochem.* 71, 1136–1144. doi: 10.1271/bbb.60551

Taira, K., Hirose, J., Hayashida, S., and Furukawa, K. (1992). Analysis of bph operon from the polychlorinated biphenyl-degrading strain of *Pseudomonas pseudoalcaligenes* KF707. *J. Biol. Chem.* 267, 4844–4853.

Turrio-Baldassarri, L., Abate, V., Alivernini, S., Battistelli, C. L., Carasi, S., Casella, M., et al. (2007). A study on PCB, PCDD/PCDF industrial contamination in a mixed urban-agricultural area significantly affecting the food chain and the human exposure. Part I: soil and feed. *Chemosphere* 67, 1822–1830. doi: 10.1016/j.chemosphere.2006.05.124

Uhlik, O., Jecna, K., Mackova, M., Vlcek, C., Hroudova, M., Demnerova, K., et al. (2009). Biphenyl-metabolizing bacteria in the rhizosphere of horseradish and bulk soil contaminated by polychlorinated biphenyls as revealed by stable isotope probing. *Appl. Environ. Microbiol.* 75, 6471–6477. doi: 10.1128/AEM.00466-09

Uhlik, O., Musilova, L., Ridl, J., Hroudova, M., Vlcek, C., Koubek, J., et al. (2013). Plant secondary metabolite-induced shifts in bacterial community structure and degradative ability in contaminated soil. *Appl. Microbiol. Biotechnol.* 97, 9245–9256. doi: 10.1007/s00253-012-4627-6

Vergani, L., Mapelli, F., Marasco, R., Crotti, E., Fusi, M., Di Guardo, A., et al. (2017a). Bacteria associated to plants naturally selected in a historical PCB polluted soil show potential to sustain natural attenuation. *Front. Microbiol.* 8:1385. doi: 10.3389/fmicb.2017.01385

Vergani, L., Mapelli, F., Zanardini, E., Terzaghi, E., Di Guardo, A., Morosini, C., et al. (2017b). Phyto-rhizoremediation of polychlorinated biphenyl contaminated soils: an outlook on plant-microbe beneficial interactions. *Sci. Total Environ.* 575, 1395–1406. doi: 10.1016/j.scitotenv.2016.09.218

Villacieros, M., Whelan, C., Mackova, M., Molgaard, J., Sanchez-Contreras, M., Lloret, J., et al. (2005). Polychlorinated biphenyl rhizoremediation by *Pseudomonas fluorescens* F113 derivatives, using a Sinorhizobium meliloti nod system to drive bph gene expression. *Appl. Environ. Microbiol.* 71, 2687–2694. doi: 10.1128/AEM.71.5.2687-2694.2005

Wald, J., Hroudova, M., Jansa, J., Vrchotova, B., Macek, T., and Uhlik, O. (2015). Pseudomonads rule degradation of polyaromatic hydrocarbons in aerated sediment. *Front. Microbiol.* 6:1268. doi: 10.3389/fmicb.2015.01268

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703. doi: 10.1128/jb.173.2.697-703.1991

Yang, X., Liu, X., Song, L., Xie, F., Zhang, G., and Qian, S. (2007). Characterization and functional analysis of a novel gene cluster involved in biphenyl degradation in *Rhodococcus* sp. strain R04. *J. Appl. Microbiol.* 103, 2214–2224. doi: 10.1111/j.1365-2672.2007.03461.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Garrido-Sanz, Manzano, Martín, Redondo-Nieto and Rivilla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 November 2018
doi: 10.3389/fmicb.2018.02815

Opportunistic Bacteria Dominate the Soil Microbiome Response to Phenanthrene in a Microcosm-Based Study

Sean Storey[1,2], *Mardiana Mohd Ashaari*[3], *Nicholas Clipson*[1,2], *Evelyn Doyle*[1,2] *and Alexandre B. de Menezes*[4*]

[1] School of Biology and Environmental Science, University College Dublin, Dublin, Ireland, [2] Earth Institute, University College Dublin, Dublin, Ireland, [3] Department of Biotechnology, Kulliyah of Science, International Islamic University Malaysia, Malaysia, Malaysia, [4] Microbiology, School of Natural Sciences, Ryan Institute, National University of Ireland, Galway, Ireland

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Wenli Chen, Huazhong Agricultural University, China
David Correa Galeote, Universidad Nacional Autónoma de México, Mexico

***Correspondence:**
Alexandre B. de Menezes
alexandre.demenezes@nuigalway.ie

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *13 December 2017*
Accepted: *02 November 2018*
Published: *21 November 2018*

Citation:
Storey S, Ashaari MM, Clipson N, Doyle E and de Menezes AB (2018) Opportunistic Bacteria Dominate the Soil Microbiome Response to Phenanthrene in a Microcosm-Based Study. Front. Microbiol. 9:2815. doi: 10.3389/fmicb.2018.02815

Bioremediation offers a sustainable approach for removal of polycyclic aromatic hydrocarbons (PAHs) from the environment; however, information regarding the microbial communities involved remains limited. In this study, microbial community dynamics and the abundance of the key gene (PAH-RHDα) encoding a ring hydroxylating dioxygenase involved in PAH degradation were examined during degradation of phenanthrene in a podzolic soil from the site of a former timber treatment facility. The 10,000-fold greater abundance of this gene associated with Gram-positive bacteria found in phenanthrene-amended soil compared to unamended soil indicated the likely role of Gram-positive bacteria in PAH degradation. In contrast, the abundance of the Gram-negative PAHs-RHDα gene was very low throughout the experiment. While phenanthrene induced increases in the abundance of a small number of OTUs from the Actinomycetales and Sphingomonadale, most of the remainder of the community remained stable. A single unclassified OTU from the *Micrococcaceae* family increased ∼20-fold in relative abundance, reaching 32% of the total sequences in amended microcosms on day 7 of the experiment. The relative abundance of this same OTU increased 4.5-fold in unamended soils, and a similar pattern was observed for the second most abundant PAH-responsive OTU, classified into the *Sphingomonas* genus. Furthermore, the relative abundance of both of these OTUs decreased substantially between days 7 and 17 in the phenanthrene-amended and control microcosms. This suggests that their opportunistic phenotype, in addition to likely PAH-degrading ability, was determinant in the vigorous growth of dominant PAH-responsive OTUs following phenanthrene amendment. This study provides new information on the temporal response of soil microbial communities to the presence and degradation of a significant environmental pollutant, and as such has the potential to inform the design of PAH bioremediation protocols.

Keywords: polycyclic aromatic hydrocarbons, microbiome, bioremediation, soil, phenanthrene

INTRODUCTION

Soil pollution is a global problem; there are about 342,000 contaminated sites in Europe alone, with this number set to increase to over 500,000 by 2025 (Anon, 2012). Polycyclic aromatic hydrocarbons (PAHs) are a class of toxic chemicals, composed of at least two benzene rings arranged in various conformations (Wickliffe et al., 2014). PAHs are released to the environment upon incomplete combustion of fossil fuels and are ubiquitous in nature, with soil as their major sink (Collins et al., 2013). Due to their carcinogenic and mutagenic potential, 16 of these compounds are listed as priority pollutants by the United States Environmental Protection Agency (Wang et al., 2010).

Bacterial communities are important components of soil environments and are essential in the provision of many ecosystem services (Brussaard, 2012), including degradation of toxic chemicals such as PAHs (Fernández-Luqueño et al., 2011). However, the hydrophobicity of PAHs renders them poorly available to microorganisms, leading to their persistence in the environment (Posada-Baquero and Ortega-Calvo, 2011). A wide range of bacteria and fungi capable of degrading PAHs have been isolated from contaminated sites (for reviews see Cerniglia, 1992, 1997; Doyle et al., 2008; Ghosal et al., 2016), with *Pseudomonas, Sphingomonas,* and *Mycobacterium* spp. amongst the most frequently isolated PAH degrading bacteria (Bastiaens et al., 2000; Johnsen and Karlson, 2005). Culture independent analyses have revealed that representatives of Proteobacteria, Actinobacteria, and Firmicutes are the bacteria most likely to increase during biodegradation of PAHs and petroleum hydrocarbons (Fuentes et al., 2014).

Although bioremediation approaches have been successfully applied to PAH-contaminated soils both in the laboratory (Sayara et al., 2011) and in the field (Lors et al., 2012), the process remains poorly understood, mainly in terms of the structure and function of microbial assemblages involved (Yang et al., 2015). PAH biodegradation studies have been analyzed using different molecular techniques (Muckian et al., 2009; Lors et al., 2010; Sun et al., 2012; Festa et al., 2016; Kuppusamy et al., 2016). In general, individual molecular microbial studies of PAH degradation targeted complex bioremediation conditions. For example, Jiao et al. (2017) showed that a variety of pollutants had consistent, broadly defined effects on soil microbial taxa, mostly repressing or inducing specific microbial populations. Zhao et al. (2016) showed that *Mycobacterium* contribute ring-hydroxylating dioxygenases involved in the initial steps of fluoranthene breakdown, while a more diverse group of bacteria contributed to the metabolism of downstream fluoranthene degradation products. Kuppusamy et al. (2016) demonstrated that metal-resistant, PAH-degrading Alphaproteobacteria can persist for longer in soils contaminated with heavy metals and PAHs compared to Gram-positive bacteria such as the Actinobacteria. Certain amendments can increase bioremediation efficiency, for example Wang et al. (2016) showed that surfactants increased PAH removal from soil and boosted the abundance of the *Pseudomonas*, *Bacillus* and *Sphingomonas*. The inoculation of PAH-degrading consortia into polluted soil can also enhance bioremediation and boost the abundance of genes associated with formation of PAH degradation products that can be further metabolized through the tricarboxylic acid cycle (Zafra et al., 2016). Festa et al. (2016) showed that inoculation of PAH-degrading *Sphingobium* increases the degradation of phenanthrene in soil but had no effect in the degradation of PAHs in a chronically contaminated soil. Other studies found less clear relationships between soil amendments, soil microbial communities and PAH degradation. For example, Thomas and Cébron (2016) found that phenanthrene amendment in the presence of plants did not lead to greater removal of phenanthrene from bulk soil, despite increases in the abundance of PAH-ring hydroxylating genes. Similarly, Delgado-Balbuena et al. (2016) did not find strong relationships between soil microbial community structure and anthracene removal from contaminated soil when applying different remediation strategies. Although these studies offer valuable insights into the microbial molecular ecology of PAH degradation in soils, they vary substantially in their objectives, target habitat, PAH type, and experimental design. Currently, studies examining the effects of a single PAH in the native soil microbiome without the confounding effects of multiple treatments and contaminants are limited.

PAH contamination causes physiological stress on the soil microbial communities and exposure to these compounds leads to the activation of detoxification and stress resistance (de Menezes et al., 2012). Therefore, PAH exposure not only triggers the growth of PAH-degrading microorganisms, but it may also affect the ecological stability of the soil microbiome, in turn potentially affecting a range of soil processes (Griffiths and Philipot, 2013). Microbial ecosystem stability is affected by species richness, evenness, and composition, however greater biodiversity levels are not necessarily a sign of a more stable ecosystem (Griffiths and Philipot, 2013; Shade, 2017). While the intermediate disturbance hypothesis was postulated to explain the frequent observation in macro-ecology of maximal biodiversity levels at intermediate levels of disturbance, the applicability of this hypothesis to microbial ecosystems is less certain (Gibbons et al., 2016). In microbial communities, disturbance is thought to cause changes in community composition, leading to a succession of species in which opportunistic, fast growing bacteria able to take advantage of transient conditions thrive initially, followed by slower growing, resource efficient microorganisms (Sigler and Zeyer, 2004). PAH contamination in soil therefore represents an ideal opportunity to study the effects of a disturbance on microbial community composition and succession.

Bacterial diversity in soil amended with the 3-ring PAH phenanthrene was compared with that in unamended soil during the course of phenanthrene degradation using 16S rRNA gene amplicon sequencing. The soil used in this experiment had previous history of exposure to PAH pollution due to the activities of a since-closed timber treatment facility, however at the time of sampling PAH levels in this soil were similar to uncontaminated soils. Although a previous study by de Menezes et al. (2012) examined gene expression in a phenanthrene-contaminated soil, only one time-point was examined and

the effect of the PAH on soil microbial community structure was not assessed. We set up two sets of microcosms in triplicate, of which one set was amended with phenanthrene and the other set used as the unamended control. These microcosms were sampled at four time points during a 17-day incubation during which time phenanthrene levels were measured and bacterial diversity analyzed by high-throughput sequencing of the 16S rRNA gene. We also assessed the potential for PAH degradation in these soils by quantifying the abundance of the gene encoding the α subunit of a PAH ring hydroxylating dioxygenase (PAH-RHDα), a key gene involved in PAH degradation. Compared to the complex experimental design of most microbiome PAH bioremediation studies, the simple experimental design used in this study allows the investigation of phenanthrene as the single treatment, without the confounding effects of multiple treatments. We observed that dominant phenanthrene-responsive bacteria showed opportunistic traits, increasing substantially in abundance in the amended samples but also to a lesser extent in the control samples, likely as a result of their capacity to adapt and grow in ecologically disturbed environments.

MATERIALS AND METHODS

Soil Collection, Microcosm Setup, and Sampling

Soil was collected from the site of a former timber treatment facility in Monard, Co. Tipperary, Ireland (52°30' N, 8°13' W). This site is elevated about 100 m above sea level and contains a mixture of gray and gray-brown podzolic soil with underlying limestone glacial till and limestone. Soil chemical analysis was carried out by ALcontrol Geochem, Dublin, Ireland, while soil particle size was determined following the method of Kettler et al. (2001). The soil had a moisture content of 18.45%, a pH of 6.58 and a total organic carbon content of 2% (w/v). The concentration of total background PAHs was 4.7 mg kg^{-1} (standard deviation 4.6%) as determined by GC-FID (Sawulski et al., 2014). The soil is a clay fine loamy drift with limestone soil, composed of 51.95% sand, 18.13% silt, and 29.91% and it is classified as luvisol.

A total of 24 microcosms were set up, 12 for each treatment (phenanthrene amendment and control) and 3 for each time point (de Menezes et al., 2012). Prior to microcosm set up, the soil was sieved to < 2 mm to remove plant matter and debris and amended with phenanthrene to a final concentration of 725 mg kg^{-1}. Fifty grams of unamended or amended soil was then placed in black polyvinylchloride containers in triplicate and incubated at 22°C for the duration of the experiment (17 days). Soil moisture content was maintained at a constant level by gravimetric addition of sterile deionized water every 2–3 days. Pots were destructively sampled on days 0, 2, 7, and 17.

Phenanthrene Extraction and Analysis

Phenanthrene was extracted from soil by adding acetone and hexane to soil samples and using mechanical agitation (Dean and Xiong, 2000). Phenanthrene concentration was determined by Gas Chromatography-Flame Ionization Detection as described by Storey et al. (2014).

DNA Extraction

DNA was extracted from triplicates of each treatment (3 microcosm pots were destructively harvested for each treatment at each time point) on days 0, 2, 7, and 17 of the experiment as outlined in de Menezes et al. (2012) using a modification of the phenol-chloroform method of Griffiths et al. (2000).

Amplicon Sequencing

Amplicon sequencing was carried out using the procedure detailed by Martínez et al. (2009). Briefly, the V1-V3 regions of the 16S rRNA gene were amplified using 8F-518R primers (Lane et al., 1985; Muyzer et al., 1993) which contained specific adapter sequences and unique barcodes for each sample. GENETOOLS software (Syngene, Cambridge, UK) was used for quality control of PCR reactions. Equal amounts of amplicons from each PCR reaction were pooled, run on a gel to ensure purity and DNA concentrations then measured using the Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher, Scientific, MS, USA) and a Qubit fluorimeter (Thermo Fisher, Scientific, MS, USA). Pyrosequencing was performed using the Roche-454 Titanium platform at the University of Nebraska-Lincoln Core for Applied Genomics and Ecology. Sequence files associated with each sample were deposited in the NCBI GenBank database under the BioProject accession number PRJNA284664.

Sequences were processed using mothur v.1.31.0. with default parameters for 454-Titanium sequence processing (Schloss et al., 2009). Briefly, after removing sequence noise, sequences smaller than 200 bp, with 1 or more nucleotide ambiguities or > 8 bp long homopolymers were removed from the dataset. Sequences were aligned against the Silva reference alignment and those sequences classified as plastid, mitochondrial, archaeal, eukaryotic or unknown at the kingdom level were discarded. Chimeras were detected using the UCHIME tool built within mothur (Edgar et al., 2011) and also removed from the dataset. After quality processing the number of sequences per sample ranged from 2,572 to 11,096 sequences. Operational taxonomic units (OTUs) were generated by calculating pairwise distances and clustering sequences with a distance cutoff of 0.03. OTUs present as a single sequence read in the dataset were removed. The OTUs were classified by obtaining the consensus taxonomy in mothur using the RDP reference files and a consensus confidence threshold cutoff of 0.8 (Schloss, 2009).

Relative abundances were determined in mothur as the abundance of each OTU to the total number of sequences in a sample. For beta-diversity analyses, the number of sequences across samples was subsampled using the sub.sample command in mothur to the number of sequences present in the sample with smallest number of sequences. The microbial community composition of one sample from day two of the control microcosms was substantially different from all other samples in the dataset and was no more similar to its two biological replicates than to samples from the other treatments. We therefore chose to remove this sample from the analysis. In

order to identify individual OTUs that were enriched in PAH-amended microcosms, we used DESeq2, which has the advantage of not requiring rarefaction or subsampling of sequence data, a procedure that leads to loss of valid sequence information (McMurdie and Holmes, 2013, 2014; Love et al., 2014). DESeq2 tests were performed using the Wald test, automatic filtering of low abundance OTUs, and an alpha of 0.05 and multiple testing adjustment of *p*-values. Bacterial alpha diversity (Shannon index) was also calculated in Phyloseq, and paired t-tests using Bonferroni adjusted *p*-values were carried out to determine the significance of differences in Shannon index between treatments.

Quantitative PCR

To generate standard curves of the RHDα-GP and the PAH-RHDα-GN genes, both genes were amplified from soil using the primer sets described in Cébron et al. (2008). End-point PCRs were performed initially in 25-μl volumes containing 12.5 μl 2X PCR master mix (Promega), 2.5 pmoles each primer, 0.2 μl ultrapure BSA (50 mg ml−1, Ambion), and 10 ng DNA template. The thermocycling conditions were as follows: 95°C for 5 min (one cycle); 95°C for 30 s, 57 (GP) or 54 °C (GN) for 30 s, 72°C for 30 s (30 cycles); 72°C for 7 min. PCR products were visualized on a 1.2% (w/v) agarose gel. The resulting products were cloned using the pGEM-T Easy Vector system (Promega), and clones containing both plasmid and insert were isolated and purified using the PureYield™ Plasmid Miniprep system (Promega). Plasmid DNA was quantified using a Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific). The copy number of the PAH-RHDα-GP and the PAH-RHDα-GN gene per volume was calculated using recombinant plasmid sizes of 3307 and 3321 bp, respectively, and a mass of 1.096 × 10–21 g bp−1. Serial dilutions were used to construct standard curves.

qPCR was then carried out using the Applied Biosystems Viia 7 qPCR machine (Life Technologies, Dublin, Ireland) in MicroAmp®optical 96-well reaction plates containing 6.25 μl of 2 × KAPA SYBR® FAST qPCR master mix, 0.25 μl ROX™ low passive reference dye (Anachem, Bedfordshire, UK), 0.2 pmoles each primer and 2 ng DNA. Nuclease-free water (Sigma-Aldrich, Arklow, Ireland) was added to give a final volume of 12.5 μl. No-template controls contained 1 μl nuclease-free water (Sigma-Aldrich, Arklow, Ireland) in place of DNA template. Amplification was carried out using a modification of the method of Cébron et al. (2008) as follows: 95°C for 1 min, then 40 cycles of 95°C for 30 s, annealing temperature for 30 s, then elongation at 72°C for 30 s. SYBR® Green signal intensity was then measured during a 10 s primer dissociation step at 80°C. A melting curve was included at the end of each reaction using a temperature increment of 0.05°C s^{-1} from 51–95°C. Standard curves were included in each experiment. All analyses were performed using ABI VIIA 7 software version 1.2.2 (Life Technologies, Carlsbad, USA), Microsoft Excel 2010 and SAS version 9.1 (SAS, Cary, USA).

Data Analysis

Multivariate statistical analysis was carried out in PRIMER version 6.1.9 with PERMANOVA add-on version 1.0.1 (Primer-E Ltd., Plymouth, UK). Statistical analysis of the relative abundance of OTUs was assessed at genus level using PRIMER. Permutational multivariate analysis of variance (PERMANOVA) was performed using a Bray Curtis dissimilarity matrix in PRIMER using 9,999 unrestricted permutation of raw data. Analysis of variance (ANOVA) was carried out using SAS version 9.1. Principal coordinates analysis (PCoA) was performed using the capscale command in the vegan package in R (Oksanen et al., 2018), using square-root transformed genus-aggregated sequence abundances and a Bray-Curtis dissimilarity matrix, with overlaid arrows representing correlations > 0.4 between genera abundance profiles and PCoA ordination axes. Pearson correlation between selected OTUs and total abundance of the PAH-RHDα gene was calculated in R using the cor.test function.

RESULTS

Bacterial Community Response to Phenanthrene

The response of the indigenous soil bacterial community during phenanthrene degradation was examined using 454 amplicon sequencing. PERMANOVA analysis (**Table 1**) indicated that bacterial community structure changed both over time and in the presence of phenanthrene. Pairwise comparisons on bacterial community structures in phenanthrene amended soil revealed that this PAH had no significant effect on bacterial community structures on days 0 and 2, but bacterial community structures on day 7 of phenanthrene amended soil were significantly different from those on day 17. In addition, bacterial community structure in phenanthrene amended soil on days 7 and 17 were significantly different to those in the unamended control soil ($P < 0.001$; **Supplementary Table 1**). When temporal changes were examined, bacterial community structure was found to change significantly between each sampling day in the unamended control ($P < 0.001$), but no changes were observed in the phenanthrene-amended soil between days 0 and 2 ($P = 0.09$).

Bacterial diversity was similar in both unamended and phenanthrene-amended soils on day 0, with soils dominated by two phyla the Actinobacteria and the Proteobacteria (**Figure 1**). Although a small increase in the relative abundance of Actinobacteria and concurrent decrease in the relative abundance of the Proteobacteria was observed between days 0 and 2 in the unamended soil, these changes were not statistically significant. In the phenanthrene amended soil, degradation proceeded

TABLE 1 | PERMANOVA results for amplicon sequencing analysis of bacterial communities at different taxa levels in soil amended or unamended with phenanthrene on days 0, 2, 7, and 17.

Taxon	Time	Phenanthrene	Time* phenanthrene
Phylum	6.87**	12.11**	2.41*
Family	12.31**	10.96**	4.07**
Genus	11.29**	10.12**	4.19**
OTU	10.41**	9.17**	4.02**

*Pseudo-F values are given, with P-values in superscript. *p < 0.05; **p < 0.01.*

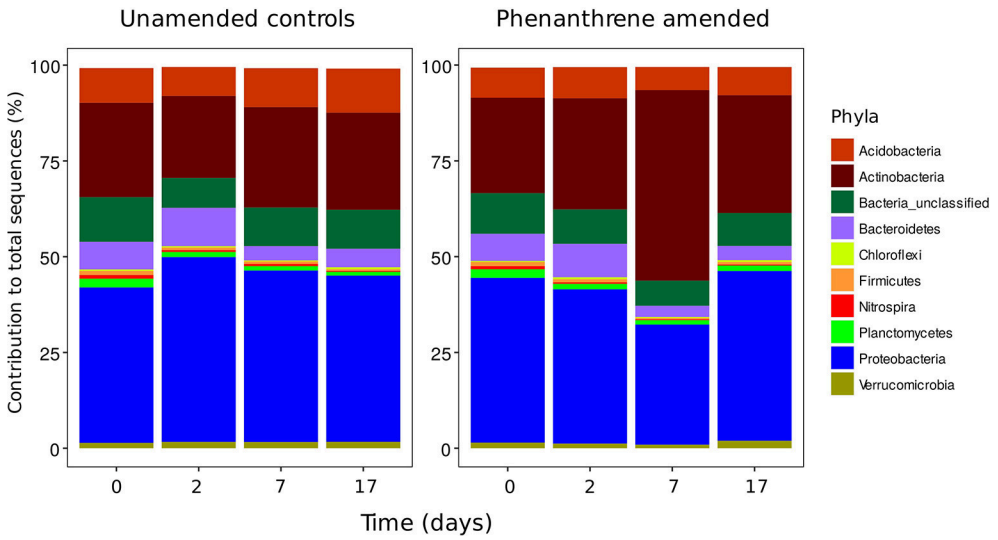


FIGURE 1 | Phyla contributing >1% to the total number of sequences in unamended or phenanthrene-amended microcosm soils on days 0, 2, 7, and 17.

rapidly following a brief lag period of approximately 5 days. 252 mg kg^{-1} (about 35%) of phenanthrene added remained after 7 days, with almost complete degradation observed by day 17 (de Menezes et al., 2012). In this soil, the relative abundance of the Actinobacteria was highest on day 7, representing 50% of the total amount of 16S rRNA gene sequences present. The relative abundance of this phylum was significantly ($P < 0.05$) lower in the unamended soil at the same time point. The relative abundance of sequences belonging to the Proteobacteria decreased significantly ($P < 0.05$) between days 2 and 7. By the end of the experiment (day 17), the relative abundance of the Actinobacteria in the phenanthrene-amended soil had returned to the levels observed on day 2, whereas, the relative abundance of the Proteobacteria returned to the levels observed initially (day 0).

ANOVA analysis showed that on day 7 unclassified *Micrococcaceae* and *Microbacteriaceae* were significantly ($P < 0.05$) relatively more abundant in phenanthrene-amended soils compared to the unamended control soil (**Table 2**). The unclassified *Micrococcaceae* and *Microbacteriaceae* increased, respectively from 9 to 32% and from 1.4 to 3% of the total 16S rRNA gene sequences in the amended soils. The relative abundance of many genera, including a range of unclassified Bacteria belonging to the *Alphaproteobacteria, Burkholderiales,* and *Bradyrhizobium* (all members of the Proteobacteria), declined significantly ($P < 0.05$) in the presence of phenanthrene. Interestingly, many genera were not significantly affected by the presence of phenanthrene at this time point.

Principal coordinates analysis (PCoA) was used to correlate changes in community structure at genus level with the presence or absence of phenanthrene (**Figure 2**). The genus *Flavobacterium* and unclassified *Actinobacteria* were correlated with day 0 samples for both control and phenanthrene-amended samples. Two days after the start of the experiment, control and phenanthrene-amended samples were not forming clearly distinct clusters, however the unclassified *Microbacteriaceae* correlated more strongly with day 2 phenanthrene-amended samples. On day 7, control and phenanthrene-amended samples can be seen to be well separated, with the unclassified *Micrococcaceae* clearly correlating with the phenanthrene-amended samples, while day 7 and 17 control samples correlated primarily with the unclassified *Myxococcales*, the *Acidobacteria* Gp6 and unclassified *Gammaproteobacteria*. On day 17, phenanthrene-amended samples cluster more closely with the day 7 and 17 control samples.

Although ANOVA showed which bacterial genera were more abundant at each treatment, it provided limited insight into changes in individual OTUs. DESeq2 analysis, which is a more appropriate statistical test for microbiome data than ANOVA, was therefore carried out to compare the differential abundance of individual OTUs in phenanthrene-amended to unamended soils at the same time points. No significantly differentially abundant OTUs were detected on either day 0 or day 2. However, on day 7, several OTUs classified to the Actinomycetales order, including *Mycobacterium* and unclassified *Micrococcaceae* were significantly more abundant in phenanthrene-amended soil than in its unamended counterpart (**Table 2**). One *Mycobacterium* OTU (OTU 82) was almost 200-fold more abundant on day 17 in phenanthrene-amended compared to unamended soil. OTUs classified to several other taxa, including *Sphingomonas*, unclassified *Microbacteraceae* and *Pseudoxanthomonas* were also significantly more abundant in the phenanthrene-amended soil compared to the unamended control soil. Interestingly, few OTUs responded negatively to phenanthrene, one unclassified OTU of the Burkholderiales and an unclassified Actinomycetales OTU were significantly less abundant in the phenanthrene-amended soil on day 17 of the experiment (**Table 2**).

TABLE 2 | Genera contributing >1 % to the total number of 16S rRNA genes sequences in phenanthrene-amended (PHE+) and unamended (PHE-) microcosm soils on day 7.

				Contribution to total sequences (%)			
Phylum	**Class**	**Family**	**Genus**	**PHE-**	**Standard error**	**PHE+**	**Standard error**
Actinobacteria	*Actinobacteria*	*Micrococcaceae*	*Micrococcaceae*_unclassified	9.35a	0.75	32.42b	1.39
*Bacteria*_unclassified	*Bacteria*_unclassified	Bacteria_unclassified	*Bacteria*_unclassified	6.73a	0.32	4.66b	0.09
Acidobacteria	*Acidobacteria*	*Acidobacteria Gp6* incertae sedis	*Acidobacteria Gp6* incertae sedis	7.19a	0.44	3.83b	0.05
Actinobacteria	*Actinobacteria*	*Actinomycetales*_unclassified	*Actinomycetales*_unclassified	4.09a	0.37	3.80a	0.07
Actinobacteria	*Actinobacteria*	*Actinobacteria*_unclassified	*Actinobacteria*_unclassified	4.47a	0.25	3.48a	0.23
Proteobacteria	*Alphaproteobacteria*	*Alphaproteobacteria*_unclassified	*Alphaproteobacteria*_unclassified	4.56a	0.36	3.32b	0.46
Actinobacteria	*Actinobacteria*	*Microbacteriaceae*	*Microbacteriaceae*_unclassified	1.42a	0.15	3.04b	0.31
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadaceae*	*Sphingomonas*	3.19a	0.06	2.90a	0.22
Actinobacteria	*Actinobacteria*	*Acidimicrobiaceae*	*Ilumatobacter*	2.65a	0.06	2.35b	0.23
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*_unclassified	*Rhizobiales*_unclassified	3.95a	0.09	2.30b	0.10
Proteobacteria	*Betaproteobacteria*	*Comamonadaceae*	*Comamonadaceae*_unclassified	1.45a	0.06	1.69b	0.03
Proteobacteria	*Gammaproteobacteria*	*Gammaproteobacteria*_unclassified	*Gammaproteobacteria*_unclassified	2.90a	0.26	1.56b	0.07
Proteobacteria	*Betaproteobacteria*	*Betaproteobacteria*_unclassified	*Betaproteobacteria*_unclassified	3.23a	0.54	1.49b	0.13
Bacteroidetes	*Flavobacteria*	*Flavobacteriaceae*	*Flavobacterium*	1.00a	0.11	1.34a	0.26
Actinobacteria	*Actinobacteria*	*Nocardiaceae*	*Nocardioides*	0.98a	0.06	1.35a	0.09
Acidobacteria	*Acidobacteria*	*Acidobacteria Gp16* incertae sedis	*Acidobacteria Gp16* incertae sedis	1.69a	0.33	1.33a	0.15
Proteobacteria	*Gammaproteobacteria*	*Xanthomonadaceae*	*Lysobacter*	0.87a	0.10	1.23a	0.20
Proteobacteria	*Alphaproteobacteria*	*Bradyrhizobiaceae*	*Bradyrhizobium*	2.65a	0.02	1.22b	0.07
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*_unclassified	*Burkholderiales*_unclassified	1.54a	0.10	1.06b	0.07
Proteobacteria	*Alphaproteobacteria*	*Phyllobacteriaceae*	*Phyllobacteriaceae*_unclassified	1.53a	0.12	1.05b	0.08

Each measurement represents the mean of three replicate samples. Genera are presented in order of abundance in phenanthrene-amended soils. Means not sharing the same letter within a row are statistically different ($p < 0.05$).

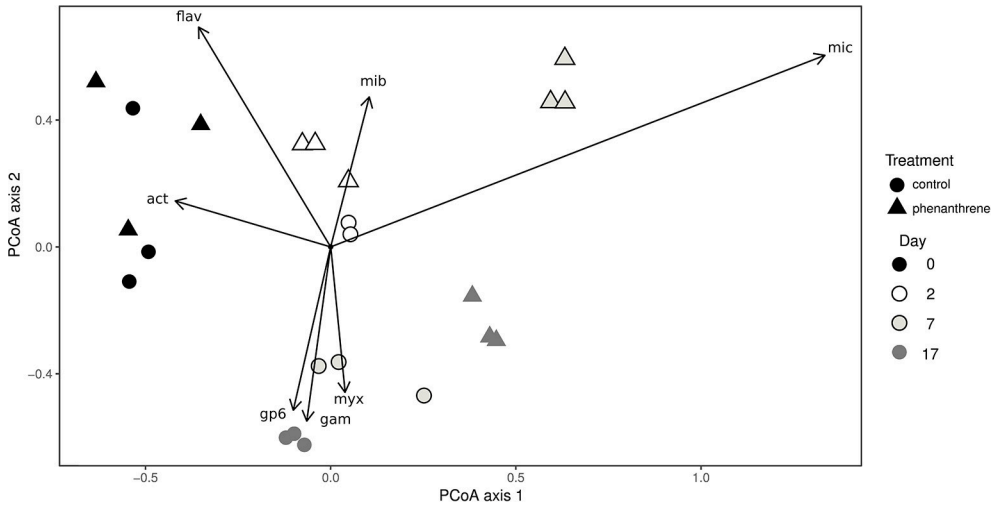


FIGURE 2 | Principal coordinates analysis (PCoA) of bacterial diversity at genus level in soil amended (triangles) or unamended (circles) with phenanthrene on days 0, 2, 7, and 17. mic, *Micrococcaceae*_unclassified; mib; *Microbacteriaceae*_unclassified; act, *Actinobacteria*_unclassified; Gp6, *Acidobacteria* Gp6 incertae sedis; fla, *Flavobacterium*; gam, *Gammaproteobacteria*_unclassified; myx, *Myxococcales*_unclassified. Arrows represent correlations > 0.4 between genera and PCoA axes.

When analyzing the relative abundance of those OTUs that were found to be significantly more abundant in the amended microcosms at any time point (phenanthrene-responsive OTUs), one single OTU classified to the Micrococcaceae family (OTU 1) increased in relative abundance from approximately 2.5% (day 0) to 32% (day 7) in the phenanthrene-amended microcosms, while in the unamended microcosms OTU 1 abundance also increased on between days 0 and 7 but to a lesser extent (10% of total sequences) (**Figure 3**). The second most relatively abundant phenanthrene-responsive OTU (OTU 2, *Sphingomonas* sp.) increased from ∼2–5% of the total sequences in the amended microcosms toward the end of the experiment, while in the unamended microcosms OTU 2 increased in abundance earlier in the experiment at day 2 to approximately 5% followed by a decline to 2.5 % of total sequences (**Figure 3**). It is noteworthy that both dominant phenanthrene-responsive OTUs increased in relative abundance in the unamended controls from day 0 and 2. The relative abundance of two other phenanthrene-responsive actinobacterial OTUs (OTU 40, unclassified Actinomycetales, and OTU 82, *Mycobacterium* sp.) increased in relative abundance from days 7 to 17 in the amended microcosms however in the controls their abundance decreased in this period (**Figure 4**).

The Shannon index of the bacterial community did not change over the course of the experiment in the unamended control microcosms. However, it was significantly lower in the phenanthrene-amended microcosms on day 7 compared to days 0 (adjusted p-value < 0.001), 2 (adjusted p-value < 0.001), and 17 (adjusted p-value < 0.05; **Figure 5**).

Abundance of RHDα-GP and RHDα-GN

The abundance of a gene (PAH-RHDα) encoding the alpha subunit of a ring hydroxylating dioxygenase involved in the first step of PAH degradation by either Gram-positive (RHDα-GP) or Gram-negative (RHDα-GN) bacteria in both amended and unamended soils is shown in **Figure 6**, respectively. The abundance of this gene associated with Gram-positive bacteria (RHDα-GP) was significantly higher (between 10^3 and 10^4 times higher) in both unamended and phenanthrene-amended soils than the gene associated with Gram-negative bacteria (RHDα-GN) at all-time points. The relative abundance of RHDα-GP changed over time and in response to the presence of phenanthrene ($P < 0.001$). Abundance increased significantly between days 2 and 7, and again between days 7 and 17 in phenanthrene-amended samples. Abundance of RHDα-GP did not change significantly in unamended soil remaining at about 100,000 copies ng^{-1} soil throughout. RHDα-GN abundance did not change significantly either over time, or in response to phenanthrene amendment.

Analysis of the Pearson's correlation between RHDα-GP and the abundance of OTUs 40, 82, 1, and 2 showed that OTUs 40 and 82 were more strongly correlated with this gene (Pearson correlation coefficients of 0.99, $p = 1.57^{-12}$ and 0.94, $p = 2.94^{-06}$ for OTUs 40 and 82, respectively), compared to OTUs 1 and 2 (Pearson correlation coefficients of 0.38, $p = 0.22$ and 0.80, p value $= 0.001$ for OTUs 1 and 2, respectively).

DISCUSSION

PAH degradation has been shown to depend on the number of aromatic rings present, with the lower molecular weight PAHs (four or less aromatic rings) more susceptible to degradation than high molecular weight PAHs (Cerniglia, 1992). The rapid rate of degradation of the 3-ring PAH, phenanthrene, observed in this study was therefore not surprising, with similar results reported elsewhere. Sawulski et al. (2014) reported that 95% of phenanthrene (200 mg kg^{-1}) was removed from a soil microcosm within two days. On day 2 of the experiment, the absence of a more pronounced phenanthrene toxic effect on

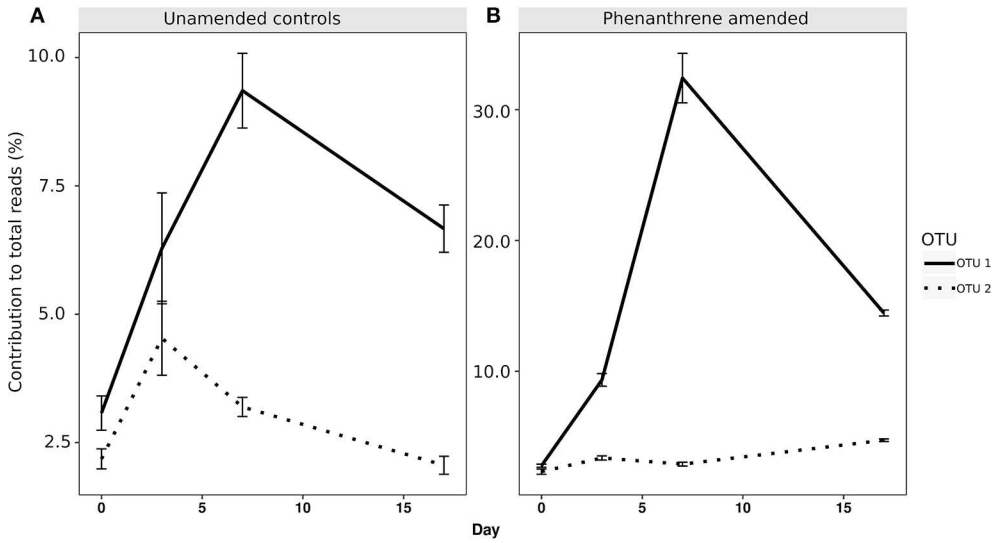


FIGURE 3 | Percentage relative abundance of the two most abundant PAH-responsive OTUs in **(A)** unamended and **(B)** phenanthrene-amended microcosm soils on days 0, 2, 7, and 17. OTU 1, unclassified *Micrococcaceae*; OTU 2, *Sphingomonas* sp.

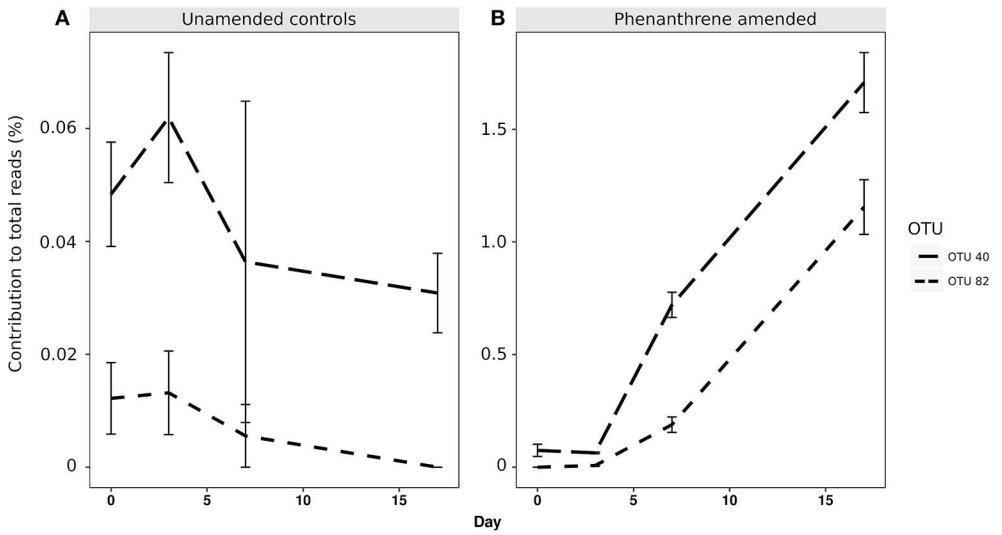


FIGURE 4 | Percentage relative abundance of PAH-responsive OTUs classified to Actinomycetales (OTU 40) and *Mycobacterium* sp. (OTU 82) in **(A)** unamended and **(B)** phenanthrene-amended microcosm soils on days 0, 2, 7, and 17.

the bacterial community, as indicated by the relatively small change in the relative abundances of the main bacterial phyla present, may be a result of previous exposure of these soils to creosote at the former timber treatment facility. Prior exposure to recalcitrant pollutants such as PAHs has been shown to impact soil microbial community structures (Johnsen and Karlson, 2005; Bargiela et al., 2015). Similarly, the fast rates of phenanthrene degradation observed on day 7 of the experiment are most likely a result of the prior exposure of this soil to PAHs. The constant leakage of PAHs into the soil from timber treatment may have stimulated any PAH degrading bacteria present or left a "seed" population of PAH-degrading bacteria capable of fast response to new pulses of PAH contamination. The greater phenanthrene degradation rate on day 7 coincided with a clear shift in bacterial community composition, indicating that phenanthrene degraders may have adapted and metabolized this compound. Actinobacteria and Proteobacteria are frequently observed as dominant phyla in soil (Janssen, 2006; Delgado-Baquerizo et al., 2016; Fierer, 2017) and representatives of both phyla have previously been associated with PAH degradation in soil (Mukherjee et al., 2014). In this study, both the bacterial diversity data and the abundance of the Gram-positive

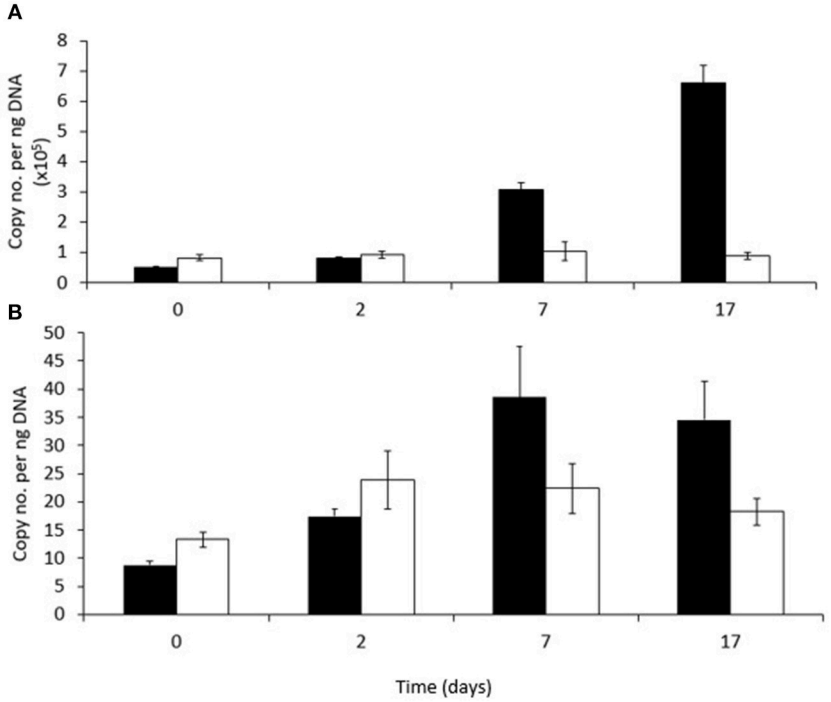


FIGURE 5 | Abundance of **(A)** RHDα-GP and **(B)** RHDα-GN genes in phenanthrene-amended (black bars) or unamended (white bars) in microcosm soils on days 0, 2, 7, and 17. Each bar represents the mean of three replicates.

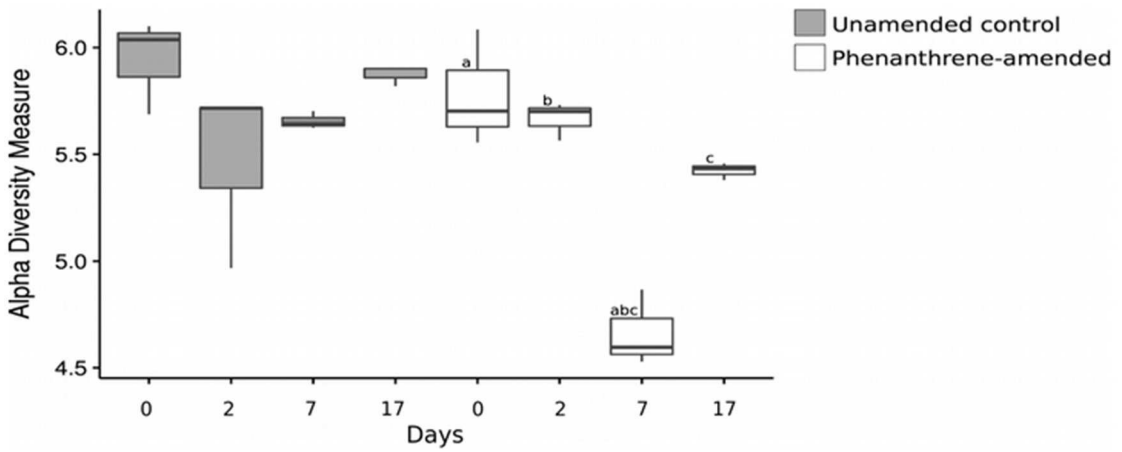


FIGURE 6 | Shannon index of the bacterial community at all-time points. Shared letters represent significant adjusted *p*-values determined by pairwise *t*-tests. The lines extend to the largest value observed.

PAH-RHDα genes suggest that the Actinobacteria were the main contributors to phenanthrene degradation. This is supported by the previous study by de Menezes et al. (2012), which reported a significant increase in transcripts associated with dioxygenases from Actinobacteria but few from the Proteobacteria in the same soil on day 7.

The higher relative abundance of Actinobacteria in PAH amended soil appears to have been driven largely by an increase in the abundance of a single unclassified OTU from the family *Micrococcaceae* (OTU 1). Members of the *Micrococcaceae* such as *Arthrobacter phenanthrneivorans* (Kallimanis et al., 2009) and *Arthrobacter oxydans* (Thion et al., 2013) have been associated with PAH degradation, both in pure culture and in the environment (Aryal and Kyriakides-Liakopoulou, 2013). Although considerably less relatively abundant than the *Micrococcaceae,* unclassified *Microbacteriaceae* were also

significantly more abundant in phenanthrene-amended soil, and this group has also been associated with PAH degradation in soil (Jacques et al., 2008; Jie et al., 2011). Although, based on relative abundance, the Proteobacteria appeared to play a smaller role in PAH-degradation in this study compared to the Actinobacteria, six OTUs classified to this phylum were significantly more abundant on days 7 and 17 in the contaminated soils compared to the controls, and this included the second most abundant phenanthrene-responsive OTU, which was classified to the *Sphingomonas* genus, often associated to PAH contamination (Leys et al., 2004). Both OTU 1 and OTU 2 were already amongst the most abundant in the test soils at the start of the experiment, which might have given them competitive advantage against other potential PAH-degrading taxa (O'Malley et al., 2010). Their original dominance in these soils may be connected to the previous history of PAH contamination of the source soil due to timber treatment, which is supported by the presence of PAH-RHDα genes in the unamended soil (Cerniglia, 1992; Johnsen and Karlson, 2005).

The involvement of OTU1 in phenanthrene degradation was suggested by the results of DeSeq2 analysis which demonstrated that the relative abundance of this OTU increased substantially at the same time as the PAH was being removed from soil. However, the relative abundance of OTU1 does not directly mirror the total abundance of the Gram-positive PAH-RHDα gene (**Figures 3**, **5**). This discrepancy may be due to changes in the total abundance of the remainder of the bacterial community in these soils, or due to this OTU expressing a PAH-RHDα gene which is not amplified by the qPCR primers used. A further possibility is that this OTU was not directly involved in the phenanthrene ring hydrolysis, but that it utilized its degradation products generated by other members of the microbial community (Festa et al., 2013). The relative abundance of two other actinobacterial OTUs (OTU 40, unclassified Actinomycetales and OTU 82, *Mycobacterium* sp.) showed stronger correlations with the total abundance of the PAH-RHDα gene than OTUs 1 and 2, indicating that OTUs 40 and 82 may therefore have been the primary taxa involved in the initial stages of phenanthrene degradation. Actinobacteria have been reported to be the primary taxa involved in PAH degradation (Uyttebroek et al., 2006; Cébron et al., 2008; Marcos et al., 2009) and a metatranscriptomic analysis of one time point during PAH degradation in soil reported that transcripts of PAH hydroxylase genes present were primarily from the Actinobacteria (de Menezes et al., 2012). The increase in the abundance of some Gram-negative OTUs of taxa associated with PAH degradation (e.g., *Sphingomonas*) without a concurrent increase in the abundance of the Gram-negative PAH-RHDα gene could also be explained by cross feeding between ring-hydroxylating bacteria and other bacteria capable of utilizing phenanthrene breakdown products (McDonald et al., 2005).

The considerable decrease in the relative abundance of OTU 1 and the concurrent rise in the relative abundance of the

TABLE 3 | Results from differential abundance analysis using DESeq2 (alpha = 0.05) showing the number of OTUs that are significant more abundant in either phenanthrene-amended or unamended samples at days 7 and 17 of the experiment.

Day	OTU	Class	Genus	Average abundance of differentially abundant OTUs	
				Control	Phenanthrene
7	OTU 1	Actinobacteria	Unclassified Micrococcaceae	9.35	32.42
	OTU 7	Actinobacteria	Unclassified Microbacteriaceae	0.85	1.83
	OTU 25	Actinobacteria	Unclassified Microbacteriaceae	0.47	0.90
	OTU 40	Actinobacteria	Unclassified Actinomycetales	0.04	0.72
	OTU 82	Actinobacteria	*Mycobacterium*	0.01	0.19
	OTU 172	Actinobacteria	Unclassified Microbacteriaceae	0.06	0.07
	OTU 187	Gammaproteobacteria	*Cellvibrio*	0.01	0.19
	OTU 320	Betaproteobacteria	Hydrogenophaga	0.00	0.14
17	OTU 1	Actinobacteria	Unclassified Micrococcaceae	6.67	14.45
	OTU 2	Alphaproteobacteria	*Sphingomonas*	2.06	4.72
	OTU 7	Actinobacteria	Unclassified Microbacteriaceae	0.68	1.31
	OTU 24	Betaproteobacteria	Unclassified Comamonadaceae	0.65	1.43
	OTU 25	Actinobacteria	Unclassified Microbacteriaceae	0.36	0.88
	OTU 32	Betaproteobacteria	Unclassified Burkholderiales	0.71	0.26
	OTU 40	Actinobacteria	Unclassified Actinomycetales	0.03	1.71
	OTU 68	Gammaproteobacteria	*Pseudoxanthomonas*	0.08	0.60
	OTU 82	Actinobacteria	*Mycobacterium*	0.00	1.15
	OTU 102	Betaproteobacteria	Unclassified Comamonadaceae	0.03	0.30
	OTU 273	Gammaproteobacteria	Unclassified Gammaproteobacteria	0.04	0.06
	OTU 244	Actinobacteria	Unclassified Actinomycetales	0.15	0.00

The numbers show average abundance for each OTU in each treatment.

Proteobacteria between days 7 and 17 in the phenanthrene-amended microcosms indicate that the community was returning to its original composition. This conclusion is further supported by the closer proximity of phenanthrene-amended day 17 samples to day 0 and 2 sample data points from phenanthrene-amended and control samples in the PCoA plot. The return of the bacterial community to its original composition demonstrates the resilience of the bacteria to the impact of phenanthrene, which is likely due to their previous exposure to PAHs in the original site from where the soil was sourced, as discussed above. Bacterial communities have been reported to recover from exposure to pollutants such as PAHs, but the length of time this recovery may take will depend on a range of factors including concentration and type of pollutant (Bordenave et al., 2007; Rodriguez et al., 2015). It may be that the growth of OTUs 1, 40, and 82 was key to the resilience of the overall bacterial community to phenanthrene on account of their possible role in removing this pollutant from the soil. With most of phenanthrene exhausted from the soil, OTU 1 may have lost its competitive advantage, and other microorganisms were then able to grow and return to their original relative abundances.

Interestingly, the same unclassified *Micrococcaceae* and *Sphingomonas* OTUs that responded positively to phenanthrene amendment also increased in relative abundance to a lesser extent in the control microcosms, following similar abundance temporal patterns as in the amended microcosms. The increase in abundance of the two dominant phenanthrene-responsive OTUs in both the PAH amended and control microcosms suggests that the ability to withstand PAH toxicity and use these compounds as growth substrates may be related to their general ecological opportunism. Although widely used to generate useful information, microcosm-based studies have several limitations (Carpenter, 1996). Harvesting of soil and set-up of microcosms can introduce oxygen and affect soil structure, which are known to affect microbial communities (Fierer et al., 2003). Opportunistic bacteria are often the first to colonize a habitat following a disturbance and show adaptations such as fast growth on limiting substrates and high tolerance to environmental stress (Sigler et al., 2002; Sigler and Zeyer, 2004; Fierer et al., 2010; Lozupone et al., 2012). It would appear that disturbance-driven ecological succession may have occurred in the phenanthrene-amended and to a lesser extent in the control soil microbiome in this experiment (Sigler and Zeyer, 2004), however the presence of phenanthrene enhanced the dominance of these opportunistic strains. The presence of an opportunistic phenotype is supported by results from the metatranscriptomic study undertaken on the same soils. Transcripts classified as heat shock proteins from Actinobacteria and Alphaproteobacteria were observed to increase when phenanthrene was added to the soil (de Menezes et al., 2012) and similar genes were identified as a genetic signature of gut early-colonizing opportunistic bacteria by Lozupone et al. (2012).

The growth and decline of the dominant PAH-responsive bacteria in the current study decreased community evenness, in agreement with the study of Thomas and Cébron (2016) who observed a decrease in bacterial community evenness two days following phenanthrene soil amendment. However, relationships between microbial evenness (Shannon diversity) and disturbance vary depending on the frequency, intensity and type of disturbance (Gibbons et al., 2016). These relationships were not tested specifically in the current experiment and warrant further investigation.

CONCLUSION

The presence of phenanthrene led to significant shifts in the structure of soil bacterial communities, and this change was dominated by an increase in the relative abundance of a small number of opportunistic taxa. The prior exposure to pollution of the soil used in this experiment may have played a role in conditioning the soil community for fast responses to pulses of PAH contamination through the growth of opportunistic, fast-growing PAH-degrading bacteria mainly from the Actinobacteria phylum. The fact that 8 out of the13 PAH-responsive bacterial OTUs in this experiment were not classified to genus level (**Table 3**) highlights the need to better characterize uncultured soil microorganisms which have beneficial roles in bioremediation. Finally, the recognition that the soil bacterial community's response to phenanthrene contamination is at least in some cases restricted to a small number of bacterial taxa may simplify the ecological modeling and design of PAH-remediation strategies.

AUTHOR CONTRIBUTIONS

SS is a postdoctoral fellow and contributed to qPCR analysis, amplicon data analysis and preparation of the manuscript. MA contributed to the qPCR analysis. ED and NC are the PIs and were responsible for experimental design. ED also contributed to the writing of the paper. AdM contributed to experimental design, set up the microcosm experiment, carried out the amplicon sequencing and the majority of the data analysis and contributed to preparation of the manuscript.

FUNDING

This work was funded under the Irish Environmental Protection Agency STRIVE programme 2007–2013 (grant no. 2008-PhD-WRM-1), the Ireland-Wales Programme 2007-2013 (GIFT project), and the SLAB scholarship programme of the Malaysian Ministry of Higher Education and the International Islamic University, Malaysia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02815/full#supplementary-material

REFERENCES

Anon (2012). *Soil Contamination: A Severe Risk for the Environment and Human Health.* Available online at: http://globalsoilweek.org/wpcontent/uploads/2014/11/GSW_factsheet_Contamination_en.pdf

Aryal, M., and Kyriakides-Liakopoulou, M. (2013). Biodegradation and kinetics of phenanthrene and pyrene in the presence of non-ionic surfactants by *Arthrobacter* strain Sphe3. *Water Air Soil Pollut.* 224, 1426–1436. doi: 10.1007/s11270-012-1426-8

Bargiela, R., Mapelli, F., Rojo, D., Chouaia, B., Tornés, J., Borin, S., et al. (2015). Bacterial population and biodegradation potential in chronically crude oilcontaminated marine sediments are strongly linked to temperature. *Sci. Rep.* 5:11651. doi: 10.1038/srep11651

Bastiaens, L., Springael, D., Wattiau, P., Harms, H., deWachter, R., Verachtert, H., et al. (2000). Isolation of adherent polycyclic aromatic hydrocarbon (PAH)-degrading bacteria using PAH-sorbing carriers. *Appl. Environ. Microbiol.* 66, 1834–1843. doi: 10.1128/AEM.66.5.1834-1843.2000

Bordenave, S., Goñi-Urriza, M. S., Caumette, P., and Duran, R. (2007). Effects of heavy fuel oil on the bacterial community structure of a pristine microbial mat. *Appl. Environ. Microbiol.* 73, 6089–6097. doi: 10.1128/AEM.01352-07

Brussaard, L. (2012). "Ecosystem services provided by the soil biota," in *Soil Ecology and Ecosystem Services*, eds R. D. Bardgett, J. E. Herrick, J. Six, T. Heffin Jones, D.R. Strong, and W. H. van der Putten (Oxford, UK: Oxford University Press), 45–58.

Carpenter, S. R. (1996). Microcosm experiments have limited relevance for community and ecosystem ecology. *Ecology* 77, 677–680. doi: 10.2307/2265490

Cébron, A., Norini, M. P., Beguiristain, T., and Leyval, C. (2008). Real-Time PCR quantification of PAH-ring hydroxylating dioxygenase genes from Gram-positive and Gram-negative bacteria in soil and sediment samples. *J. Microbiol. Methods* 73, 148–159. doi: 10.1016/j.mimet.2008.01.009

Cerniglia, C. E. (1992). Biodegradation of polycyclic aromatic hydrocarbons. *Biodegradation* 3, 351–368. doi: 10.1007/BF00129093

Cerniglia, C. E. (1997). Fungal metabolism of polycyclic aromatic hydrocarbons: past, present and future applications in bioremediation. *J. Ind. Microbiol. Biotechnol.* 19, 324–333. doi: 10.1038/sj.jim.2900459

Collins, C. D., Mosquera-Vazquez, M., Gomez-Eyles, J., Mayer, P., Gouliarmou, V., and Blum, F. (2013). Is there sufficient 'sink' in current bioaccessibility determinations of organic pollutants in soils? *Environ. Pollut.* 181, 128–132. doi: 10.1016/j.envpol.2013.05.053

de Menezes, A., Clipson, N., and Doyle, E. M. (2012). Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. *Environ. Microbiol.* 14, 2577–2588. doi: 10.1111/j.1462-2920.2012.02781.x

Dean, J. R., and Xiong, G. H. (2000). Extraction of organic pollutants from environmental matrices: selection of extraction technique. *Trac. Trends Anal. Chem.* 19, 553–564. doi: 10.1016/S0165-9936(00)00038-8

Delgado-Balbuena, L., Bello-Lopez, J. M., Navarro-Noya, Y. E., Rodríguez-Valentín, A., Luna-Guido, M. L., and Dendooven, L. (2016). Changes in the bacterial community structure of remediated anthracene-contaminated soils. *PLoS ONE* 11:e0160991. doi: 10.1371/journal.pone.0160991

Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., et al. (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* 7:10541. doi: 10.1038/ncomms10541

Doyle, E., Muckian, L., Hickey, A. M., and Clipson, N. (2008). Microbial PAH degradation. *Adv. Appl. Microbiol.* 65, 27–66. doi: 10.1016/S0065-2164(08)00602-3

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Fernández-Luqueño, F., Valenzuela-Encinas, C., Marsch, R., Martínez-Suárez, C., Vázquez-Núñez, E., and Dendooven, L. (2011). Microbial communities to mitigate contamination of PAHs in soil—possibilities and challenges: a review. *Environ. Sci. Pollut. Res.* 18, 12–30. doi: 10.1007/s11356-010-0371-6

Festa, S., Coppotelli, B. M., and Morelli, I. S. (2013). Bacterial diversity and functional interactions between bacterial strains from a phenanthrene-degrading consortium obtained from a chronically contaminated-soil. *Int. Biodeterior. Biodegrad.* 85, 42–51. doi: 10.1016/j.ibiod.2013.06.006

Festa, S., Macchi, M., Cortes, F., Morelli, I. S., and Coppotelli, B. M. (2016). Monitoring the impact of bioaugmentation with a PAH-degrading strain on different soil microbiomes using pyrosequencing. *FEMS Microbiol. Lett.* 92:125. doi: 10.1093/femsec/fiw125

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Micro.* 15, 579–590. doi: 10.1038/nrmicro.2017.87

Fierer, N., Nemergut, D., Knight, R., and Craine, J. M. (2010). Changes through time: integrating microorganisms into the study of succession. *Res. Microbiol.* 161, 635–642. doi: 10.1016/j.resmic.2010.06.002

Fierer, N., Schimel, J. P., and Holden, P. A. (2003). Variations in microbial community composition through two soil depth profiles. *Soil Biol. Biochem.* 35, 167–176. doi: 10.1016/S0038-0717(02)00251-1

Fuentes, S., Mendez, V., Aguila, P., and Seeger, R. M. (2014). Bioremediation of petroleum hydrocarbons: catabolic genes, microbial communities, and applications. *Appl. Microbiol. Biotechnol.* 98, 4781–4794. doi: 10.1007/s00253-014-5684-9

Ghosal, D., Ghosh, S., Dutta, T. K., and Ahn, Y. (2016). Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): a review. *Front. Microbiol.* 7:1369. doi: 10.3389/fmicb.2016.01369

Gibbons, S. M., Scholz, M., Hutchison, A. L., Dinner, A. R., Gilbert, J. A., and Coleman, M. L. (2016). Disturbance regimes predictably alter diversity in an ecologically complex bacterial system. *Mbio* 7:e01372. doi: 10.1128/mBio.01372-16

Griffiths, B. S., and Philipot, L. (2013). Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiol. Rev.* 37, 112–129. doi: 10.1111/j.1574-6976.2012.00343.x

Griffiths, R. I., Whiteley, A. S., O'Donnell, G., and Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analsis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* 66, 5488–5491. doi: 10.1128/AEM.66.12.5488-5491.2000

Jacques, R. J. S., Okeke, B. C., Bento, F. M., Teixeira, A. S., Peralba, M. C. R., and Camargo, F. A. O. (2008). Microbial consortium bioaugmentation of a PAH-contaminated soil. *Biores. Technol.* 99, 2637–2643. doi: 10.1016/j.biortech.2007.04.047

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi: 10.1128/AEM.72.3.1719-1728.2006

Jiao, S., Zhang, Z., Yang, F., Lin, Y., Chen, W., and Wei, G. (2017). Temporal dynamics of microbial communities in microcosms in response to pollutants. *Mol. Ecol.* 26, 923–936. doi: 10.1111/mec.13978

Jie, H. L., Yun, Y. C., Yun, T., Hui, L. G., and Ling, Z. T. (2011). Using population dynamics analysis by DGGE to design the bacterial consortium isolated from mangrove sediments for biodegradation of PAHs. *Int. Biodeterior. Biodegradation.* 65, 269–275. doi: 10.1016/j.ibiod.2010.11.010

Johnsen, A. R., and Karlson, U. (2005). PAH degradation capacity of soil microbial communities – does it depend on PAH exposure? *Microb. Ecol.* 50, 488–495. doi: 10.1007/s00248-005-0022-5

Kallimanis, A., Kavakiotis, K., Perisynakis, A., Spröer, C., Pukall, R., Drainas, C., et al. (2009). *Arthrobacter phenanthrenivorans* sp. nov., to accomodate the phenanthere-degrading bacterium *Arthrobacter* sp. strain Sphe3. *Int. J. Syst. Evol. Microbiol.* 24, 569–581. doi: 10.1099/ijs.0.000984-0

Kettler, T., Doran, J., and Gilbert, T. (2001). Simplified method for soil particle-size determination to accompany soil-quality analyses. *Soil Sci. Soc. Am. J.* 65, 849–852. doi: 10.2136/sssaj2001.653849x

Kuppusamy, S., Thavamani, P., Megharaj, M., Venkateswarlu, K., Lee, Y. B., and Naidu, R. (2016). Pyrosequencing analysis of bacterial diversity in soils contaminated long-term with PAHs and heavy metals: implications to bioremediation. *J. Hazard. Mater.* 317, 169–179. doi: 10.1016/j.jhazmat.2016.05.066

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6955–6959. doi: 10.1073/pnas.82.20.6955

Leys, N., Ryngaert, A., Bastiaens, L., Verstraete, W., Top, E. M., and Springael, D. (2004). Occurrence and phylogenetic diversity of *Sphingomonas* strains in soils contaminated with polycyclic aromatic hydrocarbons. *Appl Environ. Microbiol.* 70, 1944–1955. doi: 10.1128/AEM.70.4.1944-1955.2004

Lors, C., Damidot, D., Ponge, J. F., and Périé, F. (2012). Comparison of a bioremediation process of PAHs in a PAH-contaminated soil at field and laboratory scales. *Environ. Pollut.* 165, 11–17. doi: 10.1016/j.envpol.2012.02.004

Lors, C., Ryngaert, A., Périé, F., Diels, L., and Damidot, D. (2010). Evolution of bacterial community during bioremediation of PAHs in a coal tar contaminated soil. *Chemosphere* 81, 1263–1271. doi: 10.1016/j.chemosphere.2010.09.021

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold-change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550–571. doi: 10.1186/s13059-014-0550-8

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansen, J. K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230. doi: 10.1038/nature11550

Marcos, M. S., Lozada, M., and Dionisi, H. M. (2009). Aromatic hydrocarbon degradation genes from chronically polluted subantarctic marine sediments. *Lett. Appl. Microbiol.* 49, 602–608. doi: 10.1111/j.1472-765X.2009.02711.x

Martínez, I., Wallace, G., Zhang, C. M., Legge, R., Benson, A. K., Carr, T. P., et al. (2009). Diet induced metabolic improvements in a hamster model of hypercholesterolemia are strongly linked to alterations of the gut microbiota. *Appl. Environ. Microbiol.* 75, 4175–4184. doi: 10.1128/AEM.00380-09

McDonald, I. R., Radajewski, S., and Murrell, J. C. (2005). Stable isotope probing of nucleic acids in methanotrophs and methylotrophs: a review. *Org. Geochem.* 36, 779–787. doi: 10.1016/j.orggeochem.2005.01.005

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R package for reroducible interactive ananlysis and graphics of microbiome census data. *PLoS ONE* 8:e61217. doi: 10.1371/journal.pone.0061217

McMurdie, P. J., and Holmes, S. (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* 10:e1003531. doi: 10.1371/journal.pcbi.1003531

Muckian, L. M., Grant, R. J., Clipson, N. J., and Doyle, E. M. (2009). Bacterial community dynamics during bioremediation of phenanthrene- and fluoranthene-amended soil. *Int. Biodeterior. Biodegr.* 63, 52–56. doi: 10.1016/j.ibiod.2008.04.005

Mukherjee, S., Juottenen, H., Siivonen, P., Quesada Lloret, C., Tuomi, P., Pulkkinen, P., et al. (2014). Spatial patterns of microbial diversity and activity in an aged creosote-contaminated site. *ISME J.* 8, 2131–2142. doi: 10.1038/ismej.2014.151

Muyzer, G., Dewaal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S ribosomal RNA. *Appl. Environ. Microbiol.* 59, 695–700.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, P. R., et al. (2018). *vegan: Community Ecology Package* version 2.4-2. Available online at: https://cran.r-project.org/web/packages/vegan/index.html.

O'Malley, L., Korniss, G., and Mungara, S. S. P. (2010). Spatial competition and the dynamics of rarity in a temporally varying environment. *Evol. Ecol. Res.* 12, 279–305. Available online at: http://www.evolutionary-ecology.com/issues/v12/n03/ccar2546.pdf

Posada-Baquero, R., and Ortega-Calvo, J. J. (2011). Recalcitrance of polycyclic aromatic hydrocarbons in soil contributes to background pollution. *Environ. Pollut.* 159, 3692–3699. doi: 10.1016/j.envpol.2011.07.012

Rodriguez, R. L., Overholt, W. A., Hagan, C., Huettel, M., Kostka, J. E., and Konstantinidis, K. T. (2015). Microbial community successional patterns in beach sands impacted by the Deepwater Horizon oil spill. *ISME J* 9, 1928–1940. doi: 10.1038/ismej.2015.5

Sawulski, P., Clipson, N., and Doyle, E. (2014). Effects of polycyclic aromatic hydrocarbons on microbial community structure and PAH ring hydroxylating dioxygenase gene abundance in soil. *Biodegradation* 25, 835–847. doi: 10.1007/s10532-014-9703-4

Sayara, T., Borràs, E., Caminal, G., Sarr,à, M., and Sánchez, A. (2011). Bioremediation of PAH-contaminated soil through composting: Influence of bioaugmentation and biostimulation on contaminant biodegradation. *Int. Biodeterior. Biodegr.* 65, 859–865. doi: 10.1016/j.ibiod.2011.05.006

Schloss, P. D. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE* 4:e8230. doi: 10.1371/journal.pone.0008230

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Shade, A. (2017). Diversity is the question, not the answer. *ISME J.* 11, 1–6. doi: 10.1038/ismej.2016.118

Sigler, W. V., Crivii, S., and Zeyer, J. (2002). Bacterial succession in glacial forefield soils characterized by community structure, activity and opportunistic growth dynamics. *Microb. Ecol.* 44, 306–316. doi: 10.1007/s00248-002-2025-9

Sigler, W. V., and Zeyer, J. (2004). Colony-forming analysis of bacterial community succession in deglaciated soils indicates pioneer stress-tolerant opportunists. *Microb. Ecol.* 48, 316–323. doi: 10.1007/s00248-003-0189-6

Storey, S., Ashaari, M. M., McCabe, G., Harty, M., Dempsey, R., Doyle, O., et al. (2014). Microbial community structure during fluoranthene degradation in the presence of plants. *J. Appl. Microbiol.* 117, 74 – 84. doi: 10.1111/jam.12518

Sun, F. L., Wang, Y. S., Sun, C. C., Peng, Y. L., and Deng, C. (2012). Effects of three different PAHs on nitrogen-fixing bacterial diversity in mangrove sediment. *Ecotoxicology* 21, 1651–1660. doi: 10.1007/s10646-012-0946-8

Thion, C., Cébron, A., Beguiristain, T., and Leyval, C. (2013). Inoculation of PAH-degrading microbial interactions and PAH dissipation. *Biodegradation* 24, 569–581. doi: 10.1007/s10532-013-9628-3

Thomas, F., and Cébron, A. (2016). Short-term rhizosphere effect on available carbon sources, phenanthrene degradation, and active microbiome in an aged-contaminated industrial soil. *Front. Microbiol.* 7:92. doi: 10.3389/fmicb.2016.00092

Uyttebroek, M., Breugelmans, P., Janssen, M., Wattiau, P., Joffe, B., Karlson, U., et al. (2006). Distribution of the *Mycobacterium* community and polycyclic aromatic hydrocarbons (PAHs) among different size. *Environ. Microbiol.* 8, 836–847. doi: 10.1111/j.1462-2920.2005.00970.x

Wang, L. W., Li, F., Zhan, Y., and Zhu, L. (2016). Shifts in microbial community structure during in situ surfactant-enhanced bioremediation of polycyclic aromatic hydrocarbon-contaminated soil. *Environ. Sci. Pollut. Res.* 23, 14451–14461. doi: 10.1007/s11356-016-6630-4

Wang, W., Massey Simonich, S., Xue, M., Zhao, J., Zhang, N., Wang, R., et al. (2010). Concentrations, sources and spatial distribution of polycyclic aromatic hydrocarbons in soils from Beijing, Tianjin and surrounding areas, North China. *Environ. Pollut.*158, 1245–1251. doi: 10.1016/j.envpol.2010.01.021

Wickliffe, J., Overton, E., Frickel, S., Howard, J., Wilson, M., Simon, B., et al. (2014). Evaluation of polycyclic aromatic hydrocarbons using analytical methods, toxicology, and risk assessment research: seafood safety after a petroleum spill as an example. *Environ. Health Perspect.* 122, 6–9. doi: 10.1289/ehp.1306724

Yang, Y., Wang, J., Liao, J., Xie, S., and Huang, Y. (2015). Abundance and diversity of soil petroleum hydrocarbon-degrading microbial communities in oil exploring areas. *Appl. Microbiol. Biotechnol.* 99, 1935–1946. doi: 10.1007/s00253-014-6074-z

Zafra, G., Taylor, T. D., Absalón, A. E., and Cortés-Espinosa, D. V. (2016). Comparative metagenomic analysis of PAH degradation in soil by a mixed microbial consortium. *J. Hazard. Mater.* 318, 702–710. doi: 10.1016/j.jhazmat.2016.07.060

Zhao, J. K., Li, X. M., Ai, G. M., Deng, Y., Liu, S. J., and Jiang, C. Y. (2016). Reconstruction of metabolic networks in a fluoranthene-degrading enrichments from polycyclic aromatic hydrocarbon polluted soil. *J. Hazard. Mater.* 318, 90–98. doi: 10.1016/j.jhazmat.2016.06.055

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Storey, Ashaari, Clipson, Doyle and de Menezes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 May 2018
doi: 10.3389/fmicb.2018.00869

Taxon-Function Decoupling as an Adaptive Signature of Lake Microbial Metacommunities Under a Chronic Polymetallic Pollution Gradient

Bachar Cheaib[1]*, *Malo Le Boulch*[1,2], *Pierre-Luc Mercier*[1] and *Nicolas Derome*[1]

[1] Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec, QC, Canada, [2] GenPhySE, Institut National de la Recherche Agronomique, Université de Toulouse, INPT, ENVT, Castanet-Tolosan, France

Adaptation of microbial communities to anthropogenic stressors can lead to reductions in microbial diversity and disequilibrium of ecosystem services. Such adaptation can change the molecular signatures of communities with differences in taxonomic and functional composition. Understanding the relationship between taxonomic and functional variation remains a critical issue in microbial ecology. Here, we assessed the taxonomic and functional diversity of a lake metacommunity system along a polymetallic pollution gradient caused by 60 years of chronic exposure to acid mine drainage (AMD). Our results highlight three adaptive signatures. First, a signature of taxon—function decoupling was detected in the microbial communities of moderately and highly polluted lakes. Second, parallel shifts in taxonomic composition occurred between polluted and unpolluted lakes. Third, variation in the abundance of functional modules suggested a gradual deterioration of ecosystem services (i.e., photosynthesis) and secondary metabolism in highly polluted lakes. Overall, changes in the abundance of taxa, function, and more importantly the polymetallic resistance genes such as *copA, copB, czcA, cadR, cCusA*, were correlated with trace metal content (mainly Cadmium) and acidity. Our findings highlight the impact of polymetallic pollution gradient at the lowest trophic levels.

Keywords: function, taxon, decoupling, polymetallic gradient, Cadmium, evolution, adaptation, resistance

OPEN ACCESS

Edited by:
Florence Abram,
National University of Ireland Galway, Ireland

Reviewed by:
Zhili He,
University of Oklahoma, United States
Shishir K. Gupta,
University of Würzburg, Germany
Jonathan Elias Maldonado,
Universidad de Chile, Chile
David Georges Biron,
Centre National de la Recherche Scientifique (CNRS), France

***Correspondence:**
Bachar Cheaib
bachar.cheaib.1@ulaval.ca

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *02 September 2017*
Accepted: *16 April 2018*
Published: *03 May 2018*

Citation:
Cheaib B, Le Boulch M, Mercier P-L and Derome N (2018) Taxon-Function Decoupling as an Adaptive Signature of Lake Microbial Metacommunities Under a Chronic Polymetallic Pollution Gradient. Front. Microbiol. 9:869. doi: 10.3389/fmicb.2018.00869

INTRODUCTION

Micro-organisms represent a significant portion of global biodiversity and are the engine driving Earth's biogeochemical cycles and primary production (Falkowski et al., 2008; Green et al., 2008). Ecosystem services provided by microbes ensure optimal environmental conditions for all multicellular life forms (Robinson et al., 2010). For decades, the implications of taxon-function relationships in microbial communities have been debated by researchers (Doolittle and Zhaxybayeva, 2009; Bissett et al., 2013; Martiny et al., 2013; Louca et al., 2016b; Morrissey et al., 2016). On one hand, researchers showed that even very closely related taxa exhibited contrasting metabolic and ecological functions (e.g., distinct growth rates and metabolic substrate utilization profiles), indicating a gap between taxon phylogeny and the functional repertoires of some bacterial genera (Jaspers and Overmann, 2004; Maharjan et al., 2006; Doolittle and Zhaxybayeva, 2009). These studies employed molecular taxonomic profiling, either by sequencing SSU (small subunit

ribosomal ribonucleic acid) 16S rRNA (Jaspers and Overmann, 2004; Doolittle and Zhaxybayeva, 2009) or specific housekeeping genes (Maharjan et al., 2006). On the other hand, studies focused on microbial molecular evolution and ecology reported a significant relationship between phylogenetic groups or taxonomic composition at different hierarchical levels (i.e., Phylum and Class) with ecological and functional traits (Webb et al., 2002; Martiny et al., 2006, 2013; Ward et al., 2006; Gupta and Lorenzini, 2007; Allison and Martiny, 2008; Philippot et al., 2010; Gravel et al., 2011). The majority of these genomic studies have been limited to correlating traits with taxa abundance variation. Additional evidence at the community level is needed to predict the interplay of evolutionary processes [horizontal gene transfer (HGT), gene loss, selective pressure] and ecological processes (spatial dispersal limits, biotic interactions, neutral biogeography) drive metacommunity composition and functional repertoires in complex ecological contexts.

With advances in sequencing technologies, metagenomic approaches have the potential to advance our understanding of both the taxonomic and functional composition of complex microbial communities. In this respect, metagenomic studies have revealed significant coupling between taxonomic composition or phylogenetic lineages and ecological traits (Bouvier and del Giorgio, 2002; Philippot et al., 2010) or functional gene repertoires (Debroas et al., 2009; Goldfarb et al., 2011; Muegge et al., 2011; Bryant et al., 2012; Fierer et al., 2012b; Langille et al., 2013; Martiny et al., 2013; Forsberg et al., 2014; Mayali et al., 2014; Vanwonterghem et al., 2014; Morrissey et al., 2016; Larkin and Martiny, 2017). For example, in natural lake communities, associations are reported between taxon abundance and function (Debroas et al., 2009), and in soil communities from multiple environments, with chemical substrate variation (Goldfarb et al., 2011) and functional attributes (Fierer et al., 2012b). Most of these studies have been conducted in relatively unperturbed environments, and on microbial communities facing moderate to low selective pressure.

Other microbial community studies, mostly based on 16S rRNA gene analysis, and rarely complemented by whole metagenome shotgun sequencing, revealed either partial or marked decoupling between taxonomic composition and ecological traits (Lima-Mendez et al., 2015) or functional gene repertoires. Patterns of complete to partial decoupling are often found in natural environmental conditions (Hooper et al., 2008, 2009; Burke et al., 2011; Raes et al., 2011; Smillie et al., 2011; Barberán et al., 2012; Louca et al., 2016b). This taxon-function decoupling has rarely been discussed in extreme environments such as acid mine drainages (AMD) (Kuang et al., 2016). These findings highlight the need to further investigate environments where initial conditions have been perturbed by xenobiotic factors (Bowen et al., 2011).

The occurrence of taxon-function decoupling has been reported in other metagenomic studies as functional redundancy between phylogenetically distant taxa (Green et al., 2008; Burke et al., 2011; Stokes and Gillings, 2011) and divergent microenvironments (Hooper et al., 2008, 2009). To summarize, taxonomic and functional features could be useful in assessing adaptive response of microbial metacommunities in disturbed ecosystems. One study, to our knowledge, has focused on the outcome of microbial taxon-function relationships under selection gradients, indicating possible linkages between the structure and functioning of soil microbial communities (Fierer et al., 2012a). Thus, it remains uncertain whether taxon-function decoupling is an adaptive response to a gradual selective pressure. Xenobiotic stressors like antibiotics, chemical and metallic pollutants erode microbial biodiversity (Parnell et al., 2009), which is predicted to impair or erode ecosystem services (Sandifer and Sutton-Grier, 2014). Therefore, the characterization of taxon-function decoupling patterns will enhance our understanding of the robustness of microbial functional networks that ensure key ecosystem services. Here, the complex connections of microbial biodiversity and ecosystem services (Miki et al., 2014) were addressed at the molecular level by comparing variation in the taxonomic composition and molecular functions of microbial communities.

We hypothesized that a stress gradient, specifically a polymetallic pollution gradient over a relatively long evolutionary time scale in terms of bacterial generation time, would result in adaptive signatures in taxonomic composition and functional repertoires. Specifically, we predicted that stress gradients would gradually induce selection for microbial metacommunities with functional repertoires and a taxonomic composition capable of thriving in this harsh environment. To test our hypothesis, we targeted lakes polluted by a polymetallic gradient of acid mine waters. Heavy metals can originate either from natural sources such as volcanic activity or anthropogenically by mines tailings, an important source of AMD. Acidity gradients recorded in lake waters surrounded by natural volcanic activity (e.g., Indonesian crater lake Kawah Ijen, Argentinian volcanic lake in Patagonia), have significant effects on the microbial community composition and biodiversity (Wendt-Potthoff and Koschorreck, 2002; Löhr et al., 2006). AMD is created by the exposure of sulphidic minerals to air and water forming soluble sulfates (Almeida et al., 2008). Ferrous minerals become oxidized in contact with water producing ferric ions and H_2 (Johnson and Hallberg, 2003; Edwards and Bazylinski, 2008). Leached ions into streams generate acidic water by lowering the pH (<3). Consequently, other metal ions such as Zn, Hg, Ni, Cr, Cd, Cu, Mn, Al, As, and Pb appear in AMD waters at high concentrations. There are limited descriptions of microbial diversity in AMD in the literature, especially in impacted environments with high zinc and cadmium concentrations (Almeida et al., 2008). In AMD polluted surface water, Almeida et al. (2008) showed that bacterial diversity in Sepetiba Bay, Brazil, which is much higher than archaeal diversity, was dominated by *Proteobacteria*, *Actinobacteria*, *Cyanobacteria* and had a high abundance of unclassified bacteria (unknown strains). Similar composition (dominance of Proteobacteria) was observed over 59 microbial communities from physically and geochemically diverse AMD sites across Southeast China (Kuang et al., 2013). Kuang et al. (2013) revealed that acidity gradient is a major factor explaining community differences between AMD communities regardless of the long-distance isolation and the distinct substrate types. Likewise, the investigation of the

microbial diversity of an extremely acidic, metal-rich water lake (Lake Robule, Bor, Serbia) revealed low diversity dominated by *Proteobacteria* strains (Stankovic et al., 2014). Similar community composition was observed in bacterioplankton communities exposed to cadmium in coastal water microcosms (Wang et al., 2015). Similar to surface waters, Hemme et al. (2010, 2016) highlighted that chronic exposure to high concentrations of heavy metals (∼50 years) in groundwater caused a massive decrease in biodiversity, characterized by a high abundance of *Proteobacteria*, as well as a significant loss in allelic and metabolic diversity. More importantly, Hemme et al. (2016) pointed to the importance of HGT during the evolution of groundwater microbial communities in response to heavy metal exposure. However, very few studies were carried out on water polluted across a polymetallic gradient (Kuang et al., 2013; Desoeuvre et al., 2016). One of those studies reported the impact of an extreme poly-metallic gradient (including arsenic) on the diversity and distribution of arsenic-related genes in river waters (Desoeuvre et al., 2016). Other studies on AMD polluted freshwater sediments (Sánchez-Andrea et al., 2011; Jackson et al., 2015; Jie et al., 2016; Ni et al., 2016) showed the dominance of *Proteobacteria* in microbial communities as well as community specialization. In lake sediments exposed to AMD gradients, the effects of different metals on specific microbes and microbial activities were correlated with their respective chemical properties. All these studies used 16S rRNA gene analysis, except for one, which used deep coverage data from shotgun metagenome sequencing (Hemme et al., 2016).

In our study, we used a shotgun metagenomic sequencing approach to characterize the functional and taxonomic diversity of bacterioplankton from five lakes within a catchment that was historically exposed to a polymetallic contamination gradient (PCG) for over 60 years. As the PCG was previously correlated with taxon abundance variation (Laplante and Derome, 2011; Laplante et al., 2013), taxon-function decoupling was expected to occur in the most polluted lakes and be absent in less polluted or unpolluted lakes. Our first objective was to assess the taxonomic and functional signatures of bacterioplankton adaptation to PCG. Secondly, we aimed to provide insight into the interplay of biodiversity and ecosystem services under a stress gradient by analyzing taxon-function variation.

MATERIALS AND METHODS

Lake Characteristics and Locations

Over the last 60 years, the Rouyn-Noranda (Western Quebec, Canada) mining sites have dumped AMD with heavy polymetallic traces (Laplante and Derome, 2011) into surrounding lakes. We targeted five lakes in this area (**Supplementary Figure S1**). Among them, three have common surface water interconnected along the same hydrologic basin: Arnoux Lake (LAR-hc; highly polluted), Arnoux Bay (BAR-mc; medium levels of pollution), and Dasserat Lake (DAS-lc; the least polluted). The water polluted by AMD spreads from Arnoux to Dasserat Lake generating a polymetallic gradient over 20 km. Around 30 km to the south side of this natural system of connected lakes, Opasatica Lake (OPA-nc), which is a landlocked unpolluted site, was sampled and considered as an unpolluted negative control, and ca. 40 km to the north side, Turcotte Lake (TUR-hc), another landlocked site was selected as a highly polluted lake. Longitude and latitude coordinates are given in **Supplementary File S1**. This Western Quebec lake system is 425 km northwest of Ottawa, Ontario. The abandoned mine site is a source of tailings and eroded mine waste into the Arnoux River, which drains west to Arnoux Lake, Arnoux Bay, and then Dasserat Lake. These lakes are irregular in shape and the bathymetry reflects the relief of the underlying bedrock. The immediate surrounding area consists of hilly terrain, volcanic rocks, ultramafic rocks, mafic intrusions, granitic rocks, and early and middle Precambrian sediments (Alpay, 2016).

Metallic and Chemical Gradient Surveys

pH and Polymetallic concentration (Al, Cd, Cu, Fe, Mn, Pb, Zn) in the studied lakes was measured in June 2010, a year prior to the present study, using ICP VISTA Varian-axial mass spectrometer as described in Laplante and Derome (2011). Trace metal profiles showed a polymetallic gradient in the three interconnected lakes (**Supplementary Figure S1**). For each lake, we measured temperature (OPA-nc: 12°C; DAS-lc: 10°C; BAR-mc: 9.9°C; LAR-hc: 11.5°C; TUR-hc: 9.5°C). Dissolved organic carbon (DOC) were determined in each sample using a total organic carbon (TOC) analyzer (Shimadzu) following the non-purgeable organic carbon (NPOC) method (Laplante and Derome, 2011).

Water Sampling

Sampling was carried out in September 2011 by collecting 6 L of water per lake at a depth of 60 cm below the surface. Water samples were sequentially filtered (3 filters per sample), first through a 47-mm poly carbonate filter with 3-micron pore size, followed by a 0.22 μm nitrocellulose membrane filter (Advantec) using peristaltic filtration (Masterflex L/S Pump System with Easy-Load II Pump Head; Cole-Parmer, Vernon Hills, IL, USA). Duplicates of the 0.22 μm filter were placed into cryotubes at −80°C.

DNA Extraction and Metagenome Sequencing

Filters duplicates were pooled, then genomic DNA was extracted as described by Laplante et al. (2013). Library preparation (TruSeq DNA Illumina) of paired-end reads (2 × 100 bp read length) was performed by the McGill University/Genome Quebec Innovation Center for whole metagenomic shotgun sequencing using a HiSeq™ 2000 Sequencing System. A total of 30 Gbps were obtained and the sequencing data summary is shown in **Supplementary File S2**. The sequence files are available from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra), *BioProject ID*: PRJNA449990.

Bioinformatic and Statistical Analysis

Reads-Based Approach (Figure 1)

To first discard methodological biases including sequencing artifacts, we pre-processed data for quality filtering, chimeric sequences, homopolymers, and short reads (cutoff: 50 bp)

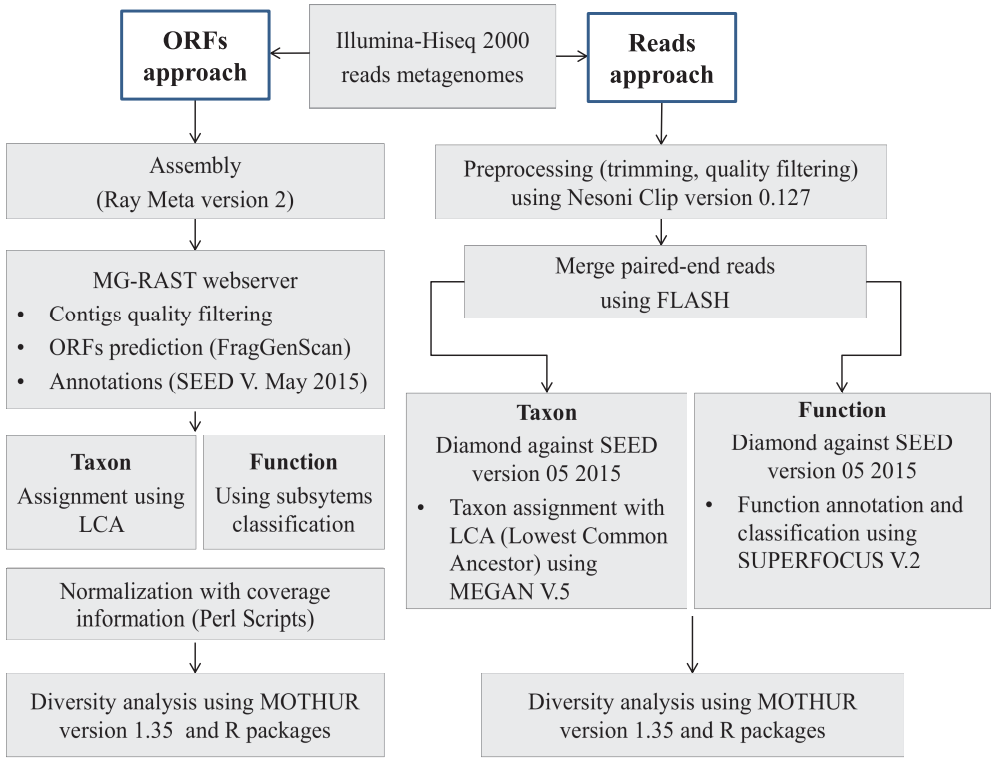


FIGURE 1 | Bioinformatics analysis pipeline. Two approaches were developed for this work. With the ORF-based approach, *de novo* assembly was performed on raw reads data using *Ray Meta* software. Then, the predicted ORFs (Open Reading Frames) were annotated using Diamond similarity research tool against SEED, which is a curated database. With the reads-based approach, merged (with FLASH algorithm) and filtered reads (with Nesoni) (length ~200 pb in average) were annotated using the same tool as the ORF-based approach, Diamond algorithm and SEED database. For both approaches, we used the lowest common ancestor (LCA) algorithm in the taxonomic assignment and the subsystems hierarchy in function classification. Diversity measures were computed using mothur software.

using the Nesoni Clip tool (https://github.com/Victorian-Bioinformatics-Consortium) version 0.133. Overall, the quality of forward reads (R1) was better than reverse reads (R2). This difference is related to sequencing quality decrease over the length of reads, in addition to the loss of enzymatic specificity overtime in the paired-end platform technology. Base calling quality was selected at a Phred or Q score of 33 (**Supplementary File S2**). FLASH software v1.2.11 (Magoč and Salzberg, 2011) was used with default parameters (10–65 bp overlapping window) to merge paired-end reads.

As a second step, following the selection of good quality reads for all five metagenomic samples, a sequence similarity search was performed against the SEED database (Version: May 2015) (Overbeek et al., 2014) using Diamond v0.7.9.58 (Buchfink et al., 2015). The taxonomic content of each sample was assigned using the Lowest Common Ancestor (LCA) method (Huson et al., 2007). Functional abundance was estimated using the SUPERFOCUS software (Silva et al., 2015) with Diamond ($1e^{-12}$ as *p*-value, 60 identity as threshold, 30 base pairs as minimum alignment length). To cope with missing biological replicates and unequal read numbers across all five lake samples (varying from 37–80 million paired-end reads before filtration), a read subsampling approach without replacement was used instead of rarefying or simulating reads from complete genomes. Accordingly, each metagenomic sample was subsampled 12 times with an equal number of reads (1 million reads). The uniformity in terms of number of subsampled reads from all samples were largely respected as in previous studies using simulated metagenomes (Mavromatis et al., 2007; Garcia-Etxebarria et al., 2014). The 60 generated metagenomic pseudo-replicates of equal size were submitted to our custom pipeline of taxonomic and functional abundance annotations.

Thirdly, to measure alpha and beta diversity based on feature abundances, we employed the OTU concept of taxonomic units (Schloss et al., 2009). Considering each feature (genus, function) as an OTU, alpha and beta diversity were computed using Mothur software (Schloss et al., 2009). UniFrac distances based on shared and unshared features were computed for each compared pair of samples. To inspect how environmental factors impact metacommunity composition, subsampled sites were first plotted based on their feature abundances with non-metric multidimensional scaling (NMDS) using Bray–Curtis distance between samples. Hence, the OPA-nc sample was used as control reference to compute the differential abundance of genera. Then, the computed distance matrix was clustered with Ward's method based on minimum variance. Clusters of genus abundance were distinguished with different colors on the NMDS plot. Next, mixed metals metadata were projected on NMDS axes by fitting a

regression model. The significance of the "regression coefficient" of the model was computed using a random permutation test (1,000 iterations). Then the regression coefficient between the randomized response and the fitted values from the model was computed. The NMDS model was run using the VEGAN package (Oksanen et al., 2016) in the R statistical environment (R Foundation for Statistical Computing, 2008). To test for a correlation between taxonomical and functional composition, a Canonical Correlation Analysis using the CCA (González et al., 2008) and mixOmics (Rohart et al., 2017) packages in R were applied. With CCA, the function-taxon cross-correlation was computed by maximizing the linear combinations between the two matrix vectors. Then a regularization function of CCA from mixOmics was used to deal with the high number of features (genus, function) compared to the low number of samples (60 subsamples). Regularization parameters (λ1 and λ2) were determined through a standard cross-validation (CV) procedure on a two-dimensional surface. The optimal value for λ was obtained by searching for the largest *CV-score* on the 2D surface that requires intensive computing time to converge for the optimal cross-validation value. Choice of canonical dimensions and graphical representation of features and samples were performed with mixOmics package.

ORFs-Based Approach (Figure 1)

To improve annotation accuracy in terms of length and coverage, an Open Reading Frame (ORF) prediction approach was used after *de novo* assembly. Collinear metagenomic reads belonging to the same genetic unit were merged into contiguous sequences (contigs). Firstly, *de novo* assemblies of raw reads were performed using the *RAY Meta* (Boisvert et al., 2012) assembler. Secondly, to explore contig features and gene contents, contigs were submitted to the MG-RAST webserver (Glass et al., 2010) and ORF prediction was conducted using the FragGeneScan tool (Rho et al., 2010). Afterwards, contigs were annotated with the BLAT tool implemented in MG-RAST against the SEED database using stringent filtering parameters ($1e^{-12}$ as *p*-value, 85% identity as threshold, 50 base pairs as minimum alignment length). Statistical summaries of annotated contigs are available in **Supplementary File S2**. Customized microbial annotations from the MG-RAST webserver were improved using the RESTful API tool (Wilke et al., 2015). An additional similarity research step based on BLASTx (parameters; identity threshold of 85%, *e*-value of 10^{-12} and minimum alignment length of 50 base pairs) (Camacho et al., 2009) was performed on contigs against the BacMet database (Pal et al., 2014) for annotating all polymetallic resistance genes (hereafter termed PMRGs). After annotations, contig coverage information determined by the Ray Meta assembler was added to normalize abundance information. Then, both abundance matrices of taxon and function coverage (both normalized and non-normalized) were analyzed with the STAMP software using a differential proportion comparisons test (Parks and Beiko, 2010). In a second additional workflow analysis, the ORFs were locally annotated with Diamond as described above in the "Reads-based approach" section. BLAT and Diamond provided similar annotation results. To measure *alpha* and *beta* diversity within and between communities, abundance matrices were adapted for the Mothur software. At the third step, metabolic abundance was analyzed using MG-RAST metabolite annotations. The metabolic differential abundance was surveyed using iPATH (Yamada et al., 2011); this tool offers the visualization of shared and specific pathways between pairs of samples.

RESULTS

Decoupling Taxon-Function

To investigate the impact of the polymetallic selection gradient on lake metacommunity composition, we measured the pattern of decoupling between taxon and function along the contamination gradient of the five lakes. We hypothesized that taxon-function decoupling pattern is an adaptive response of lake metacommunities. To detect this pattern, we performed two independent analyses: (i) taxonomic structure vs. functional diversity and (ii) canonical correlation of taxon and function.

Detangled Taxonomic Structure and Function Diversity

According to *alpha-diversity* analysis, the highest value of community *richness (chao index)* at the genus level was recorded in BAR-mc (OPA-nc: 126.8, DAS-lc: 115.07, BAR-mc: 212.6, LAR-hc: 121.66, TUR-hc: 68). In contrast to richness, community *evenness (Shannon index)* was lowest in TUR-hc (0.116), intermediate in OPA-nc (2.24), gradually decreasing along the metallic gradient from DAS-lc (2.85) < BAR-mc (2.58) < LAR-hc (2.47). However, community *evenness* of functions (OPA-nc: 2.36, DAS-lc: 2.32, BAR-mc: 2.92, LAR-hc: 2.62, TUR-hc: 2.87) was higher in BAR-mc, LAR-hc and TUR-hc then OPA-nc and DAS-lc. Then, *beta-diversity* analysis at the genus level (**Figure 2F**) revealed two patterns of structural convergence: (i) between the two independent lake communities, namely the unpolluted control (OPA-nc) and the low polluted lake (DAS-lc); (ii) between the interconnected BAR-mc-LAR-hc and the polluted control TUR-hc communities. Concerning functional diversity distribution, *beta-diversity* of all subsystems (**Figure 3B**) revealed two convergent patterns: (i) between the polluted control (TUR-hc), the highly-polluted gradient lake (LAR-hc), and the medium-polluted lake (BAR-mc); (ii) between the independent lake communities, namely the low-polluted DAS-lc and the negative control (OPA-nc) communities.

Canonical Correlations of Taxon and Function

Regularized canonical correlation analysis (rCCA) of function (subsystems level 1, 2, 3) with taxon was assessed using a maximal cross-validation criterion (see Materials and Methods). To detect linear combinations between function and taxon we separately performed the same rCCA analysis for OPA-nc/DAS-lc, and then for BAR-mc/LAR-hc/TUR-hc. For OPA-nc/DAS-lc (**Figures 6A–D**), we found a maximum variance of only 1% explained by the first axis computed from the taxon covariance matrix, and 1% explained by the first canonical correlation principal component computed from the function subsystems level 1 (results not shown) covariance matrix, even

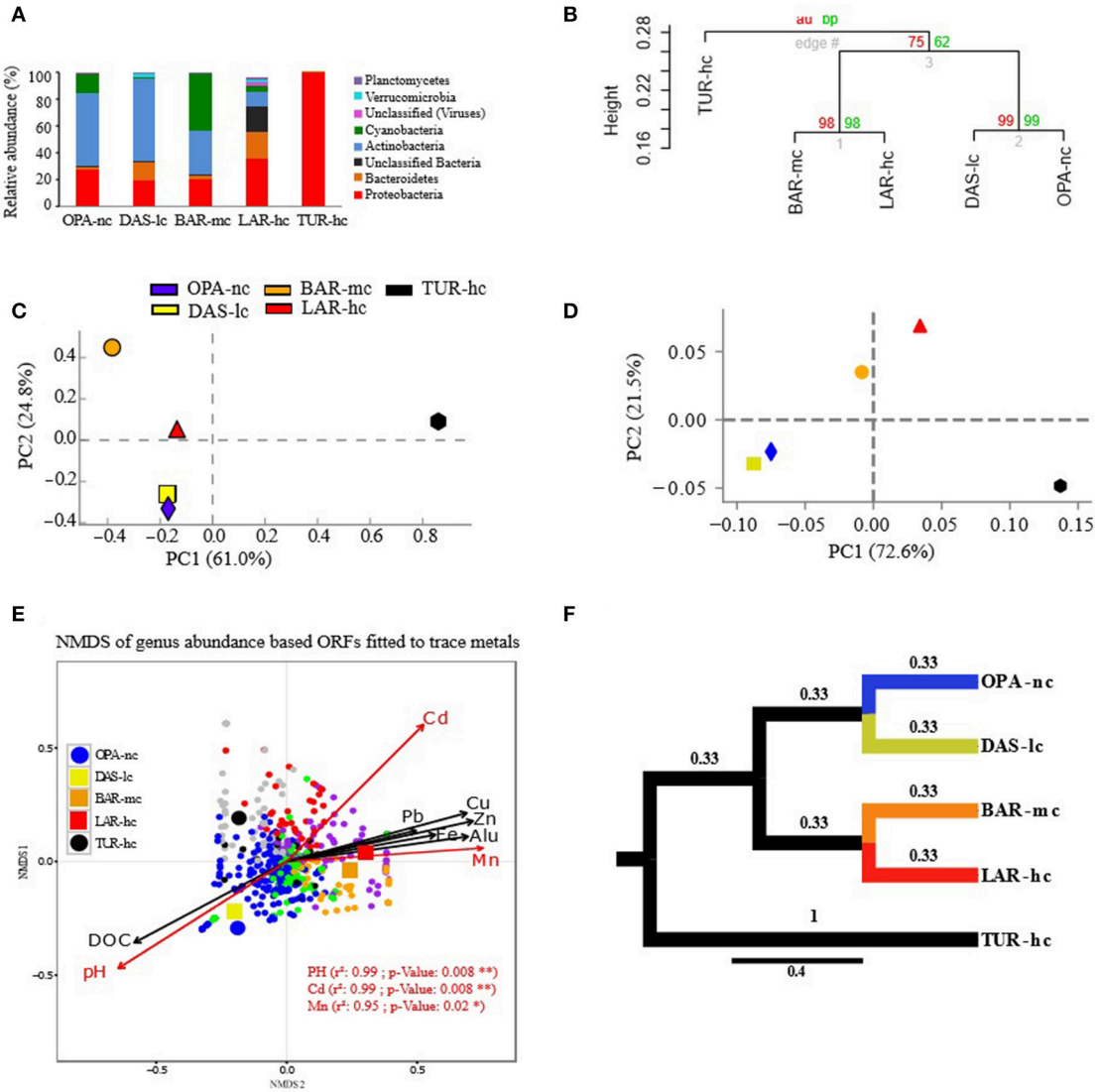


FIGURE 2 | Composition of metacommunities based on the ORF approach. **(A)** Metacommunity composition (y-axis) is shown in stacked bars for each lake metagenome (x-axis). Only phyla with relative abundance (RA) greater than 1% are shown. **(B)** Hierarchical clustering of samples based on genus RA using Ward's method and Bray–Curtis dissimilarity distance, bootstrap AU (Approximately Unbiased) *p*-value and BP (Bootstrap Probability) values are shown on the nodes. **(C,D)** Principal Component Analysis (PCA) of samples based on genus RA with different annotation parameters of alignment length cutoffs (50 pb in c and 30 bp in d) and identity threshold (85% in **C** and 60% in **D**). **(E)** NMDS of genera abundance fitted to trace metals was performed with Bray–Curtis distance, three dimensions were *a priori* defined for distance rank ordination and stress value was below 0.05. Cadmium (Cd), Manganese (Mn), and pH significantly fitted with NMDS axes are highlighted in red. NMDS loadings (NMDS1, NMDS2), and *P*-value of correlation r^2 of trace metals were reported in **Supplementary File S6**. Each small dot represents the ordinated genus, while each large point represents the lake communities' samples using a circle for OPA-nc in blue and the control TUR-hc in black, and the connected lakes are illustrated with squares (LAR-hc in red, BAR-mc in orange and DAS-lc in yellow). Genus plot coordinates, clusters and dot labels are shown in **Supplementary File S4**. **(F)** Tree based Unifrac distance computed with mothur is indicated by branch lengths. All these results were obtained using the ORF based approach with 85% identity threshold, *e*-value of 10^{-12}, minimum alignment length of 50 base pairs, and the lowest common ancestor (LCA) algorithm for taxonomic assignment. OPA-nc (Opasatica Lake) is the negative control; DAS-lc (Dasserat Lake) is low polluted; BAR-mc (Arnoux Bay) is medium polluted; LAR-hc (Arnoux Lake) is highly polluted, and TUR-hc (Turcotte Lake) is the positive control of contamination.

when we tested the canonical model at the most accurate hierarchical functional resolution (subsystems level 3). Supported by a high cross-validation score (0.975), this result suggested a strong coupling between taxon and function. Conversely, in BAR-mc/LAR-hc/TUR-hc (**Figures 5A–D**) we found the first axis explained 25% of the variance computed from the taxon covariance matrix, and 3% (Subsystem level 1) to 6% (subsystems level 3) explained by the first canonical axis computed from the functional covariance matrix. This result, supported with a high cross-validation score (0.99), revealed a weak correlation between taxon and function, thus suggesting a strong taxon-function decoupling. Using the first two canonical axes, in BAR-mc,

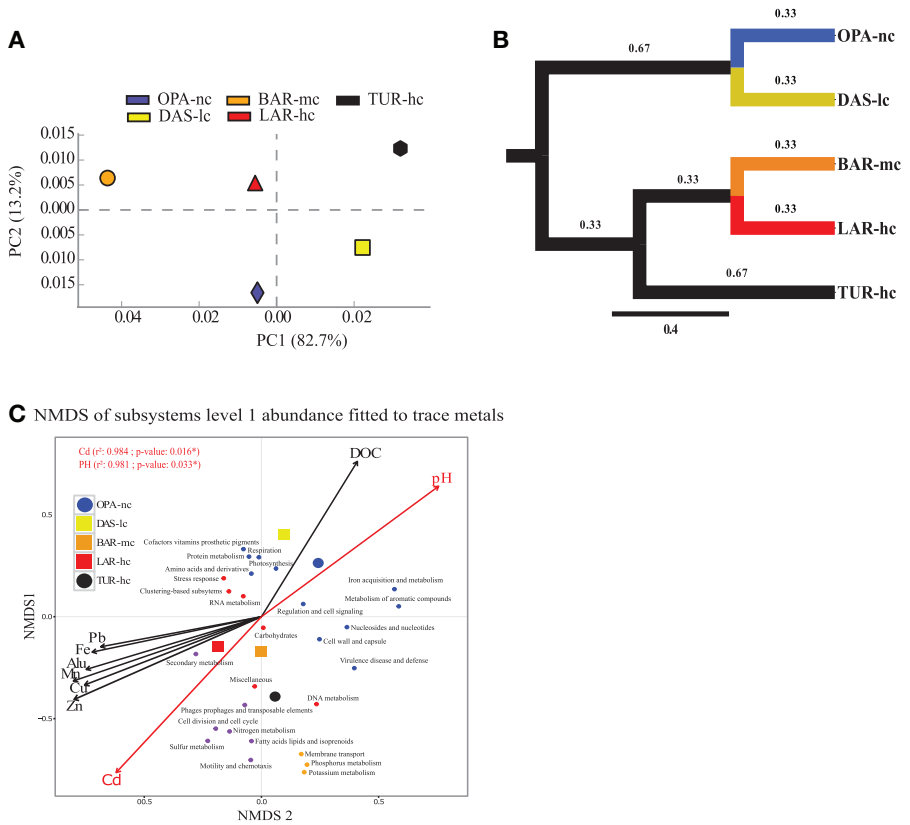


FIGURE 3 | Function abundance classification based on ORF approach. Composition analysis of metacommunity functions based on relative abundance (RA) of subsystems using principal component analysis (PCA) **(A)**, tree based Unifrac distance **(B)** and NMDS **(C)** identified the same pattern. **(A)** PCA figure was obtained from STAMP software (Parks et al., 2014). **(B)** NMDS (three *a priori* predefined dimensions projected into two dimensions, stress value < 0.05, Bray–Curtis distance) axes of all annotated subsystems level 1 fit significantly with Cadmium (Cd), and pH using the ORF approach. In d, each small dot represents a subsystem, while the large dot does represent the lake metagenome indicated with a circle for the negative control lake (OPA-nc) in blue and the positive control lake (TUR-hc) in black. The connected lakes are illustrated with squares (LAR-hc in red, BAR-mc in orange and DAS-lc in yellow). NMDS loadings (NMDS1, NMDS2), and *P*-value of correlation r^2 of trace metals were reported in **Supplementary File S6**. Subsystems plot coordinates, clusters, and dot labels are resumed in **Supplementary File S4**. **(C)** Tree based Unifrac distance computed with mothur is indicated by branch lengths In the ORF based approach the following parameters were strictly respected; 85% identity threshold, *e*-value of 10^{-12}, minimum alignment length of 50 base pairs, and the lowest common ancestor (LCA) algorithm for taxonomic assignment. OPA-nc (Opasatica Lake) is the negative control; DAS-lc (Dasserat Lake) is low polluted; BAR-mc (Arnoux Bay) is medium polluted; LAR-hc (Arnoux Lake) is highly polluted, and TUR-hc (Turcotte Lake) is the positive control of contamination.

LAR-hc, and TUR-hc, a clear separation was observed between taxon and function (**Figure 5C**), while the axes are superimposed in OPA-nc and DAS-lc (**Figure 6C**).

Taxonomic Variation Signatures

The metacommunity composition analysis emphasized three major patterns marked by abundance shifts within and between *Proteobacteria*, *Cyanobacteria* and *Actinobacteria* phylum (**Figure 2A**). In the first pattern, ***Proteobacteria*** mostly dominated by *Betaproteobacteria* (**Supplementary File S3**) reached a higher relative abundance in the highly-polluted (hc) lakes TUR-hc (99%) and LAR-hc (35%) compared to the moderately-polluted (mc) lake BAR-mc (20%), the least-polluted (lc) lake DAS-lc (19%), and the unpolluted (nc) lake OPA-nc (27%). At the genus level, *Polynucleobacter*, *unclassified Burkholderia,* and *Burkholderia* were the most dominant within polluted lakes TUR-hc, LAR-hc, and BAR-mc, respectively, while *Polaromonas* was the most dominant in DAS-lc and OPA-nc (**Supplementary File S3**). In the second pattern, ***Actinobacteria*** were the most dominant phylum (**Supplementary File S3**) in less polluted lakes [OPA-nc (53%) and DAS-lc (62%)] and their relative abundance gradually decreased in more polluted lakes [BAR-mc (33%), LAR-hc (10%) and completely disappeared in TUR-hc], mainly for the five most abundant genera: *Streptomyces*, *Frankia*, *Mycobacterium*, *Kribbella,* and *Nocardioides* (**Supplementary File S3**). In the third pattern, *Cyanobacteria* (**Supplementary File S3**) were abundant in OPA-nc (15.4%) and BAR-mc (42%), and much less frequent in LAR-hc (4.4%), DAS-lc (0.2%), and TUR-hc (< 0.01%). At genus level, *Synechococcus* was most dominant, accounting for 98% and 92% of *Cyanobacteria* genera in OPA-nc and DAS-lc, respectively. In contrast, distinct *Cyanobacteria* genera were dominant in polluted lakes: the filamentous *Anabaena* in BAR-mc, unclassified *Cyanobacteria* in LAR-hc, and both

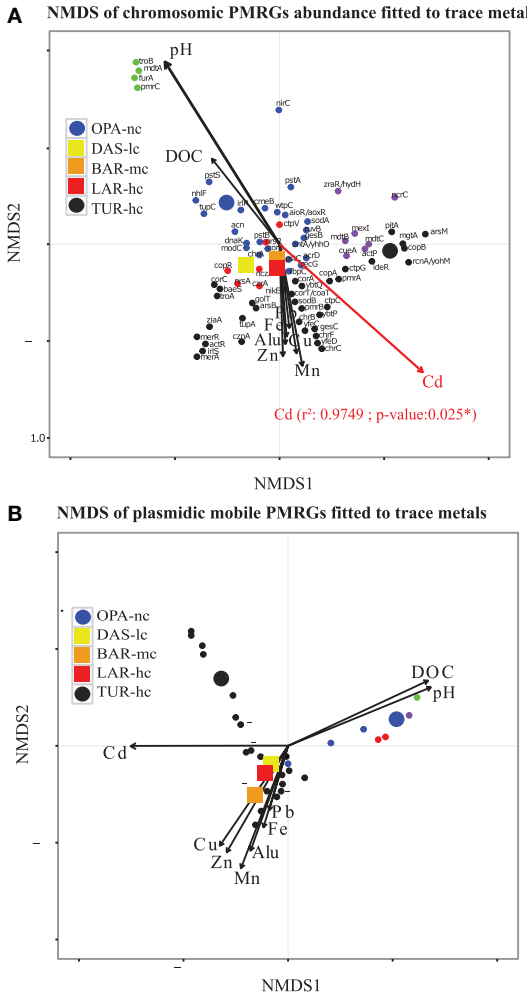


FIGURE 4 | Polymetallic resistance genes (PMRG) abundance correlation with trace metals. **(A)** For PMRG on chromosomes (72 genes), Cadmium (Cd) was significantly correlated with NMDS axes and it was the main explanatory factor of abundance variation of these genes between metacommunities. **(B)** NMDS axes based on relative abundance of PMRG located on plasmids (27 genes) do not significantly fit with any trace metal arrows. This NMDS analysis was performed with Bray–Curtis distance, three dimensions were *a priori* defined for distance rank ordination and stress value was below 0.05. NMDS loadings (NMDS1, NMDS2), and *P*-value of correlated trace metals are reported in **Supplementary File S6**. Each small dot represented an individual PMRG, while each large point represents the lake communities' samples using circles for OPA-nc in blue and the control TUR-hc in black, and the connected lakes were illustrated with squares, LAR-hc in red, BAR-mc in orange and DAS-lc in yellow. PMRG plot coordinates, clusters and dot labels are shown in **Supplementary File S4**. Thresholds of 75% of identity, minimum alignment length of 50 base pairs and *e*-value of 10^{-12} parameters were strictly respected. PMRG were annotated by performing Blastn of ORFs against BacMet database using Diamond software.

the diazotrophic *Cyanothece,* and the filamentous *Anabaena* in TUR-hc.

Lake metacommunity abundance shifts were further documented using bootstrapped hierarchical classification and PCA. At the genus level, both methods showed similar pattern of clustering with high statistical support (bootstrap values above 75%; more than 95% of explained variation by the first two PCA components), with BAR-mc, LAR-hc, and TUR-hc grouped separately from OPA-nc, DAS-lc (**Figures 2B–D**).

Role of Trace Metals in Taxonomic Variation Signatures

NMDS analysis based on ORFs (**Figure 2E**) revealed interesting relationships (significant R-squared indicating regression model's goodness of fit) between taxonomic abundance and different factors such as pH, DOC and trace metals (mainly Cadmium). OPA-nc and DAS-lc were significantly correlated with DOC and pH axes, while all other sites exposed to polymetallic gradient (BAR-mc, LAR-hc, and TUR-hc) were significantly correlated with trace metals axes (**Figure 2E**). To further analyze the link between abundance shifts at different taxonomic ranks and the trace metal gradient, the same NMDS analysis was performed using the ORFs approach. The abundance of *Proteobacteria* (**Supplementary Figure S3a**), *Actinobacteria* (**Supplementary Figure S3b**), and *Cyanobacteria* (**Supplementary Figure S3c**) were studied separately. NMDS analyses of abundance shifts at the genus level revealed significant correlations with different metal axes, pH and DOC. The shifts in composition within lake metacommunities were not explained by the same factors. For example, variation in the abundance of *Proteobacteria* among lakes was mainly explained by Cd, pH, Mn, Alu, while Cd and Fe explained variation in the abundance of *Actinobacteria*, and Alu and Mn were the main factors explaining variation in the abundance of *Cyanobacteria* among lake metacommunities.

Function Variation Signatures

Our results showed 6,801 annotated functions from all communities distributed into 988 subsystems in level 3, 192 subsystems in level 2, and 28 subsystems in level 1 (see sheet 2 in **Supplementary Figure S5**). At the first level (see **Supplementary Figure S5** and **Supplementary File S5**), our results of cross-metagenomes comparison suggested that the relative abundance of "Photosynthesis," "Cofactors, Vitamins, Prosthetic Groups, Pigments," and "Respiration" subsystems was significantly highest in OPA-nc while the "Stress response" was the lowest in this lake. However, Subsystems of "RNA metabolism," and mobile elements (Phages, prophages, plasmids, and transposable elements) showed the highest abundance in BAR-mc, followed by LAR-hc and TUR-hc, and low abundance in OPA-nc and DAS-lc. Furthermore, the relative abundance of the "carbohydrates" subsystem decreased gradually in all lakes except from DAS-lc to OPA-nc (**Supplementary File S5**). Interestingly, among the 28 subsystems (Level 1), four subsystems "Nitrogen metabolism," "Cell cycle and division," "Sulfur metabolism," "and "Motility and Chemotaxis" decreased gradually along the contamination gradient. In addition, three subsystems (Phosphorus and Potassium metabolism, Membrane transport) were absent in LAR-hc and showed specific profiles of low abundance (**Supplementary File S5**) varying between 0.2 and 3.8% in

BAR-mc and TUR-hc. For multiple subsystems in Level 1 ($n = 12$), no gradual abundance variation was observed. However, at a deeper resolution, many important functions related to metals transport and resistance from the "Virulence defense and disease," "Membrane transport," and "Iron acquisition and metabolism" subsystems showed few gradual (i.e., Cobalt-Zinc-Cadmium resistance) abundance profiles and high specific abundance per lake (**Supplementary Figure S7**). At the functional level, variation abundance was detectable within all subsystems where three profiles of abundance variation were observed from OPA-nc to TUR-hc: (i) profile 1 (FP1) represents gradual function abundance decrease (106 functions) along the contamination gradient (**Supplementary File S5** and **Supplementary Figure S10**), (ii) profile 2 (FP2) represents gradual function abundance increase (123 functions) along the contamination gradient **Supplementary File S5** and **Supplementary Figure S10**, and (iii) profile 3 (FP3) represents specific functional abundance (**Supplementary File S5** and **Supplementary Figure S11**) in control negative OPA-lc (167 functions), or in polluted lakes (225 functions). These functional profiles were not necessarily observed in one subsystem, but rather multiple profiles were detectable within one subsystem (**Supplementary File S5**). For example, under the "Virulence, Disease and Defense" subsystem, we observed all these profiles with functions related to metal resistance FP2 (i.e., Cobalt-zinc-cadmium CzcA protein, Cation efflux system protein CusA), and FP1 (i.e., Magnesium and cobalt efflux protein CorC), and FP3 (i.e., Copper homeostasis) OPA-nc (see Virulence subsystem in **Supplementary File S5**). However, functions related to mobile genes and HGT agents (**Supplementary Figure S8**) were significantly more abundant in polluted lakes (e.g., Gene transfer agent proteins, conjugative transfer proteins, DNA repair, CRISPR associated proteins, integrons). Classification of functional abundance (subsystem levels 1, 2, 3) identified two independent clusters. The first cluster grouped BAR-mc, LAR-hc and TUR-hc, and the second grouped DAS-lc and OPA-nc (**Supplementary Figures S5**, **S6**). Similar topologies were obtained using both approaches: ORF (**Supplementary Figure S4b**) and reads subsampling (**Supplementary Figures S4c,d**). PCA analysis based on the ORF approach produced the same results, where at least 71% of variance was explained on the first PC for all subsystem function levels. We only presented a PCA plot for subsystems abundance in level 1, where more than 82% of variation in functional abundance was explained by the first component (**Figure 3A**). At the metabolic level, analysis of enzymes abundance profiles cross-metagenomes showed different topology which was a dichotomy between OPA-nc and all others pollution gradient lakes (See **Supplementary File S7** and **Supplementary Figure S9**).

Role of Trace Metals in Function Variation Signatures

NMDS analysis of functional abundance highlighted two main patterns of correlation (significant R-squared indicating regression model's goodness of fit) with metadata (**Figure 3C**). First, BAR-mc, LAR-hc, and TUR-hc were correlated with Cadmium axis ($p \leq 0.05$). Second, OPA-nc and DAS-lc were correlated with pH axis ($p \leq 0.05$). The same analysis performed on the subsystems in level 2 (192 functional modules) suggested a significant contribution of all studied factors (results not shown). At the finest functional level, lakes ordination based on the NMDS of polymetallic resistance genes (PMRG)s abundance showed a fit with the cadmium concentration gradient (**Figure 4**), where DAS-lc was ordinated near BAR-mc and LAR-hc. In NMDS analysis of PMRGs located on chromosomes (**Figure 4A**), only Cadmium played a significant role in explaining abundance variation. Similarly, the NMDS analysis of PMRGs located on plasmids provided the same classification profile even though they do not fit significantly with any metal traces (**Figure 4B**).

DISCUSSION

Decoupling Taxon-Function as a Signature of Adaptive Strategies

Comparing the compositional signatures of taxon and function, we observed that relative shifts in taxon abundance could only partially predict the impact of metallic toxicity on metacommunity structure (see section Role of Trace Metals in Taxonomic Variation Signatures). By considering the signatures of functional abundance of the subsystems explained by pH and Cadmium in polluted lakes, we could more accurately predict the impact of metallic contaminants on ecosystem services of lake metacommunities. In this respect, the contamination gradient explained much variation in community function structure and provided a powerful way to further assess the relationship between the distribution of functional abundance and selective pressure, which may increase gradually with the expelled AMD flow over time. The impact of the selection gradient on lake metacommunity composition was tested through two independent analyses, first using diversity measures, and second by detecting taxon-function decoupling patterns. Alpha taxonomic diversity suggest a switch in BAR-mc, while the gradual decrease in evenness based both taxon and function in OPA-nc: (2.2^{t}; 2.3^{f}), DAS-lc (2.8^{t}; 2.3^{f}), BAR-mc (2.5^{t}; 2.9^{f}), LAR-hc (2.4^{t}; 2.6^{f}), TUR-hc (0.1^{t}; 2.8^{f}) could be a potential consequence of composition homogeneity in community type (e.g., *Proteobacteria* in TUR-hc). Indeed, this observation may be related to the low complexity in AMD communities previously documented for the same lake system (Laplante and Derome, 2011; Laplante et al., 2013), and for other AMD metacommunities (Allen and Banfield, 2005; Huang et al., 2016).

The rCCA analysis allowed for the detection of significant spatial correlation between taxon and function in OPA-nc/DAS-lc, reflecting a coupling between taxon and function. In these unpolluted lakes, as mentioned above, NMDS analysis showed that environmental factors (Cadmium, pH, and DOC) explained variation in the overall taxonomic and functional composition. At high resolution (subsystems level 2, 3) NMDS showed a slight difference between OPA-nc and DAS-lc, but we cannot unequivocally associate these variations to trace metal ratios. We

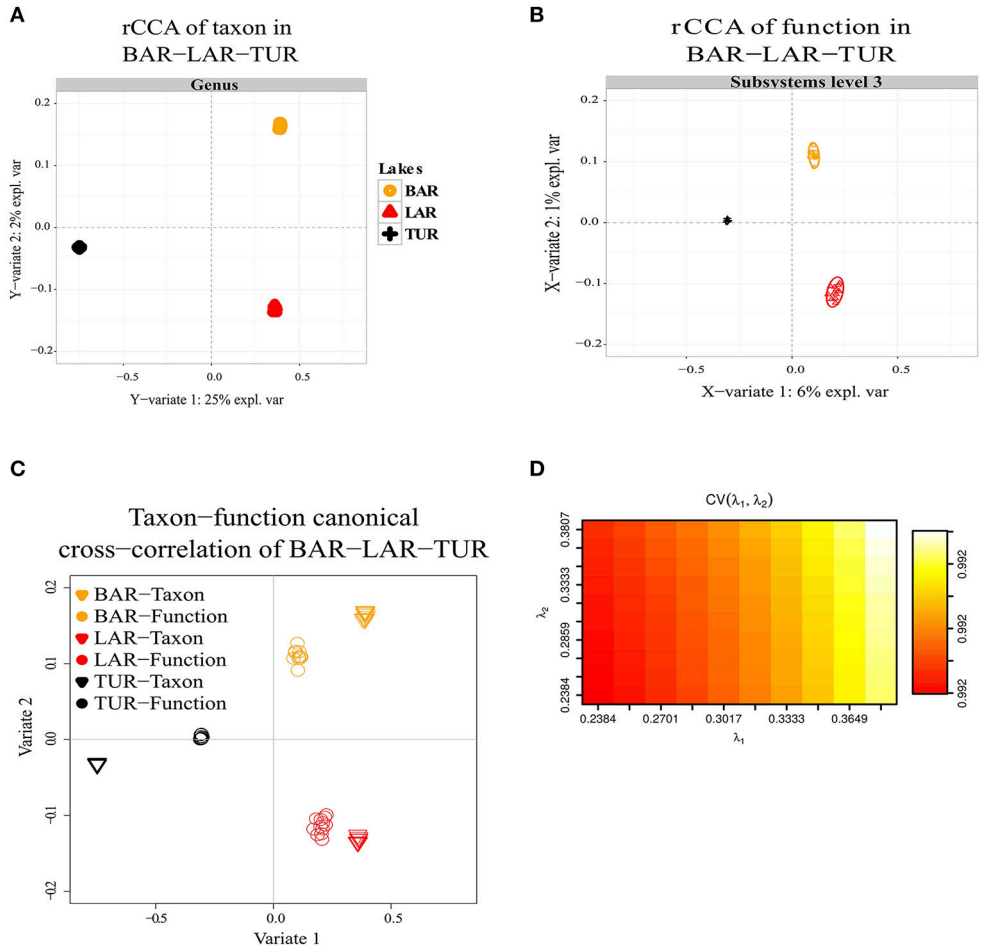


FIGURE 5 | Decoupling of taxon and function between metacommunities based on the subsampled reads approach. **(A–D)** Regularized canonical correlation analysis (rCCA) performed on BAR-mc, LAR-hc, and TUR-hc. **(A)** rCCA of a taxon (relative genus abundance) showed 25% of the explained variance on the first canonical component. **(B)** rCCA of function (subsystems level3 relative abundance) showed only 6% of the explained variance on the first canonical component. **(C)** The canonical cross-correlation of taxon-function identified a decoupling pattern. **(D)** Cross-validation score converged to a maximum value of 0.99 when regularization parameters λ1 and λ2 were both fixed at 0.375.

may have missed other explanatory environmental and chemical variables (i.e., NFigure$_2$, NO_3, SO_4, PO_4), or the potential variation resulting from neutral ecological process, drift or random reproduction as observed in wastewater habitats (Ofiteru et al., 2010). Such coupling is not necessarily absolute but partial, owing to the presence of some differentiated sub-communities performing the same ecosystem services. In pristine natural conditions (without stressful anthropogenic inputs), coupling between taxon and function was observed in freshwater lakes (Langenheder et al., 2005; Debroas et al., 2009), and decoupling was observed in oceanic bacterial communities from contrasted environments (Louca et al., 2016b).

Overall, in the present study, we found that functional variation between polluted and unpolluted lakes was better explained by environmental factors than taxonomic variation between and within functional groups. Concerning the three lake communities facing exposure to a polymetallic gradient (BAR-mc/LAR-hc/TUR-hc), the explained variance between taxon (25%) and function (6%) strongly suggests a decoupling between taxa and functions. The shared functions in these three polluted lakes reflect a convergent pattern, which in turn could be interpreted as a predictive signature of the ecosystem service's impairment associated with acid mine lake water. This conclusion is further supported by the NMDS results, where the distribution of polluted lakes fitted closely to Cadmium. In addition to rCCA, when comparing tree topologies of structure and function (**Figures 2F**, **3C**), we detected additional patterns of taxon-function decoupling, like the PCG. Such an approach offers interesting insights into the adaptive strategies used by metacommunities facing long-term exposure to polymetallic pollution. Often interpreted as an indicator of HGT in natural communities (Ram et al., 2005; Green et al., 2008; Burke et al., 2011; Louca et al., 2016a,b) and AMD communities (Navarro et al., 2013; Devarajan et al., 2015; Chen et al., 2016; Hemme et al., 2016), taxon-function decoupling may provide evidence for selective pressure on microbial communities (e.g., exerted by metallic exposure). Indeed, as mentioned above, multiple proteins playing a role in

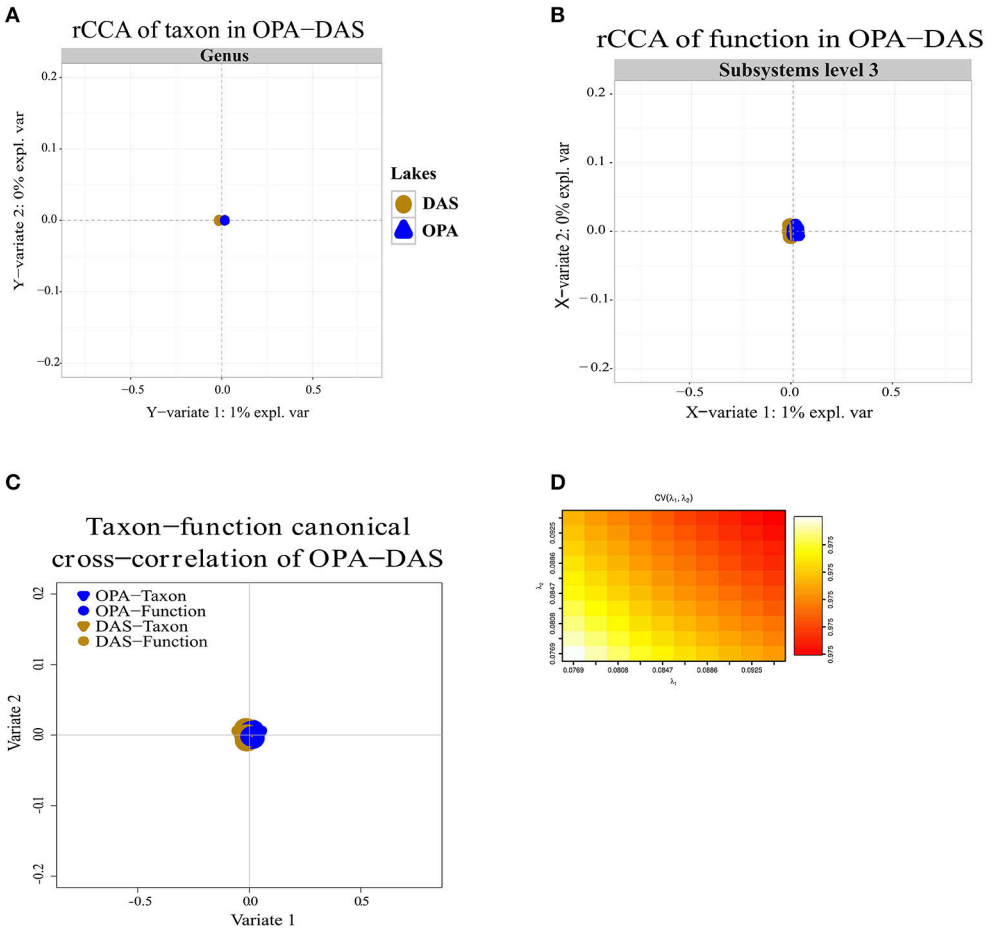


FIGURE 6 | Coupling of taxon and function between metacommunities based on the subsampled reads approach. **(A–D)** Regularized canonical correlation analysis (rCCA) performed on OPA-nc, and DAS-lc. **(A)** rCCA of the taxon (relative genus abundance) showed 1% of the explained variance on the first canonical component. **(B)** rCCA of function (subsystems level3 relative abundance) also showed 1% of the explained variance on the first canonical component. **(C)** The canonical cross-correlation of taxon-function identified a coupling pattern between taxon and function. **(D)** Cross-validation score converged to a maximum value of 0.975 when regularization parameters λ1 and λ2 were both fixed at 0.0925. The rCCA method was applied using mixOmics and CCA package in R.

HGT, such as cassettes of integrons and transposable elements, were present in polluted lakes, and absent in an unpolluted lake (OPA-nc). We observed more than 14 mobile PMRGs located on plasmids, and only two PMRGs on both plasmids and chromosomes. The plasmid location of these PMRGs indicates that bacterial conjugation may be a vector for HGT. Interestingly, a heatmap of abundance clustering from chromosomal and plasmid PMRGs (figure not shown) produced a similar topology of functional profiles (i.e., OPA-nc; DAS-lc-BAR-mc; LAR-hc-TUR-hc).

Evolutionarily speaking, such taxon-function decoupling patterns are expected to be signature of adaptation within communities between closely, but also distantly, related bacterial strains. Consequently, community composition in BAR-mc, LAR-hc or TUR-hc may have independently evolved via HGT events of resistance and regulatory genes. According to functional abundance results, the potential occurrence of HGT is higher in LAR-hc and TUR-hc compared to BAR-mc, which is closer to DAS-lc and OPA-nc in terms of functional distribution. A subset of adaptive beneficial transferred genes is expected to reach fixation (Lind et al., 2010), but the long term metallic contamination may have funneled the "metal resistance gene pool" into different evolutionary trajectories due to the mounting selective pressure.

Taxonomic Adaptive Signatures

In this study, the overall taxonomic variation suggests three salient patterns of abundance distribution. First, a "composition gradient" pattern constituted three shifts in taxonomic structure: (i) high abundance of *Proteobacteria* in polluted sites (TUR-hc, LAR-hc; BAR-mc), (ii) high abundance of *Actinobacteria* in unpolluted sites (OPA-nc, DAS-lc), (iii) intermediate levels of *Cyanobacteria* in all sites, with *Nostocales* being abundant in polluted lakes and *Chroococcales* abundant in unpolluted

lakes (**Supplementary File S3**). Second, a "community type" pattern suggests that the overall metacommunity exhibited compositional shifts along the five lakes from wide (phylum) to narrow (genus) taxonomic levels. Third, a "taxonomic convergence" pattern highlights parallel changes of community taxonomic structure, thus confirming previous results based on semi-quantitative and quantitative studies (Laplante and Derome, 2011; Laplante et al., 2013).

To further reinforce the taxonomic composition analysis, we examined genera abundance and ORF distributions. Similar ratios of ORFs/Genus were observed in the five studied metagenomes (**Supplementary Figure S2**). The number of annotated ORFs in all metagenomes was comparable. Furthermore, random subsampling analysis without replacement produced similar results (slightly different in topology) compared to the ORFs approach, with remarkable clustering fidelity of subsampled replicates from each metagenome (**Supplementary Figure S4a**). Here, the subsampling approach revealed consistency in the molecular signal of each lake. We acknowledge that the subsampling approach used in our analysis cannot replace real biological replications, but it is rather an indicator of the metagenomic data robustness to the metacommunity structure.

To understand the sources of variation in contributing to the three major shifts of relative abundance in community type, combined NMDS and correlational analyses were performed for each pattern of taxonomic variation. First, the *Proteobacteria* genus distribution of eight predefined clusters (**Supplementary File S4**) showed that abundance variation between communities was mainly explained by synergistic interactions of Cd, pH, Mn, and Alu (**Supplementary Figure S3a**). According to previous studies, *Proteobacteria* were among the most abundant phyla in acid mine water (Laplante et al., 2013; Streten-Joyce et al., 2013) and in freshwater lake sediments polluted by "heavy metals" (Ni et al., 2016). Second, in contrast to *Proteobacteria*, our results divided *Actinobacteria* into four genus abundance clusters (**Supplementary File S4**) constrained by two main and opposite explanatory factors, Cd and Fe (**Supplementary Figure S3b**). In fact, the most abundant *Actinobacteria* genera (*Streptomyces, Frankia, Mycobacterium*), which varied between polluted and unpolluted lakes, fall in the same abundance cluster (see *Actinobacteria* in **Supplementary File S4**). Indeed, some *Actinobacteria* (e.g., *Streptomyces*) strains are known to have different metal-resistance profiles (Álvarez et al., 2013). Interestingly, strains like *Mycobacterium* were able to transport and uptake Cd (Dimkpa et al., 2009). On the other hand, *Cyanobacteria* abundance showed different patterns of abundance in polluted and unpolluted lakes (**Supplementary File S4** and **Supplementary Figure S3c**) suggesting that *Chroococcales* (*Cyanothece, Microcystis, Synechocystis, Thermosynechococcus*) and *Synechococcales* (*Synechococcus, Prochlorococcus*) are much more affected by trace metals compared to the *Nostocales (Anabaena, Aphanizomenon, Cylindrospermopsis, Dolichospermum, Nodularia, Nostoc, Raphidiopsis)*. Although Cd was not identified here as a significant explanatory factor, diverse strains of *Nostocales* were documented to have the capacity to adsorb Cadmium (Pokrovsky et al., 2008) and trace metals (Mota et al., 2015). Interestingly, the sudden break of *Nostocales* lineages (**Supplementary File S3**) between the connected lakes DAS-lc, BAR-mc and LAR-hc is potentially related to resistance thresholds to trace metals, as higher levels become toxic to *Synechococcus* (Ludwig et al., 2015). Furthermore, the high relative abundance of *Chroococcales* and *Cyanobacteria* in OPA-nc and DAS-lc is potentially related to their role in photosynthesis and DOC mineralization (Bittar et al., 2015). Overall, our results show that metallic toxicity impacts metacommunity structure and provides a partial explanation for the relative shifts in abundance found in the lakes we studied. The dominance of *Proteobacteria* in over polluted communities confirms the result previously observed in the same lake system (Laplante et al., 2013), and from various acid mine waters in the world (Almeida et al., 2008; Hemme et al., 2010; Kuang et al., 2013; Stankovic et al., 2014; Wang et al., 2015).

Functional Adaptive Signatures

At the general level (subsystems level 1), only four subsystems showed gradual variation. At the function level, our results suggest deterioration in ecosystem services along the contamination gradient, as relative abundance of functional modules in 18 subsystems such as "*Carbohydrates,*" "*Photosynthesis,*" "*Cell division and cycle,*" "*DNA metabolism,*" and "*Respiration*" decrease gradually. However, under "*Virulence defense and disease*" and "*Membrane transport*" subsystems (level 1), many important metals transport and resistance functions (i.e., Cobalt-Zinc-Cadmium resistance) increased between OPA-nc and other lakes (**Supplementary Figure S7**). These profiles of gradual changes were less observable at the general level (subsystems level 1), and more detectable at the functional level resolution of many subsystems. The gradual decrease and increase in relative abundance proportions was clearly observed at le the lowest molecular function (i.e., *Photosynthesis* functions) along the polymetallic gradient. Overall, variation in the functional composition of metacommunities suggests convergence between BAR-mc/LAR-hc and TUR-hc, two geographically distant and independent lakes affected by independent AMD sources.

In contrast to the community classification based on taxonomic composition, BAR-mc is functionally closer to LAR-hc-TUR-hc than OPA-DAS. NMDS of functional composition, community hierarchical clustering, and PCA analysis all find the same classification results. Cadmium and pH were the main factors explaining functional composition variability among lakes. However, independent analysis performed on both PRMGs and enzymatic functions abundance showed that DAS-lc fitted within the polluted lakes (BAR-mc-LAR-hc-TUR-hc) instead of OPA-nc. PMRGs located on plasmids (**Figure 4B**) were differentiated from those located on chromosomes (**Figure 4A**) since plasmid genes are known to house more adaptive genes acquired via bacterial conjugation (Li et al., 2015). Only two experimentally confirmed genes (*copA* and *actP)* were found in both plasmids and chromosomes. *CopA* is involved in silver/copper export and homeostasis (Cha and Cooksey, 1991; Outten et al., 2001; Banci et al., 2003; Behlau et al.,

2011). Acetate Permease (ActP) controls copper homeostasis in rhizobium preventing low pH-induced copper toxicity (Reeve et al., 2002). NMDS analysis based on Chromosomal PMRG abundance revealed that Cadmium plays a significant role ($p \leq 0.05$) in shaping the differential abundance of these genes. Alternatively, analysis of plasmid PMRGs did not highlight any significant fit with metal axes (**Figure 4B**), owing to the low number of annotated PMRGs on plasmids. Using OPA-nc as an unpolluted reference in our comparative framework, differential metabolic abundance variation revealed an erosion of biosynthesis pathways along the contamination gradient (results of compared pathways not shown). Eroded metabolic functions were associated to degradation of aromatic compounds, amino acid biosynthesis, and carbohydrates, thus leading to the loss of major bacterial mediated ecosystem services. As bacterial communities experienced a consistent metallic stress over 60 years of mining activities, many functions associated with ecosystem services likely became energetically too expensive to be maintained. Such a selective environment may have led to community specialization. Community specialization has recently been demonstrated in soil AMD communities (Volant et al., 2014) and natural freshwater communities (Pernthaler, 2013; Salcher, 2013; Pérez et al., 2015). In summary, the two main elements (or factors) that explained the majority of the functional variation between polluted vs. unpolluted communities were pH and Cadmium concentration. Nonetheless, other metal trace gradients offered partial explanations for functional variation.

CONCLUSIONS

In this study, we examined adaptive signatures within natural lacustrine microbial communities living under a gradient of selective pressure induced by trace metal contamination from over 60 years of mining. Using a metagenomic approach based on whole genome shotgun sequencing, we identified a convergence in both taxonomic and function responses, thus providing evidence for genotypic signatures of adaptive evolution. Strong selective pressure may drive overall taxon-function decoupling, which may reflect the occurrence of gene loss and HGT induced by AMD gradient, or the result of strong selection exerted on existing strains possessing the necessary resistance genetic background. This study remains a preliminary assessment of decoupling phenomenon and further studies are eventually needed to understand in a deeper manner the nature of convergence between unpolluted environments vs. polluted environments in a context of stress gradient. At the taxonomic scale, metacommunity composition showed marked relative abundance shifts of major phyla, but was much more marked at the genus level, suggesting a "community type" adaptation to the metallic gradient within each ecological niche. At the function scale, we observed the erosion of metabolic pathways along the metallic gradient despite the higher abundance of functional categories like stress response, regulation, protein metabolism, and metallic resistance in polluted lakes compared to unpolluted lakes. Investigating the relationship of both taxonomic and functional signatures, we detected a decoupling pattern between taxon and function in polluted lakes as an indicator of adaptation potentially via HGT. These results suggest, for the first time, a decoupling pattern of taxon-function within natural communities adapted to a gradient of polymetallic contamination. This decoupling pattern highlights the gap between microbial biodiversity and ecosystem services in polluted environments.

AUTHOR CONTRIBUTIONS

ND conceived the experiment. P-LM conducted the experiment. ML performed the data assembly. BC produced and analyzed the results. BC and ND wrote the manuscript. This project was under supervision of ND. All authors reviewed the manuscript.

ACKNOWLEDGMENTS

We thank Prof. Connie Lovejoy, Dr. Anne Dalziel, Dr. Martin Llewellyn, Dr. Amanda Xuereb, and Dr. Mohamed Alburaki for reading comments, Eric Normandeau for python scripts of data subsampling. We thank all Ph.D. students in ND laboratory for reading and comments. We also thank Hayan Hmidan for generating geographical maps and Dr. Sebastien Boutin for helps in water sampling. This project was funded by a grant from the Natural Science and Engineering Research Council of Canada (NSERC grant no. 6663) obtained by ND.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00869/full#supplementary-material

Supplementary Figure S1 | Geographical localisation and metallic profiles of sampled lakes. **(a)** Geographical localisation of the sampling sites located in Ryoun-Noranda (West Quebec, Canada) visited in June 2011. Latitute and longitude coordinates of sampling sites are 48.25005489 and −79.40574646 in Opasatica lake (OPA-nc); 48.07601448 and −79.3082428 in Dasserat lake (DAS-lc); 48.24090959 and −79.35012817 in Arnoux Bay (BAR-mc); 48.25051211 and −79.333992 in Arnoux lake (LAR-hc); 48.30474963 and −79.07742262 in Turcotte lake (TUR-hc). This map was produced using Arc GIS Esri® Arc Map™ 10.1 under academic license certification. **(b)** Trace metals concentrations measured in the five sampled lakes 1 year before this study (Laplante and Derome, 2011). The x-axis represents the log ratio of trace metal concentrations (mg/l) and the y-axis represents detection limit in each lake. The metallic gradient showed that Cadmium was under the detection limit in OPA-nc (negative control), at the detection limit in DAS-lc (low contamination), three times more than the detection limit in BAR-mc (medium contamination), LAR-hc (high contamination), and TUR-hc (positive control). Contamination gradient classification refers to the Cadmium log ratio across the five lakes.

Supplementary Figure S2 | Classification of lake metacommunities based ORF approach at genus and phylum levels. **(a)** Distribution of ORF and annotated genus in the five metagenomes. This figure showed that not only the number of predicted ORFs (**Supplementary File S2**), comparable between metagenomes but also the genus count (**Supplementary File S3**). **(b)** Hierarchical clustering of samples using Ward's method and Bray–Curtis dissimilarity distance, bootstrap AU (Approximately Unbiased) *p*-value and BP (Bootstrap Probability) value are shown on nodes. **(c,d)** principal component analysis (PCA) of samples based on genus relative abundance (RA) assigned with coverage **(c)** and without coverage **(d)** normalization. Metacommunity clustering based on genus abundance is different at phylum level where BAR-mc was closer to OPA-nc and DAS-lc. **(e)** PCA analysis of samples based on function RA with different annotation

parameters of alignment length cutoff (30 bp) and identity threshold (60%). **(f)** Distribution of filtered ORFs on different alignment length cutoffs. For the ORF based approach **(a–d)**, the 85% identity threshold, *e*-value of 10^{-12} and minimum alignment length of 50 base pairs parameters were selected in filtering annotations, and the LCA (Lowest Common ancestor) algorithm was used to assign taxonomy.

Supplementary Figure S3 | Composition of metacommunities based on the ORFs approach. NMDS (with Bray-Curtis distance) of genera abundance for major abundant phyla fitted to trace metals for *Proteobacteria* **(a)**, *Actinobacteria* **(b)**, *Cyanobacteria* **(c)**, the water pH, and trace metals which correlated significantly with NMDS axes were highlighted in red. Each small point in figures a, b, and c represented the genus abundance, while each big point does represent the lake metacommunities samples using circle shape for OPA-nc in blue and the control TUR-hc in black, and the connected lakes were illustrated with square shape, LAR-hc in red, BAR-mc in orange and DAS-lc in yellow. NMDS loadings (NMDS1, NMDS2), and *P*-value of correlation r^2 of trace metals were reported in **Supplementary File S6**. Genus plot coordinates, clusters and dot labels are resumed in **Supplementary File S4**.

Supplementary Figure S4 | Hierarchical clustering of taxon and function. **(a)** Hierarchical clustering of artificial replicates based on genus abundance using the subsampled reads approach. **(b)** Hierarchical clustering of samples based on abundance of subsystem level 1 using the ORF approach. **(c, d)** Hierarchical clustering of subsampled replicates based on subsystems level 1 and 3 using the reads approach. Hierarchical clustering was performed using Ward's method and Bray–Curtis dissimilarity distance; bootstrap AU (Approximately Unbiased) *p*-value and BP (Bootstrap Probability) value are shown on the nodes.

Supplementary Figure S5 | Heatmap of subsystems in level 1. This heatmap represents metagenomes classification based on subsystems in level 1 (See **Supplementary File S5**). Dendrogram's topology identified two clusters. The first cluster grouped BAR-mc, LAR-hc and TUR-hc, and the second grouped DAS-lc and OPA-nc. The hierarchical clustering of relative abundance proportions of subsystems, and of samples was performed using Ward's method and Bray–Curtis dissimilarity distance. The ORF approach was used with identity threshold of 85%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary Figure S6 | Heatmap of subsystems in all levels. Subsystems relative abundance were clustered cross-metagenomes in different levels, level2 (981 modules), level3 (192 modules) and function level (6801 functions) (See **Supplementary File S5**). The same topology was observed in level 1 (See **Supplementary Figure S5**) and in all levels. The hierarchical clustering of relative abundance proportions of subsystems, and of samples was performed using Ward's method and Bray–Curtis dissimilarity distance. The ORF approach was used with identity threshold of 60%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary Figure S7 | Heatmap of multiple subsystems abundant in function level. This heatmap represents cross-metagenomes, the common and most abundant functions (>2%) in 22 subsystems (See **Supplementary File S5**). Functions of polymetallic resistance (Cation efflux system protein CusA and Cobalt–zinc–cadmium resistance protein CzcA) showed a profile of gradual abundance increase along the pollution gradient. The hierarchical clustering of relative abundance proportions of functions, and of samples was performed using Ward's method and Bray–Curtis dissimilarity distance. The ORF approach was used with identity threshold of 60%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary Figure S8 | Subsystem of "Phages, prophages, plasmids, and transposable elements" cross-metagenomes. Under this subsystem multiple relevant functions (level 3) related to mobile elements and transfer vectors (Gene transfer agents, transposons, prophages, conjugative plasmids, integrons) were shared between DAS-lc, BAR-mc, LAR-hc, TUR-hc, and depleted in OPA-nc. However, each metagenome contains specific profile of mobile elements functions such like agents of gene transfers and conjugative elements in TUR-hc. The hierarchical clustering of relative abundance proportions of this subsystem modules, and of samples was performed using Ward's method and Bray–Curtis dissimilarity distance. The ORF approach was used with identity threshold of 60%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary Figure S9 | Metabolic abundance cross-metagenomes. This heatmap represents 1,842 annotated enzymes (See EC number in **Supplementary File S7**) in all samples. The hierarchical clustering of relative abundance proportions of enzymes, and of samples was performed using Ward's method and Bray–Curtis dissimilarity distance. The dendrogram shows dichotomy between OPA-nc metagenome and all others. The ORF approach was used with identity threshold of 60%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary Figure S10 | Gradual variation of functions cross-metagenomes. Two heatmaps represent gradual function abundance FP1 (106 functions) and FP2 (123 functions) along the contamination gradient. The hierarchical clustering of relative abundance proportions of functions was performed using Ward's method and Bray–Curtis dissimilarity distance. The ORF approach was used with identity threshold of 60%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary Figure S11 | Specific variation of functions cross-metagenomes. Two heatmaps represent specific function abundance FP3-OPA-nc (167 functions) and FP3 specific to pollution gradient (225 functions). The hierarchical clustering of relative abundance proportions of functions was performed using Ward's method and Bray–Curtis dissimilarity distance. The ORF approach was used with identity threshold of 60%, *e*-value of 10–12 and minimum alignment length of 50 base pairs parameters. Vegan package and heatmap () function in R were used to produce this figure.

Supplementary File S1 | This file contains two tables. The first table resumed the geographical coordinates of sampled sites. The second table presented abiotic parameters measured for each sampled site 1 year before this study (Laplante and Derome, 2011).

Supplementary File S2 | This file summarized statistics of reads, contigs, and ORFs MG-RAST annotations per lake metagenome.

Supplementary File S3 | This file summarized in one table the relative abundance of major taxa at phylum, class, and genus levels.

Supplementary File S4 | This file resumed details of NMDS plots. NMDS loadings, abundance clusters and points labels of all genus (Table 1), *Proteobacteria* (Table 2), *Actinobacteria* (Table 3), *Cyanobacteria* (Table 4), then of all subsystems (Table 5) were reported for assuming a better understanding of NMDS figures and Supplementary Figures.

Supplementary File S5 | This file reported subsystems annotations and data analysis; All subsystems data (dataset output of STAMP software) in Table 1, subsystems level 1 relative abundance (RA) in Table 2, list of all annotated subsystems in Table 3, function profiles classification of RA proportions in Table 4, different function profiles (FP) (Tables 5, 6, 7, 8), resume of FP occurrence in subsystems level 1 (Table 9), summary of most abundant functions (Table 10) and summary of the subsystem "*virulence, disease, and defense*" (Table 11).

Supplementary File S6 | This file summarized details of NMDS correlation analysis with metadata. Each table resumed NMDS loadings (NMDS1, NMDS2), *P*-value of correlated trace metals of taxa (Table 1), subsystems level 1 (Table 2) and PMRGs.

Supplementary File S7 | This file resumed diversity measures based relative abundance of taxa (genus) and function (subsystems) in Table 1 and all annotated enzymes by their EC-number in Table 2.

REFERENCES

Allen, E. E., and Banfield, J. F. (2005). Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* 3, 489–498. doi: 10.1038/nrmicro1157

Allison, S. D., and Martiny, J. B. H. (2008). Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11512–11519. doi: 10.1073/pnas.0801925105

Almeida, W. I., Vieira, R. P., Cardoso, A. M., Silveira, C. B., Costa, R. G., Gonzalez, A. M., et al. (2008). Archaeal and bacterial communities of heavy metal contaminated acidic waters from zinc mine residues in Sepetiba Bay. *Extremophiles* 13, 263–271. doi: 10.1007/s00792-008-0214-2

Alpay, S. (2016). "Multidisciplinary Environmental Science Investigations Surrounding the Former Aldermac Mine," in *The Lac Dasserat Study Workshop Summarized* (No. 7993) (Abitibi, QC)

Álvarez, A., Catalano, S. A., and Amoroso, M. J. (2013). Heavy metal resistant strains are widespread along Streptomyces phylogeny. *Mol. Phylogenet. Evol.* 66, 1083–1088. doi: 10.1016/j.ympev.2012.11.025

Banci, L., Bertini, I., Ciofi-Baffoni, S., Gonnelli, L., and Su, X.-C. (2003). Structural basis for the function of the N-terminal domain of the ATPase CopA from Bacillus subtilis. *J. Biol. Chem.* 278, 50506–50513. doi: 10.1074/jbc.M307389200

Barberán, A., Fernández-Guerra, A., Bohannan, B. J. M., and Casamayor, E. O. (2012). Exploration of community traits as ecological markers in microbial metagenomes. *Mol. Ecol.* 21, 1909–1917. doi: 10.1111/j.1365-294X.2011.05383.x

Behlau, F., Canteros, B. I., Minsavage, G. V., Jones, J. B., and Graham, J. H. (2011). Molecular characterization of copper resistance genes from *Xanthomonas citri* subsp. citri and Xanthomonas alfalfae subsp. citrumelonis? *Appl. Environ. Microbiol.* 77, 4089–4096. doi: 10.1128/AEM.03043-10

Bissett, A., Brown, M. V., Siciliano, S. D., and Thrall, P. H. (2013). Microbial community responses to anthropogenically induced environmental change: towards a systems approach. *Ecol. Lett.* 16, 128–139. doi: 10.1111/ele.12109

Bittar, T. B., Stubbins, A., Vieira, A. A. H., and Mopper, K. (2015). Characterization and photodegradation of dissolved organic matter (DOM) from a tropical lake and its dominant primary producer, the cyanobacteria Microcystis aeruginosa. *Mar. Chem.* 177, 205–217. doi: 10.1016/j.marchem.2015.06.016

Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., and Corbeil, J. (2012). Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biol.* 13:R122. doi: 10.1186/gb-2012-13-12-r122

Bouvier, T. C., and del Giorgio, P. A. (2002). Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol. Oceanogr.* 47, 453–470. doi: 10.4319/lo.2002.47.2.0453

Bowen, J. L., Ward, B. B., Morrison, H. G., Hobbie, J. E., Valiela, I., Deegan, L. A., et al. (2011). Microbial community composition in sediments resists perturbation by nutrient enrichment. *ISME J.* 5, 1540–1548. doi: 10.1038/ismej.2011.22

Bryant, J. A., Stewart, F. J., Eppley, J. M., and DeLong, E. F. (2012). Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone. *Ecology* 93, 1659–1673. doi: 10.1890/11-1204.1

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011). Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14288–14293. doi: 10.1073/pnas.1101591108

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Cha, J. S., and Cooksey, D. A. (1991). Copper resistance in Pseudomonas syringae mediated by periplasmic and outer membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* 88, 8915–8919. doi: 10.1073/pnas.88.20.8915

Chen, L., Huang, L., Méndez-García, C., Kuang, J., Hua, Z., Liu, J., et al. (2016). Microbial communities, processes and functions in acid mine drainage ecosystems. *Curr. Opin. Biotechnol.* 38, 150–158. doi: 10.1016/j.copbio.2016.01.013

Debroas, D., Humbert, J.-F., Enault, F., Bronner, G., Faubladier, M., and Cornillot, E. (2009). Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget - France). *Environ. Microbiol.* 11, 2412–2424. doi: 10.1111/j.1462-2920.2009.01969.x

Desoeuvre, A., Casiot, C., and Héry, M. (2016). Diversity and distribution of arsenic-related genes along a pollution gradient in a river affected by acid mine drainage. *Microb. Ecol.* 71, 672–685. doi: 10.1007/s00248-015-0710-8

Devarajan, N., Laffite, A., Graham, N. D., Meijer, M., Prabakar, K., Mubedi, J. I., et al. (2015). Accumulation of clinically relevant antibiotic-resistance genes, bacterial load, and metals in freshwater lake sediments in Central Europe. *Environ. Sci. Technol.* 49, 6528–6537. doi: 10.1021/acs.est.5b01031

Dimkpa, C. O., Merten, D., Svatoš, A., Büchel, G., and Kothe, E. (2009). Siderophores mediate reduced and increased uptake of cadmium by Streptomyces tendae F4 and sunflower (*Helianthus annuus*), respectively. *J. Appl. Microbiol.* 107, 1687–1696. doi: 10.1111/j.1365-2672.2009.04355.x

Doolittle, W. F., and Zhaxybayeva, O. (2009). On the origin of prokaryotic species. *Genome Res.* 19, 744–756. doi: 10.1101/gr.086645.108

Edwards, K. J., and Bazylinski, D. A. (2008). Intracellular minerals and metal deposits in prokaryotes. *Geobiology* 6, 309–317. doi: 10.1111/j.1472-4669.2008.00156.x

Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *Science* 320, 1034–1039. doi: 10.1126/science.1153213

Fierer, N., Lauber, C. L., Ramirez, K. S., Zaneveld, J., Bradford, M. A., and Knight, R. (2012a). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J.* 6, 1007–1017. doi: 10.1038/ismej.2011.159

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et al. (2012b). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21390–21395. doi: 10.1073/pnas.1215210110

Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., et al. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature* 509, 612–616. doi: 10.1038/nature13377

Garcia-Etxebarria, K., Garcia-Garcer,à, M., and Calafell, F. (2014). Consistency of metagenomic assignment programs in simulated and real data. *BMC Bioinform.* 15:90. doi: 10.1186/1471-2105-15-90

Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D., and Meyer, F. (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* 2010:pdb.prot5368. doi: 10.1101/pdb.prot5368

Goldfarb, K. C., Karaoz, U., Hanson, C. A., Santee, C. A., Bradford, M. A., Treseder, K. K., et al. (2011). Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Terr. Microbiol.* 2, 94. doi: 10.3389/fmicb.2011.00094

González, I., Déjean, S., Martin, P. G. P., and Baccin, A. (2008). *CCA: An R Package to Extend Canonical Correlation Analysis. J. Stat. Softw.* 23, 1–14. doi: 10.18637/jss.v023.i12

Gravel, D., Bell, T., Barbera, C., Bouvier, T., Pommier, T., Venail, P., et al. (2011). Experimental niche evolution alters the strength of the diversity-productivity relationship. *Nature* 469, 89–92. doi: 10.1038/nature09592

Green, J. L., Bohannan, B. J. M., and Whitaker, R. J. (2008). Microbial biogeography: from taxonomy to traits. *Science* 320, 1039–1043. doi: 10.1126/science.1153475

Gupta, R. S., and Lorenzini, E. (2007). Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the Bacteroidetes and Chlorobi species. *BMC Evol. Biol.* 7:71. doi: 10.1186/1471-2148-7-71

Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., et al. (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J.* 4, 660–672. doi: 10.1038/ismej.2009.154

Hemme, C. L., Green, S. J., Rishishwar, L., Prakash, O., Pettenato, A., Chakraborty, R., et al. (2016). Lateral gene transfer in a heavy metal-contaminated-groundwater microbial community. *mBio* 7, e02234–e02215. doi: 10.1128/mBio.02234-15

Hooper, S. D., Mavromatis, K., and Kyrpides, N. C. (2009). Microbial co-habitation and lateral gene transfer: what transposases can tell us. *Genome Biol.* 10:R45. doi: 10.1186/gb-2009-10-4-r45

Hooper, S. D., Raes, J., Foerstner, K. U., Harrington, E. D., Dalevi, D., and Bork, P. (2008). A molecular study of microbe transfer between distant environments. *PLoS ONE* 3:e2607. doi: 10.1371/journal.pone.0002607

Huang, L.-N., Kuang, J.-L., and Shu, W.-S. (2016). Microbial ecology and evolution in the acid mine drainage model system. *Trends Microbiol.* 24, 581–593. doi: 10.1016/j.tim.2016.03.004

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Jackson, T. A., Vlaar, S., Nguyen, N., Leppard, G. G., and Finan, T. M. (2015). Effects of bioavailable heavy metal species, arsenic, and acid drainage from mine tailings on a microbial community sampled along a pollution gradient in a freshwater ecosystem. *Geomicrobiol. J.* 32, 724–750. doi: 10.1080/01490451.2014.969412

Jaspers, E., and Overmann, J. (2004). Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl. Environ. Microbiol.* 70, 4831–4839. doi: 10.1128/AEM.70.8.4831-4839.2004

Jie, S., Li, M., Gan, M., Zhu, J., Yin, H., and Liu, X. (2016). Microbial functional genes enriched in the Xiangjiang River sediments with heavy metal contamination. *BMC Microbiol.* 16:179. doi: 10.1186/s12866-016-0800-x

Johnson, D. B., and Hallberg, K. B. (2003). The microbiology of acidic mine waters. *Res. Microbiol.* 154, 466–473. doi: 10.1016/S0923-2508(03)00114-1

Kuang, J., Huang, L., He, Z., Chen, L., Hua, Z., Jia, P., et al. (2016). Predicting taxonomic and functional structure of microbial communities in acid mine drainage. *ISME J.* 10, 1527–1539 doi: 10.1038/ismej.2015.201

Kuang, J.-L., Huang, L.-N., Chen, L.-X., Hua, Z.-S., Li, S.-J., Hu, M., et al. (2013). Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J.* 7, 1038–1050. doi: 10.1038/ismej.2012.139

Langenheder, S., Lindström, E. S., and Tranvik, L. J. (2005). Weak coupling between community composition and functioning of aquatic bacteria. *Limnol. Oceanogr.* 50, 957–967. doi: 10.4319/lo.2005.50.3.0957

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Laplante, K., and Derome, N. (2011). Parallel changes in the taxonomical structure of bacterial communities exposed to a similar environmental disturbance: parallel changes in the structure of bacterial communities. *Ecol. Evol.* 1, 489–501. doi: 10.1002/ece3.37

Laplante, K., Boutin, S., and Derome, N. (2013). Parallel changes of taxonomic interaction networks in lacustrine bacterial communities induced by a polymetallic perturbation. *Evol. Appl.* 6, 643–659. doi: 10.1111/eva.12050

Larkin, A. A., and Martiny, A. C. (2017). Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ. Microbiol. Rep.* 9, 55–70. doi: 10.1111/1758-2229.12523

Li, A.-D., Li, L.-G., and Zhang, T. (2015). Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. *Front. Microbiol.* 6:1025. doi: 10.3389/fmicb.2015.01025

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015). Ocean plankton. Determinants of community structure in the global plankton interactome. *Science* 348:1262073. doi: 10.1126/science.1262073

Lind, P. A., Tobin, C., Berg, O. G., Kurland, C. G., and Andersson, D. I. (2010). Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol. Microbiol.* 75, 1078–1089. doi: 10.1111/j.1365-2958.2009.07030.x

Löhr, A. J., Laverman, A. M., Braster, M., Straalen, N. M., and van, Röling, W. F. M. (2006). Microbial Communities in the World's Largest Acidic Volcanic Lake, Kawah Ijen in Indonesia, and in the Banyupahit River Originating from It. *Microb. Ecol.* 52, 609–618. doi: 10.1007/s00248-006-9068-2

Louca, S., Jacques, S. M. S., Pires, A. P. F., Leal, J. S., Srivastava, D. S., Parfrey, L. W., et al. (2016a). High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* 1:0015. doi: 10.1038/s41559-016-0015

Louca, S., Parfrey, L. W., and Doebeli, M. (2016b). Decoupling function and taxonomy in the global ocean microbiome. *Science* 353, 1272–1277. doi: 10.1126/science.aaf4507

Ludwig, M., Chua, T. T., Chew, C. Y., and Bryant, D. A. (2015). Fur-type transcriptional repressors and metal homeostasis in the cyanobacterium *Synechococcus* sp. PCC 7002. *Front Microbiol.* 6:1217.doi: 10.3389/fmicb.2015.01217

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinforma. Oxf. Engl.* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Maharjan, R., Seeto, S., Notley-McRobb, L., and Ferenci, T. (2006). Clonal adaptive radiation in a constant environment. *Science* 313, 514–517. doi: 10.1126/science.1129865

Martiny, A. C., Treseder, K., and Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* 7, 830–838. doi: 10.1038/ismej.2012.160

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341

Mavromatis, K., Ivanova, N., Barry, K., Shapiro, H., Goltsman, E., McHardy, A. C., et al. (2007). Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat. Methods* 4, 495–500. doi: 10.1038/nmeth1043

Mayali, X., Weber, P. K., Mabery, S., and Pett-Ridge, J. (2014). Phylogenetic patterns in the microbial response to resource availability: amino acid incorporation in San Francisco Bay. *PLoS ONE* 9:e95842. doi: 10.1371/journal.pone.0095842

Miki, T., Yokokawa, T., and Matsui, K. (2014). Biodiversity and multifunctionality in a microbial community: a novel theoretical approach to quantify functional redundancy. *Proc. R. Soc. B* 281:20132498. doi: 10.1098/rspb.2013.2498

Morrissey, E. M., Mau, R. L., Schwartz, E., Caporaso, J. G., Dijkstra, P., van Gestel, N., et al. (2016). Phylogenetic organization of bacterial activity. *ISME J.* 10, 2336–2340. doi: 10.1038/ismej.2016.28

Mota, R., Pereira, S. B., Meazzini, M., Fernandes, R., Santos, A., Evans, C. A., et al. (2015). Effects of heavy metals on Cyanothece sp. CCY 0110 growth, extracellular polymeric substances (EPS) production, ultrastructure and protein profiles. *J. Proteomics* 120, 75–94. doi: 10.1016/j.jprot.2015.03.004

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., et al. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332, 970–974. doi: 10.1126/science.1198719

Navarro, C. A., von Bernath, D., and Jerez, C. A. (2013). Heavy metal resistance strategies of acidophilic bacteria and their acquisition: importance for biomining and bioremediation. *Biol Res* 46, 363–371. doi: 10.4067/S0716-97602013000400008

Ni, C., Horton, D. J., Rui, J., Henson, M. W., Jiang, Y., Huang, X., et al. (2016). High concentrations of bioavailable heavy metals impact freshwater sediment microbial communities. *Ann. Microbiol.* 66, 1003–1012. doi: 10.1007/s13213-015-1189-8

Ofiteru, I. D., Lunn, M., Curtis, T. P., Wells, G. F., Criddle, C. S., Francis, C. A., et al. (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15345–15350. doi: 10.1073/pnas.1000604107

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., and O'Hara, R. (2016). *vegan: Community Ecology Package*. R Package 2.3-3. doi: 10.4135/9781412971874.n145. Available online at: https://CRAN.R-project.org/package=vegan

Outten, F. W., Huffman, D. L., Hale, J. A., and O'Halloran, T. V. (2001). The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. *J. Biol. Chem.* 276, 30670–30677. doi: 10.1074/jbc.M104122200

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. G. J. (2014). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, D737–D743. doi: 10.1093/nar/gkt1252

Parks, D. H., and Beiko, R. G. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* 26, 715–721. doi: 10.1093/bioinformatics/btq041

Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinforma. Oxf. Engl.* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494

Parnell, J. J., Crowl, T. A., Weimer, B. C., and Pfrender, M. E. (2009). Biodiversity in microbial communities: system scale patterns and

mechanisms. *Mol. Ecol.* 18, 1455–1462. doi: 10.1111/j.1365-294X.2009.04128.x

Pérez, M. T., Rofner, C., and Sommaruga, R. (2015). Dissolved organic monomer partitioning among bacterial groups in two oligotrophic lakes. *Environ. Microbiol. Rep.* 7, 265–272. doi: 10.1111/1758-2229.12240

Pernthaler, J. (2013). "Freshwater microbial communities," in *The Prokaryotes,* eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin; Heidelberg: Springer Berlin Heidelberg), 97–112.

Philippot, L., Andersson, S. G. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., et al. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat. Rev. Microbiol.* 8, 523–529. doi: 10.1038/nrmicro2367

Pokrovsky, O. S., Martinez, R. E., Golubev, S. V., Kompantseva, E. I., and Shirokova, L. S. (2008). Adsorption of metals and protons on Gloeocapsa sp. cyanobacteria: a surface speciation approach. *Appl. Geochem.* 23, 2574–2588. doi: 10.1016/j.apgeochem.2008.05.007

R Foundation for Statistical Computing (2008). *R: A Language and Environment for Statistical Computing*. Vienna

Raes, J., Letunic, I., Yamada, T., Jensen, L. J., and Bork, P. (2011). Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol. Syst. Biol.* 7:473. doi: 10.1038/msb.2011.6

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., et al. (2005). Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920. doi: 10.1126/science. 1109070

Reeve, W. G., Tiwari, R. P., Kale, N. B., Dilworth, M. J., and Glenn, A. R. (2002). ActP controls copper homeostasis in Rhizobium leguminosarum bv. viciae and *Sinorhizobium meliloti* preventing low pH-induced copper toxicity. *Mol. Microbiol.* 43, 981–991. doi: 10.1046/j.1365-2958.2002.02791.x

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38, e191–e191. doi: 10.1093/nar/gkq747

Robinson, C. J., Bohannan, B. J. M., and Young, V. B. (2010). From structure to function: the ecology of host-associated microbial communities. *Microbiol. Mol. Biol. Rev.* 74, 453–476. doi: 10.1128/MMBR.00014-10

Rohart, F., Gautier, B., Singh, A., and Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Comput. Biol.* 13:e1005752. doi: 10.1371/journal.pcbi.1005752

Salcher, M. M. (2013). Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *J. Limnol.* 73, 74–87. doi: 10.4081/jlimnol.2014.813

Sánchez-Andrea, I., Rodríguez, N., Amils, R., and Sanz, J. L. (2011). Microbial diversity in anaerobic sediments at Río Tinto, a naturally acidic environment with a high heavy metal content. *Appl. Environ. Microbiol.* 77, 6085–6093. doi: 10.1128/AEM.00654-11

Sandifer, P. A., and Sutton-Grier, A. E. (2014). Connecting stressors, ocean ecosystem services, and human health. *Nat. Resour. Forum* 38, 157–167. doi: 10.1111/1477-8947.12047

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Silva, G. G. Z., Green, K. T., Dutilh, B. E., and Edwards, R. A. (2015). SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* 32, 354–361. doi: 10.1093/bioinformatics/btv584.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244. doi: 10.1038/nature10571

Stankovic, S., Moric, I., Pavic, A., Vasiljevic, B., Johnson, B., and Cvetkovic, V. (2014). Investigation of the microbial diversity of an extremely acidic, metal-rich water body (Lake Robule, Bor, Serbia). *J. Serbian Chem. Soc.* 79, 729–741. doi: 10.2298/JSC130605071S

Stokes, H. W., and Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens: lateral gene transfer and antibiotic resistance evolution. *FEMS Microbiol. Rev.* 35, 790–819. doi: 10.1111/j.1574-6976.2011.00273.x

Streten-Joyce, C., Manning, J., Gibb, K. S., Neilan, B. A., and Parry, D. L. (2013). The chemical composition and bacteria communities in acid and metalliferous drainage from the wet–dry tropics are dependent on season. *Sci. Total Environ.* 443, 65–79. doi: 10.1016/j.scitotenv.2012.10.024

Vanwonterghem, I., Jensen, P. D., Dennis, P. G., Hugenholtz, P., Rabaey, K., and Tyson, G. W. (2014). Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J.* 8, 2015–2028. doi: 10.1038/ismej.2014.50

Volant, A., Bruneel, O., Desoeuvre, A., Héry, M., Casiot, C., Bru, N., et al. (2014). Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage. *FEMS Microbiol. Ecol.* 90, 247–263. doi: 10.1111/1574-6941.12394

Wang, K., Zhang, D., Xiong, J., Chen, X., Zheng, J., Hu, C., et al. (2015). Response of bacterioplankton communities to cadmium exposure in coastal water microcosms with high temporal variability. *Appl. Environ. Microbiol.* 81, 231–240. doi: 10.1128/AEM.02562-14

Ward, D. M., Bateson, M. M., Ferris, M. J., Kühl, M., Wieland, A., Koeppel, A., et al. (2006). Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1997–2008. doi: 10.1098/rstb.2006.1919

Webb, C. O., Ackerly, D. D., McPeek, M. A., and Donoghue, M. J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33, 475–505. doi: 10.1146/annurev.ecolsys.33.010802.150448

Wendt-Potthoff, K., and Koschorreck, M. (2002). Functional groups and activities of bacteria in a highly acidic volcanic mountain stream and lake in Patagonia, Argentina. *Microb. Ecol.* 43, 92–106. doi: 10.1007/s00248-001-1030-8

Wilke, A., Bischof, J., Harrison, T., Brettin, T., D'Souza, M., Gerlach, W., et al. (2015). A RESTful API for accessing microbial community data for MG-RAST. *PLoS Comput. Biol.* 11:e1004008. doi: 10.1371/journal.pcbi.1004008

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., and Bork, P. (2011). iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* 39, W412–W415. doi: 10.1093/nar/gkr313

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cheaib, Le Boulch, Mercier and Derome. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 November 2017
doi: 10.3389/fmicb.2017.02297

pH Stress-Induced Cooperation between *Rhodococcus ruber* YYL and *Bacillus cereus* MLY1 in Biodegradation of Tetrahydrofuran

Zubi Liu[1†], Zhixing He[2†], Hui Huang[1], Xuebin Ran[1], Adebanjo O. Oluwafunmilayo[1] and Zhenmei Lu[1]*

[1] College of Life Sciences, Zhejiang University, Hangzhou, China, [2] College of Basic Medical Science, Zhejiang Chinese Medical University, Hangzhou, China

Microbial consortia consisting of cooperational strains exhibit biodegradation performance superior to that of single microbial strains and improved remediation efficiency by relieving the environmental stress. Tetrahydrofuran (THF), a universal solvent widely used in chemical and pharmaceutical synthesis, significantly affects the environment. As a refractory pollutant, THF can be degraded by some microbial strains under suitable conditions. There are often a variety of stresses, especially pH stress, that inhibit the THF-degradation efficiency of microbial consortia. Therefore, it is necessary to study the molecular mechanisms of microbial cooperational degradation of THF. In this study, under conditions of low pH (initial pH = 7.0) stress, a synergistic promotion of the THF degradation capability of the strain *Rhodococcus ruber* YYL was found in the presence of a non-THF degrading strain *Bacillus cereus* MLY1. Metatranscriptome analysis revealed that the low pH stress induced the strain YYL to up-regulate the genes involved in anti-oxidation, mutation, steroid and bile acid metabolism, and translation, while simultaneously down-regulating the genes involved in ATP production. In the co-culture system, strain MLY1 provides fatty acids, ATP, and amino acids for strain YYL in response to low pH stress during THF degradation. In return, YYL shares the metabolic intermediates of THF with MLY1 as carbon sources. This study provides the preliminary mechanism to understand how microbial consortia improve the degradation efficiency of refractory furan pollutants under environmental stress conditions.

Keywords: metatranscriptome, *Rhodococcus ruber* YYL, *Bacillus cereus* MLY1, cooperation, tetrahydrofuran degradation, low pH stress

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Bernd Wemheuer, University of New South Wales, Australia
Svetlana Yurgel, Dalhousie University, Canada

***Correspondence:**
Zhenmei Lu
lzhenmei@zju.edu.cn

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *25 July 2017*
Accepted: *07 November 2017*
Published: *21 November 2017*

Citation:
Liu Z, He Z, Huang H, Ran X, Oluwafunmilayo AO and Lu Z (2017) pH Stress-Induced Cooperation between Rhodococcus ruber YYL and Bacillus cereus MLY1 in Biodegradation of Tetrahydrofuran. Front. Microbiol. 8:2297. doi: 10.3389/fmicb.2017.02297

INTRODUCTION

Bioremediation is an eco-friendly waste management technique that uses naturally-occurring organisms to break down hazardous substances into less toxic or non-toxic substances. Microbial bioremediation has been the most significant remediation method for the mineralization of pollutants in contaminated environments (Vilchez-Vargas et al., 2010; Krastanov et al., 2013). For this reason, many degrading microorganisms have been isolated from the polluted environments and tested for their degradation potential and mechanisms. However, degrading microorganisms do not often exist in mono-culture in the environment. Instead, many bacteria are found in close

association with one another and exhibit a superior biodegradation performance than that by a single species (Mikeskova et al., 2012). For example, compared to single bacteria, a consortium of *Pseudomonas* sp. SUK1 and *Aspergillus ochraceus* NCIM-1146 shows an increased ability to break down the textile dye reactive navy blue HE2R in 24 h (Kadam et al., 2011). The biochemical cooperation of different microorganisms accelerates the complete degradation of some pollutants (Wu et al., 2010). In addition, some bacteria may display no degradation ability but protect the bacteria with degradation ability or change the structure of the contaminants to enhance the degradation process (Di Gioia et al., 2004; Zhao and Wong, 2009). Therefore, interactions between microbes play significant roles in hazardous substance degradation.

Tetrahydrofuran (THF) is a heterocyclic ether with the formula $(CH_2)_4O$. As a polar, versatile solvent, THF is widely used as a reaction media for the synthesis of polymers. Additionally, THF is easily detected in the groundwater (Isaacson et al., 2006) and can penetrate the skin, cause rapid dehydration, inhibit cytochrome P450, and induce central nervous system irritation, narcosis, edema, and colonic muscle spasms in animals (Chhabra et al., 1990; Malley et al., 2001). As an eco-friendly and cost-effective strategy for pollutant removal, the application of THF-degrading microorganisms has been rarely investigated. To date, only a few strains of the genera *Rhodococcus* (Daye et al., 2003; Yao et al., 2009; Tajima et al., 2012), *Pseudonocardia* (Kohlweyer et al., 2000; Masuda et al., 2012), and *Pseudomonas* (Chen et al., 2010) have been reported to bear the ability to use THF as the sole carbon source. In our previous studies, we isolated a THF-degrading strain, named *Rhodococcus ruber* YYL, which had a maximum THF degradation rate of 137.60 $mg{\cdot}h^{-1}{\cdot}g^{-1}$ YYL dry weight (Yao et al., 2009, 2013). In addition, the symbiotic non-THF-degrading *Bacillus cereus* MLY1 was isolated with strain YYL from the activated sludge and could successfully augment the strain YYL colonization in activated sludge and remarkably improve the THF removal in the reactor (Yao et al., 2013). However, the mechanism of the non-THF-degrading microorganism's survival in the mineral medium and the interactions with strain YYL remain unclear.

The environmental pH has been one of the key factors affecting the microbial degradation of pollutants (Lowe et al., 1993; Hall-Stoodley et al., 2004; Huertas et al., 2010). Strain YYL shows its optimal degradation efficiency at initial pH 8.3 (Yao et al., 2009). However, actual environment conditions cannot meet this optimal pH. Strain YYL produces acidic substances continuously during the degradation of THF and decreases the pH of its environment, negatively affecting its degradation efficiency. We found that the THF degradation efficiency of YYL could be improved when under low pH stress (initial pH = 7.0) by the addition of the strain MLY1.

To study the interactions between strains YYL and MLY1, these two strains were co-cultured under the initial pH of 7.0 or 8.3 and subjected to metatranscriptome analysis. Here we focus on the difference in THF removal and gene expression in strain YYL and MLY1 between the co-culture and mono-culture systems with different pH conditions.

MATERIALS AND METHODS

Strains, Culture Conditions, and Co-culture Experiments

The THF-degrading strain *R. ruber* YYL was cultured and maintained in 100 mL liquid optimal base mineral medium (BMM) at initial pH of 8.3 with 20 mM THF (Yao et al., 2009); meanwhile, *B. cereus* MLY1 was also cultured and maintained in 100 mL BMM with 1.0 g/L yeast extract. For the initial phase of each co-culture experiment, the stationary phase cells of strain YYL or MLY1 were centrifuged (5 min, 8000 rpm), the growth media was decanted, and the cells were suspended in BMM at the initial pH of 8.3 or 7.0. Subsequently, 100 mL BMM (initial pH = 8.3 or 7.0) with 20 mM THF was inoculated with 2 mL of strain YYL (OD_{600} = 1.5) and 1 mL of MLY1 (OD_{600} = 1.5) to make a cell ratio of 2:1. The corresponding mono-culture systems were incubated with the same inoculum of YYL and 1 mL of BMM. One Erlenmeyer flask of culture was used for each sample, and four replicates were prepared for every sampling time point in the four treatment groups. All the flasks were cultured at 140 rpm and a temperature of 30°C.

Sample Collection and Detection

Samples were collected for the determination of the THF concentration and the pH during the incubation. THF concentrations were measured by GC-2014C gas chromatography equipped with a flame ionization detector (FID) and an AOC-20i auto injector (SHIMADZU, Japan). The injector, oven, and detector temperatures were set to 200, 160, and 200°C, respectively. The THF peak was observed at a retention time of 1.7 min. For the transcriptional analysis, the sampling time for the experiment was chosen based on the THF removal efficiency and strain growth stage in the four treatment groups. In this study, samples were collected at the initial stage, the first 4 days, during which the cells of all the treatments were at the exponential growth phase. Also, the transcriptional changes in strain YYL that could describe its response within the symbiotic system were identified. All the collected samples were immediately centrifuged at 4°C, flash-frozen in liquid nitrogen, and stored at −80°C. To detect strain MLY1 growth, the ratio of the DNA levels of the specific genes *thm/GerM* were quantified using quantitative PCR (qPCR) to detect the ratio of strain YYL to strain MLY1 in co-culture (Rosenthal et al., 2011; Benomar et al., 2015); *thm* is the THF monooxygenase gene responsible for the first step of THF degradation as described in previous studies (He et al., 2014), and the *GerM* gene encodes a lipoprotein that stabilizes the GerA-GerQ complex through an interaction with the remodeled cell wall during spore formation in bacilli (Rodrigues et al., 2016). The primers for the qPCR are shown in Supplementary Table S1.

RNA Isolations, Library Construction, and Sequencing

Total RNA was extracted from four biological replicates per growth condition using the RNeasy Mini Kit (Cat#74106, QIAGEN, Germany) according to the manufacturer's

instructions. The RNA quality and yield were measured with gel electrophoresis using 1% agarose gel and NanoDrop 2000 Spectrophotometer (Thermo Scientific, Wilmington, DE, United States). Subsequently, the extracted quadripartite RNA was pooled equally and purified with the RNeasy Micro Kit (Cat#74004, QIAGEN, Germany) and the RNase-Free DNase Set (Cat#79254, QIAGEN, Germany) (Zhang and Gant, 2005; Kainkaryam et al., 2010; Devi et al., 2016; Zhao et al., 2017). The success of DNA removal was confirmed by PCR of the 16S rRNA gene with primers 27F and 1492R (DeLong, 1992) (Supplementary Table S1). The reaction contained 5 μL of the 10x LA Taq buffer (Mg^{+2}), 4 μL of dNTPs (2.5 mM), 1 μL of primer 27F (10 μM), 1 μL of primer 1492R (10 μM), 1 μL of the template DNA/RNA, 0.5 μL of the LA Taq polymerase, and 37.5 μL of ddH_2O. The PCR conditions used were 94°C for 30 s and 28 cycles of 55°C for 30 s and 72°C for 90 s. Again, the integrity and quantity were assessed using NanoDrop 2000 Spectrophotometer and Agilent Bioanalyzer 2100 (Agilent Technologies, United States). The depletion of ribosomal RNA before cDNA synthesis was performed using the Ribo-Zero kit (Epicentre Biotechnologies, United States) for meta-bacteria. The mRNA was fragmented using the Ambion RNA fragmentation kit (Ambion, United States), and double-stranded cDNA was generated using the Qubit dsDNA HS Assay Kit (Invitrogen, United States). Subsequently, four mRNA libraries were constructed, evaluated using the Agilent Bioanalyzer 2100, and sequenced using 2 × 125 paired-end reads on an Illumina HiSeq2500 sequencer using the HiSeq SBS Kit v4 (Illumina, United States). Approximately 5 Gbp of clean data was targeted for each library. There were 57,246,312; 38,944,960; 43,907,548; and 47,725,396 reads of total clean data obtained, and 57,189,768; 38,862,480; 43,867,792; and 47,697,494 reads of clean data were left after rRNA removal for the mono- and co-culture performed with initial pH of 7.0 and 8.3.

RNA-Seq Data Analysis

Raw reads were filtered to remove low quality reads using Seqtk[1] with the following standards: (1) reads with adaptors were removed; (2) reads containing more than 50 bases with low quality (Q20) were removed; (3) reads with more than 3 N bases were removed; (4) low quality bases or N bases assigned at the 3′ tail were removed; and (5) reads shorter than 20 bp were removed. To eliminate all ribosomal RNA sequences, reads mapping to the rRNA (5S rRNA, 16S rRNA, and 23S rRNA) of YYL and MLY1 were removed, and the remaining clean reads were used for the subsequent analysis.

All the clean reads from the mono- and co-culture systems were pooled and assembled using the Trinity assembly algorithm for Primary UniGene and CAP3 EST for Final UniGene (First_contig and Second_contig) (Huang and Madan, 1999; Grabherr et al., 2011). The acquired Final UniGene sequences were searched against the NCBI non-redundant (nr) database (March 2014) using BLASTx (version BLAST-2.2.28+) (Altschul et al., 1990) for the detection of protein-coding genes with the parameter *e*-value $< 10^{-30}$, identity >60%, and genes were identified using self-developed Perl scripts; the maximum overlap against adjacent genes was 100 bp. Subsequently, further redundancy was removed using CD-HIT-EST v4.6 (Li and Godzik, 2006) with a sequence identity threshold of 99% in every 1000 bp (Wong et al., 2015). A collection of 9,483 non-redundant genes were used as a reference genome for the differential gene analysis.

Clean reads from each condition were aligned to the whole-body transcriptome of YYL. The normalized output for each gene expression was calculated as Reads Per Kilobase per Million mapped reads (RPKM) (Mortazavi et al., 2008). For Differentially Expressed Gene (DEG) sets, hierarchical clustering analysis was performed using the complete linkage and Euclidean distance as a measure of similarity. All data analyses and visualization of DEGs were conducted using R3.0.2[2]. Gene expression under different pH conditions in the mono- and co-culture systems were compared, and the differential expression was considered as a fold-change ≥2 and false discovery rate (FDR) < 0.05.

All the gene sequences were searched against the Gene Ontology (GO) (*e*-value $< 10^{-5}$), Kyoto Encyclopedia of Genes and Genomes (KEGG), and evolutionary genealogy of genes Non-supervised Orthologous Groups (eggNOG) using BLASTx (version: BLAST-2.2.28^{+}) (Altschul et al., 1990; Ashburner et al., 2000). To confirm which GO terms and metabolic pathways from the mono- and co-culture systems responded to the different pH conditions, GO and KEGG enrichment analyses were performed using DAVID (Huang et al., 2009). The significantly enriched GO terms and KEGG pathways in the DEGs were identified with the hypergeometric test in the entire genome background. The calculated *p*-value was corrected by the Bonferroni method with a threshold of *p*-value ≤ 0.05 to define the GO terms and KEGG pathways as significantly enriched. Subsequently, the major biological functions, the most important biochemical metabolic pathways, and signal transduction pathways with differentially expressed genes (DEGs) were identified by GO and pathway significance enrichment.

Sequencing data are available in the NCBI database under the accession numbers SRR5723771, SRR5723772, SRR5723773, and SRR5723774.

RESULTS

THF Degradation in the Incubation Systems

As previously mentioned, the pH is a significant variable affecting the THF degradation efficiency of strain YYL, and its optimal initial pH for THF degradation is 8.3 (Yao et al., 2009). The degradation efficiency of strain YYL at the sub-optimal initial pH of 7.0 can be improved by co-culture of YYL with MLY1. Therefore, strains YYL and MLY1 were mono- or co-cultured at initial pH of 7.0 and 8.3 and subjected to degrading physiological analysis (**Figure 1**). When the THF concentration and pH between the co-culture and mono-culture systems were compared, there was no significant difference found under the

[1]http://github.com/lh3/seqtk

[2]http://www.r-project.org

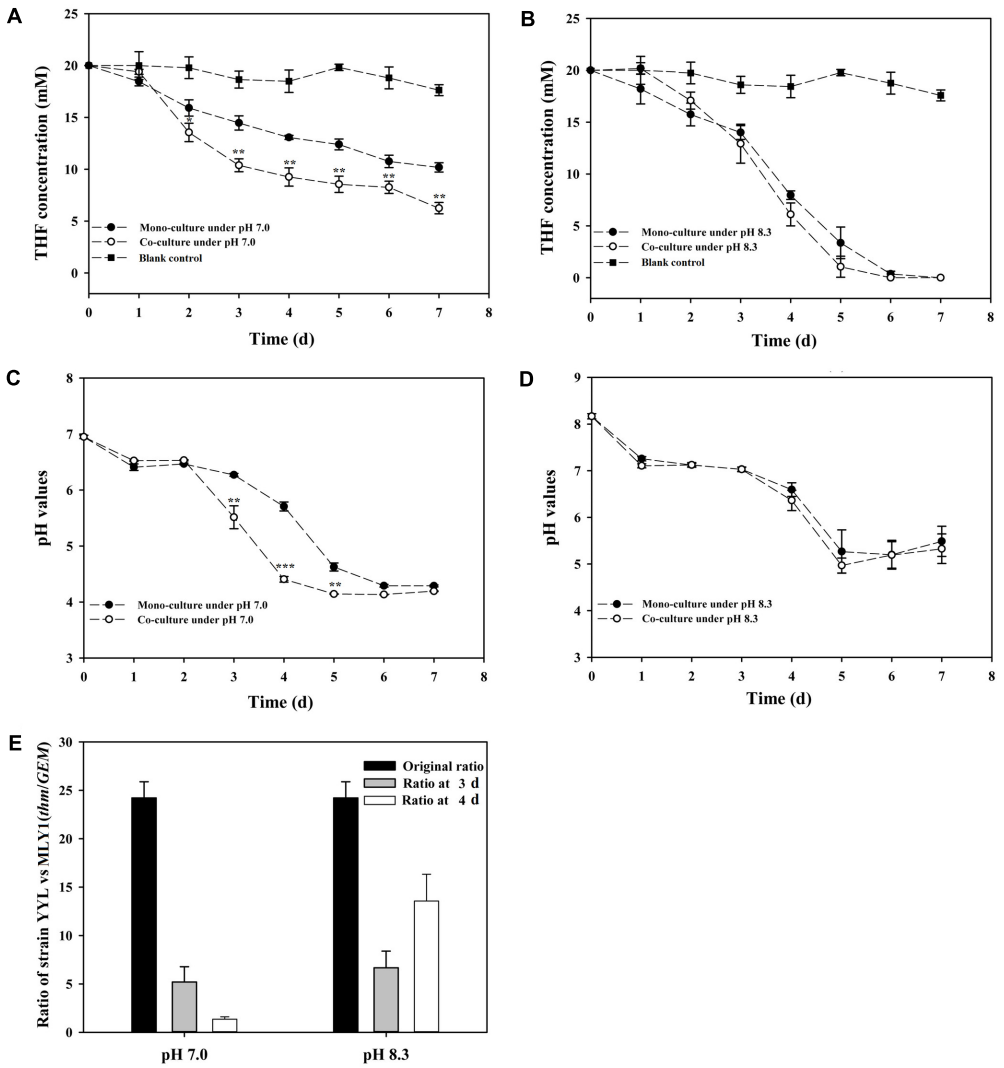


FIGURE 1 | Tetrahydrofuran (THF) degradation curves under the initial pH of 7.0 **(A)** and 8.3 **(B)**, pH change curves under the initial pH of 7.0 **(C)** and 8.3 **(D)**, ratio of strain YYL to strain MLY1 **(E)** in the mono-culture and co-culture systems under pH stress (initial pH = 7.0) and the optimal initial pH of 8.3. Significance was analyzed by the Student's *t*-test ($n = 4$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

initial pH of 8.3, but a lower THF concentration and pH were detected under the initial pH of 7.0 in the co-culture than in the mono-culture (**Figures 1A–D**). Meanwhile, the ratio of strain YYL/strain MLY1 was lower when under the initial pH of 7.0 than that observed under the initial pH of 8.3 (**Figure 1E**), suggesting that strain MLY1 plays a more significant role under low pH conditions. Additionally, a more obvious synergistic promotion of the THF degradation was detected between strain YYL and MLY1 with low pH stress (initial pH = 7.0).

Transcriptional Response of Strain YYL to pH Stress in Mono-culture

As discussed above, the low initial pH of 7.0 was unfavorable for the THF degradation by strain YYL. To explore the effects of pH on the gene expression of strain YYL, transcriptional differences in the mono-culture systems of strain YYL under the initial pH of 7.0 and 8.3 were compared. As shown in **Figure 2A**, clear differences were observed in the general patterns of the transcriptional responses to the two-different pH. Growing strain YYL under the initial pH of 7.0 led to both increases and decreases in gene expression when compared to growing at the initial pH of 8.3. To explore the functions of the DEGs from strain YYL, KEGG, and GO annotation were analyzed.

The KEGG enrichment analysis revealed that steroid degradation and biosynthesis, degradation of the aromatic compounds, xylene, and dioxin, RNA transport, spliceosome, non-homologous end-joining, alpha-linoleic acid metabolism, and bile acid biosynthesis were up-regulated when strain YYL was grown under the initial pH of 7.0 (**Figure 2B**). Also, the GO functional analysis showed that pH stress (initial pH = 7.0) induced up-regulation of genes involved in the transposase

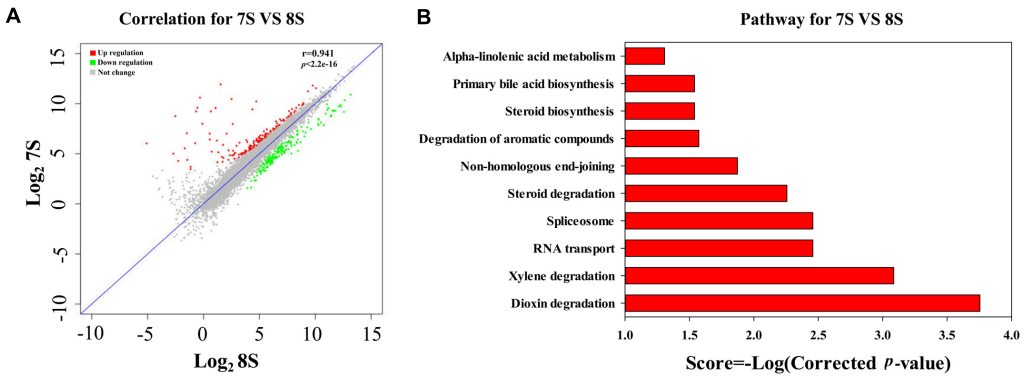


FIGURE 2 | Correlation plot **(A)** and metabolism pathway (KEGG) analyses **(B)** of differentially expressed genes (DEGs) from strain YYL under the initial pH of 7.0 and 8.3. 7S and 8S represent the mono-culture of strain YYL under the initial pH of 7.0 and 8.3, respectively.

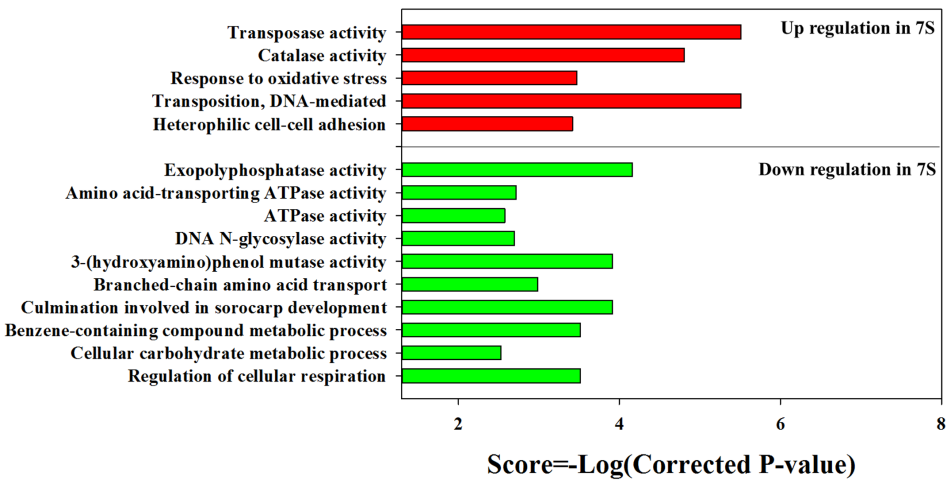


FIGURE 3 | Gene Ontology (GO) enrichment results for DEGs from strain YYL in 7S vs. 8S. 7S and 8S represent mono-culture under the initial pH of 7.0 and 8.3, respectively.

and catalase activity, response to oxidative stress, transposition, and heterothallic cell-cell adhesion, with down-regulation of the ATPase activity, exopolyphosphatase activity, DNA N-glycosylase activity, branched-chain amino acid transport, and cellular carbohydrate metabolic processes occurring (**Figure 3**).

In summary, strain YYL responded to the low initial pH (7.0) stress by enhancing steroid concentration, anti-oxidation, additional compound degradation, translation, and mutation activity. Meanwhile, the low initial pH (7.0) stress decreased the ATPase activity, carbohydrate metabolic activity, and cell respiration of strain YYL.

Transcriptional Response of Strain YYL to Co-culture Under Different pH Values

According to the THF degradation curves (**Figures 1A,B**), strain MLY1 exhibited a synergistic promotion of the THF degradation by strain YYL when grown with initial pH of 7.0 but had no such effect when grown with initial pH of 8.3. Therefore, we hypothesized that strain MLY1 would induce different effects on the gene expression of strain YYL in co-cultured systems grxown at a different pH. As shown in **Figures 4A,C**, the correlation of the gene expression from strain YYL between the mono-culture and co-culture indicates that strain MLY1 influences the gene expression of strain YYL.

The DEGs from strain YYL derived from comparing mono- and co-culture under the initial pH of 7.0 were also analyzed by KEGG and GO annotation. As shown in **Figures 4B,D**, significantly different KEGG pathways were up-regulated in mono-culture when grown with initial pH of 7.0. The depleted KEGG pathways from the co-culture system included RNA transport, primary bile acid biosynthesis, glycolysis, 2-oxocarboxylic acid metabolism, and amino acid biosynthesis. The GO functional analysis revealed that the depleted GO functions from the co-culture system included 113 terms, such as the fatty acid biosynthesis process, response to acid, amino acid biosynthesis process, ATPase activity, and malate dehydrogenase activity (Supplementary Figure S1A); the enriched GO functions from the co-culture system included 25

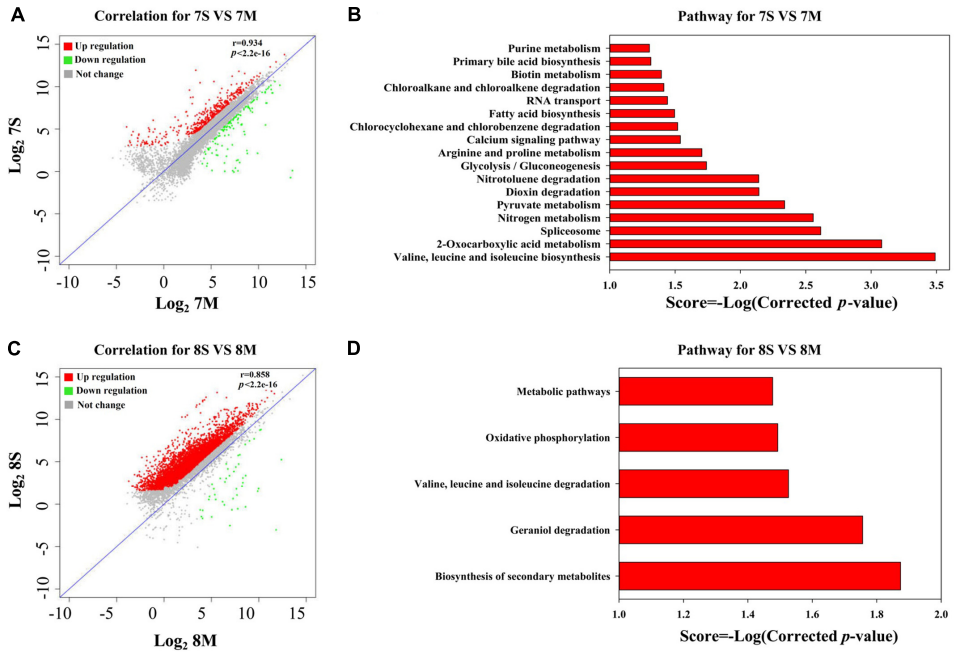


FIGURE 4 | Analysis of DEGs between mono-culture and co-culture in strain YYL. Correlation plot **(A)** and metabolism pathway (KEGG) analyses **(B)** under the initial pH of 7.0. Correlation plot **(C)** and metabolism pathway (KEGG) analyses **(D)** under the initial pH of 8.3. 7S and 7M represent the mono-culture and co-culture under the initial pH of 7.0, respectively. 8S and 8M represent the mono-culture and co-culture under the initial pH of 8.3, respectively.

terms, such as the preribosome, phosphoprotein phosphatase activity, and isopentenyl diphosphate biosynthetic process (Supplementary Figure S1B).

In the co-culture grown with initial pH of 8.3, the depleted KEGG pathways from strain YYL included metabolic pathways, oxidative phosphorylation, valine, leucine, and isoleucine degradation, geraniol degradation, and biosynthesis of secondary metabolites (**Figure 5D**). The GO functional analysis revealed that the depleted GO functions from the co-culture system included nine terms, such as RNA binding, proton-transporting ATP synthase activity, and growth (Supplementary Figure S2); the enriched GO function terms from the co-culture system included only the ribonucleo-protein complex and cytosolic large ribosomal subunit.

In summary, strain MLY1 imposed more effects on the gene expression of strain YYL when grown with initial pH of 7.0 than when grown with initial pH of 8.3. With initial pH of 7.0, strain YYL exhibited a lower level of amino acid biosynthesis, fatty acid biosynthesis, ATPase activity, response to acid, and translation activity because of the influence of strain MLY1; with initial pH of 8.3, strain YYL exhibited a lower level of the functions involved in cellular structure and biological function, such as cell wall synthesis, RNA binding, and cell growth.

Transcriptional Response to pH Stress in Co-culture

In the mono-culture system, pH stress (initial pH = 7.0) induced significant alterations in the gene expression of strain YYL. To explore the differences in gene expression between the two pH values in the co-culture systems, the transcriptomes of strain YYL and MLY1 were analyzed.

As shown in **Figure 5A**, most of the DEGs from strain YYL were up-regulated when YYL was grown with initial pH of 7.0 in the co-culture system. The KEGG enrichment analysis of DEGs indicated that oxidative phosphorylation was up-regulated in strain YYL because of the pH stress (initial pH = 7.0) (**Figure 5B**). The GO functional analysis showed the enrichment of GO function terms involved in the structural constituent of the ribosome, ATP synthesis, translation, and growth; the depleted GO function terms included rRNA binding, ribosome, ribonucleo-protein complex, and cytosolic large ribosomal subunit (Supplementary Figure S3).

As shown in **Figure 5C**, strain MLY1 in the co-culture system exhibited different gene expression when grown under the initial pH of 7.0 and 8.3. The KEGG analysis of differential genes showed the up-regulation of pathways from strain MLY1 grown with initial pH of 7.0 such as the biosynthesis of amino acids, 2-oxocarboxylic acid metabolism, sucrose metabolism, and ribosome (**Figure 5D**). The GO functional analysis revealed only two GO functional terms that were up-regulated in strain MLY1 grown with initial pH of 7.0 (Supplementary Figure S4).

In summary, the effects of low initial pH (7.0) stress on the gene expression of strain YYL were less notable in the co-culture system than in the mono-culture system. Meanwhile, strain MLY1 enhanced 2-oxocarboxylic acid metabolism and the synthesis of amino acids, ATP, and ribosomes to help strain YYL respond to the low initial pH (7.0) stress in the co-culture system.

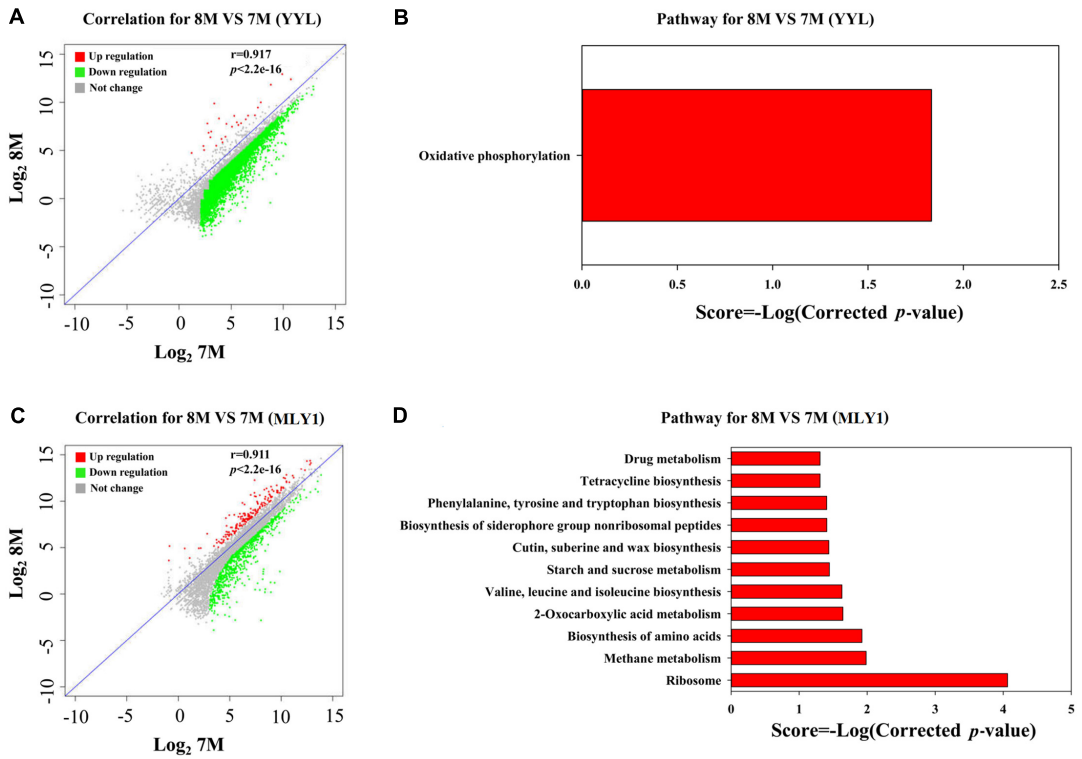


FIGURE 5 | Analysis of DEGs between the initial pH of 7.0 and 8.3 in the co-culture systems. Correlation plot **(A)** and metabolism pathway (KEGG) analyses **(B)** in strain YYL. Correlation plot **(C)** and metabolism pathway (KEGG) analyses **(D)** in strain MLY1. 7M and 8M represent co-culture under the initial pH of 7.0 and 8.3, respectively.

DISCUSSION

The response to pH stress plays a vital role in the survival of Gram-positive bacteria, and the mechanisms utilized by the gram-positive bacteria in response to pH stress operate in different ways (Cotter and Hill, 2003). However, the THF-degrading strain YYL exhibits weak resistance to low pH stress, with a low growth rate and THF degradation efficiency under low pH stress (Yao et al., 2009). In this study, two genes involved in anti-oxidation and mutation in strain YYL were unregulated when YYL was grown with initial pH of 7.0 suggesting that this pH renders an environmental stress on strain YYL. To deal with this stress, the synergistic relationships among microorganisms can be beneficial in the natural environment (Ma et al., 2016; Ren et al., 2016). Based on our analysis of the THF degradation efficiency, a synergistic relationship between strain YYL and the non-THF-degrading strain MLY1 exists when strain YYL experiences pH stress (initial pH = 7.0) but not when strain YYL is not stressed (initial pH = 8.3).

Based on these results, we propose a model for the cooperation between strains YYL and MLY1, shown in **Figure 6**. In co-culture system with THF as the sole carbon source, THF is degraded by YYL, and the easily usable intermediates are utilized by MLY1 as carbon sources (**Figure 6**). The unfavorable pH environment causes a significant change in the expression of genes involved in the metabolism of fatty acids required for lipid synthesis, ATP for energy, and amino acids for translation in strain YYL (**Figure 3** and Supplementary Figure S1A), which is lethal for YYL if the pH stress cannot be relieved. THF degradation is an acid-producing process, and the protons produced need be transported out of the cell, a process that is dependent upon ATP (Buch-Pedersen et al., 2009) (**Figure 6**). Fortunately, the symbiotic strain MLY1, surviving on the intermediates of the THF degradation, provides lipids, ATP, and amino acids for strain YYL to relieve the pH stress (**Figure 6**). Based on the above interactions between strain YYL and MLY1, the symbiotic system is stable to completely degrade the THF.

Strain MLY1 Could Contribute Fatty Acids and Lipids to Strain YYL

In response to pH stress (initial pH = 7.0), strain YYL up-regulates the genes involved in steroid biosynthesis and degradation and bile acid biosynthesis. Steroids and bile acids are both sterols, and interactive transformation exists between the steroids and bile acids in the genus *Rhodococcus* (Donova, 2007; Yam et al., 2011). Up-regulation of steroid metabolism plays a significant role in the response to environmental stress by *Rhodococcus* sp. (Larkin et al., 2006; Orro et al., 2015). This may be because of the fact that steroid metabolism is related to energy supply and membrane structure in microbial cells (Haussmann et al., 2013). In addition, previous literature has reported that

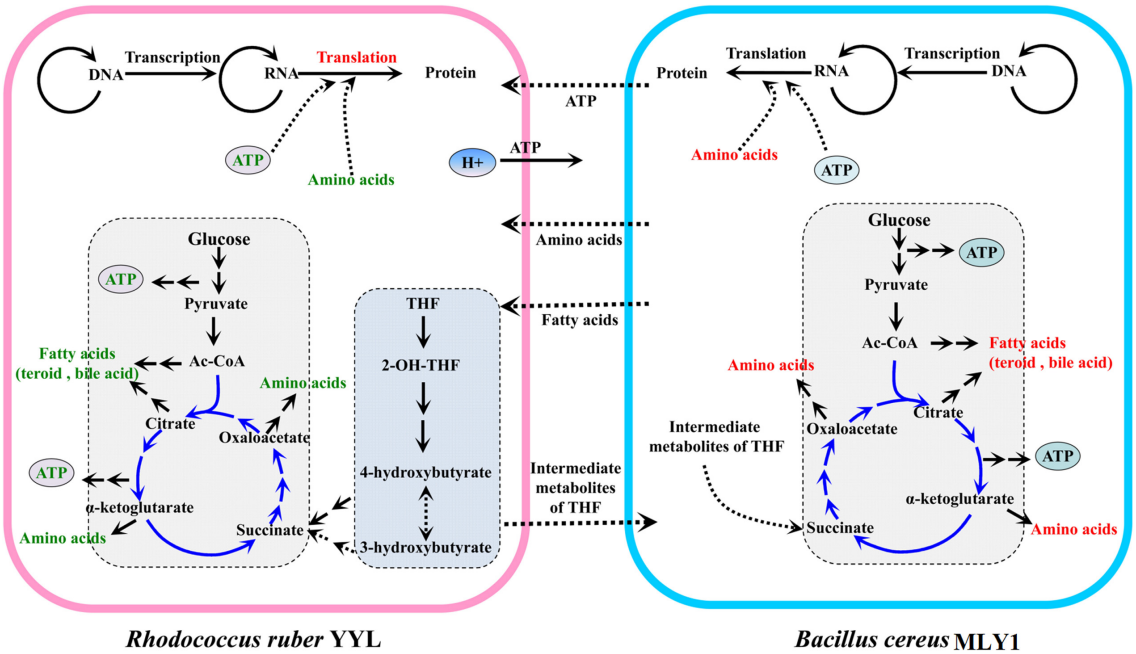


FIGURE 6 | The schematic mechanism of the interspecific interactions between strain YYL and strain MLY1 under pH stress (initial pH = 7.0) during THF biodegradation. Strain YYL degrades THF to intermediates, providing carbon sources for strain MLY1. THF degradation is an acid-producing process, and the protons produced need be transported to the extracellular space, a process dependent upon ATP. Low pH stress causes a decrease in the metabolism of ATP, fatty acids, and amino acids in strain YYL, while strain MLY1 provides the ATP, fatty acids, and amino acids for strain YYL to relieve the pH stress. Red text represents up-regulated metabolism; green text represents down-regulated metabolism. The middle-dashed arrows point in the direction of metabolic flow between strain YYL and MLY1.

most of the enzymes in charge of steroid transformation belong to the family of cytochrome P450 monooxygenases (P450s) in the Actinobacteria (Shtratnikova et al., 2016); this enzyme activity can be inhibited by THF (Urlacher and Girhard, 2012). Therefore, the THF inhibition of the P450s in strain YYL grown with initial pH of 7.0 might induce the up-regulation of the genes encoding for the P450s to ensure adequate P450 function for steroid transformation, for instance.

At initial pH of 7.0, the co-culture of strain YYL and strain MLY1 causes down-regulation of the bile acid and fatty acid biosynthesis necessary for lipid production in strain YYL (Supplementary Figure S1A). Simultaneously, strain MLY1 up-regulates the genes involved in the biosynthesis of fatty acids and lipids, including cutin, suberine, and wax (**Figure 5**). We propose that strain MLY1 might contribute the fatty acids and lipids to help strain YYL respond to the low pH stress in the co-culture system.

Flow of ATP from Strain MLY1 to Strain YYL

Bacteria frequently encounter environmental stress that generates a severe demand for ATP (Kobayashi et al., 1986); for example, exposure to low pH in the case of *Escherichia coli* requires ATP (Bearson et al., 1997). However, strain YYL exposure to pH stress (initial pH = 7.0) causes down-regulation of the genes involved in ATP production, such as the ATPase, cell respiration, and carbohydrate metabolic functions. Generally, ATPase activity, a good marker for the condition of the cell membrane and its enzymatic activity (Garbay and Lonvaudfunel, 1994), is essential for growth; the ATPase counteracts variations in the cytoplasmic pH by pumping protons out of the cells during low pH stress (Kullen and Klaenhammer, 1999; Kuhnert et al., 2004). The down-regulation of ATPases by strain YYLunder low pH stress (initial pH = 7.0) is not beneficial for responding to the pH stress. Therefore, the down-regulation of genes involved in ATP production likely explains the poor efficiency of THF degradation by strain YYL when grown at initial pH of 7.0.

The lack of ATPase activity and extrusion of protons during low pH stress in strain YYL leads YYL to require adequate support from strain MLY1. Comparing strain YYL co-cultured with strain MLY1 using the optimal pH (initial pH = 8.3), we noted the down-regulation of genes involved in ATP synthesis coupled with proton transport and the tricarboxylic acid cycle in the strain YYL co-cultured under pH stress (initial pH = 7.0) (**Figure 3**). Simultaneously, strain YYL up-regulated the genes involved in starch and sucrose metabolism, which could generate ATP in response to the low pH stress.

Strain MLY1 Could Provide Amino Acids and Synthetic Proteins to Strain YYL

During environmental stress, organisms must tightly regulate the activity of translation because of the high-energy consumption of

protein synthesis (Sorensen and Loeschcke, 2007; Yamasaki and Anderson, 2008). To deal with pH stress (initial pH = 7.0), strain YYL up-regulates the genes responsible for RNA transport but down-regulates the genes responsible for amino acid transport. These alterations in gene expression in strain YYL might illustrate a significant reprogramming of protein translation for strain YYL to respond to the pH stress.

In a synergistic relationship of bacteria, the exchange of amino acids frequently occurs (Nobu et al., 2015). When under pH stress (initial pH = 7.0), co-culture with strain MLY1 causes the down-regulation of valine, leucine, and isoleucine biosynthesis in strain YYL (Supplementary Figure S1B). Simultaneously, strain MLY1 up-regulates the genes involved in the biosynthesis of amino acids and translation (**Figure 5D** and Supplementary Figure S4). It is likely that strain MLY1 provides amino acids and some synthetic proteins for strain YYL in the co-culture system.

In summary, this work demonstrates that pH stress imposes an inhibitory effect on the THF-degrading strain YYL. Transcriptome analysis reveals that strain YYL up-regulates the genes involved in anti-oxidation, mutation, steroid and bile acid metabolism, and translation while simultaneously down-regulating the genes involved in ATP production when experiencing the low pH stress. MLY1 has no THF degradation activity, but it could provide fatty acids, ATP, and amino acids for strain YYL in response to the pH stress in the co-culture systems which might relieve the inhibition of strain YYL.

LIMITATION STATEMENT

The results reported in this article are only preliminary as four biological replicates were pooled for each condition investigated prior to sequencing.

AUTHOR CONTRIBUTIONS

Performed experiments: ZuL and ZH. Analyzed data: ZuL, ZH, HH, and XR. Conceived and designed experiments: ZuL, ZH, HH, XR, AO, and ZhL. All authors have agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (Nos. 41630637 and 31422003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02297/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bearson, S., Bearson, B., and Foster, J. W. (1997). Acid stress responses in enterobacteria. *FEMS. Microbiol. Lett.* 147, 173–180. doi: 10.1111/j.1574-6968.1997.tb10238.x

Benomar, S., Ranava, D., Cardenas, M. L., Trably, E., Rafrafi, Y., Ducret, A., et al. (2015). Nutritional stress induces exchange of cell material and energetic coupling between bacterial species. *Nat. Commun.* 6:6283. doi: 10.1038/ncomms7283

Buch-Pedersen, M. J., Pedersen, B. P., Veierskov, B., and Palmgren, M. G. (2009). Protons and how they are transported by proton pumps. *Pflugers Arch.* 457, 573–579. doi: 10.1007/s00424-008-0503-8

Chen, J. M., Zhou, Y. Y., Chen, D. Z., and Jin, X. J. (2010). A newly isolated strain capable of effectively degrading tetrahydrofuran and its performance in a continuous flow system. *Bioresour. Technol.* 101, 6461–6467. doi: 10.1016/j.biortech.2010.03.064

Chhabra, R. S., Elwell, M. R., Chou, B., Miller, R. A., and Renne, R. A. (1990). Subchronic toxicity of tetrahydrofuran vapors in rats and mice. *Fundam. Appl. Toxicol.* 14, 338–345. doi: 10.1016/0272-0590(90)90213-4

Cotter, P. D., and Hill, C. (2003). Surviving the acid test: responses of gram-positive bacteria to low pH. *Microbiol. Mol. Biol. Rev.* 67, 429–453. doi: 10.1128/MMBR.67.3.429-453.2003

Daye, K. J., Groff, J. C., Kirpekar, A. C., and Mazumder, R. (2003). High efficiency degradation of tetrahydrofuran (THF) using a membrane bioreactor: identification of THF-degrading cultures of *Pseudonocardia* sp. strain M1 and *Rhodococcus ruber* isolate M2. *J. Ind. Microbiol. Biotechnol.* 30, 705–714. doi: 10.1007/s10295-003-0103-8

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5685–5689. doi: 10.1073/pnas.89.12.5685

Devi, K., Mishra, S. K., Sahu, J., Panda, D., Modi, M. K., and Sen, P. (2016). Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of *Cymbopogon winterianus*. *Sci. Rep.* 6:21026. doi: 10.1038/srep21026

Di Gioia, D., Fambrini, L., Coppini, E., Fava, F., and Barberio, C. (2004). Aggregation-based cooperation during bacterial aerobic degradation of polyethoxylated nonylphenols. *Res. Microbiol.* 155, 761–769. doi: 10.1016/j.resmic.2004.05.015

Donova, M. V. (2007). Transformation of steroids by actinobacteria: a review. *Appl. Biochem. Microbiol.* 43, 5–18. doi: 10.1134/S0003683807010012

Garbay, S., and Lonvaudfunel, A. (1994). Characterization of membrane-bound ATPase activity of leuconostoc-oenos-growth-conditions. *Appl. Microbiol. Biotechnol.* 41, 597–602. doi: 10.1007/BF00178496

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.* 2, 95–108. doi: 10.1038/nrmicro821

Haussmann, U., Wolters, D. A., Franzel, B., Eltis, L. D., and Poetsch, A. (2013). Physiological adaptation of the *Rhodococcus jostii* RHA1 membrane proteome to steroids as growth substrates. *J. Proteome Res.* 12, 1188–1198. doi: 10.1021/pr300816n

He, Z., Yao, Y., Lu, Z., and Ye, Y. (2014). Dynamic metabolic and transcriptional profiling of *Rhodococcus* sp. strain YYL during the degradation of tetrahydrofuran. *Appl. Environ. Microbiol.* 80, 2656–2664. doi: 10.1128/AEM.04131-13

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Huang, X., and Madan, A. (1999). CAP3: a DNA sequence assembly program. *Genome Res.* 9, 868–877. doi: 10.1101/gr.9.9.868

Huertas, M., Sáez, L., Roldán, M., Luque-Almagro, V., Martínez-Luque, M., Blasco, R., et al. (2010). Alkaline cyanide degradation by *Pseudomonas pseudoalcaligenes* CECT5344 in a batch reactor. Influence of pH. *J. Hazard. Mater.* 179, 72–78. doi: 10.1016/j.jhazmat.2010.02.059

Isaacson, C., Mohr, T. K., and Field, J. A. (2006). Quantitative determination of 1,4-dioxane and tetrahydrofuran in groundwater by solid phase extraction GC/MS/MS. *Environ. Sci. Technol.* 40, 7305–7311. doi: 10.1021/es061 5270

Kadam, A. A., Telke, A. A., Jagtap, A. S., and Govindwar, S. P. (2011). Decolorization of adsorbed textile dyes by developed consortium of *Pseudomonas* sp. SUK1 and *Aspergillus ochraceus* NCIM-1146 under solid state fermentation. *J. Hazard. Mater.* 189, 486–494. doi: 10.1016/j.jhazmat.2011. 02.066

Kainkaryam, R. M., Bruex, A., Gilbert, A. C., Schiefelbein, J., and Woolf, P. J. (2010). PoolMC: smart pooling of mRNA samples in microarray experiments. *BMC Bioinformatics* 11:299. doi: 10.1186/1471-2105-11-299

Kobayashi, H., Suzuki, T., and Unemoto, T. (1986). Streptococcal cytoplasmic pH is regulated by changes in amount and activity of a proton-translocating ATPase. *J. Biol. Chem.* 261, 627–630.

Kohlweyer, U., Thiemer, B., Schrader, T., and Andreesen, J. R. (2000). Tetrahydrofuran degradation by a newly isolated culture of *Pseudonocardia* sp. strain K1. *FEMS Microbiol. Lett.* 186, 301–306. doi: 10.1111/j.1574-6968.2000. tb09121.x

Krastanov, A., Alexieva, Z., and Yemendzhiev, H. (2013). Microbial degradation of phenol and phenolic derivatives. *Eng. Life Sci.* 13, 76–87. doi: 10.1002/elsc. 201100227

Kuhnert, W. L., Zheng, G. L., Faustoferri, R. C., and Quivey, R. G. (2004). The F-ATPase operon promoter of *Streptococcus mutans* is transcriptionally regulated in response to external pH. *J. Bacteriol.* 186, 8524–8528. doi: 10.1128/ Jb.186.24.8524-8528.2004

Kullen, M. J., and Klaenhammer, T. R. (1999). Identification of the pH-inducible, proton-translocating F_1F_0-ATPase (*atpBEFHAGDC*) operon of *Lactobacillus acidophilus* by differential display: gene structure, cloning and characterization. *Mol. Microbiol.* 33, 1152–1161. doi: 10.1046/j.1365-2958.1999.01557.x

Larkin, M. J., Kulakov, L. A., and Allen, C. C. R. (2006). Biodegradation by members of the genus *Rhodococcus*: biochemistry, physiology, and genetic adaptation. *Adv. Appl. Microbiol.* 59, 1–29. doi: 10.1016/S0065-2164(06)59001-X

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Lowe, S. E., Jain, M. K., and Zeikus, J. G. (1993). Biology, ecology, and biotechnological applications of anaerobic bacteria adapted to environmental stresses in temperature, pH, salinity, or substrates. *Microbiol. Rev.* 57, 451–509.

Ma, X. K., Ding, N., Peterson, E. C., and Daugulis, A. J. (2016). Heavy metals species affect fungal-bacterial synergism during the bioremediation of fluoranthene. *Appl. Microbiol. Biotechnol.* 100, 7741–7750. doi: 10.1007/s00253-016-7595-4

Malley, L. A., Christoph, G. R., Stadler, J. C., Hansen, J. F., Biesemeier, J. A., and Jasti, S. L. (2001). Acute and subchronic neurotoxicological evaluation of tetrahydrofuran by inhalation in rats. *Drug Chem. Toxicol.* 24, 201–219. doi: 10.1081/DCT-100103720

Masuda, H., McClay, K., Steffan, R. J., and Zylstra, G. J. (2012). Biodegradation of tetrahydrofuran and 1,4-dioxane by soluble diiron monooxygenase in *Pseudonocardia* sp. strain ENV478. *J. Mol. Microbiol. Biotechnol.* 22, 312–316. doi: 10.1159/000343817

Mikeskova, H., Novotny, C., and Svobodova, K. (2012). Interspecific interactions in mixed microbial cultures in a biodegradation perspective. *Appl. Microbiol. Biotechnol.* 95, 861–870. doi: 10.1007/s00253-012-4234-6

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nobu, M. K., Narihiro, T., Rinke, C., Kamagata, Y., Tringe, S. G., Woyke, T., et al. (2015). Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* 9, 1710–1722. doi: 10.1038/ ismej.2014.256

Orro, A., Cappelletti, M., D'Ursi, P., Milanesi, L., Di Canito, A., Zampolli, J., et al. (2015). Genome and phenotype microarray analyses of *Rhodococcus* sp. BCP1 and *Rhodococcus opacus* R7: genetic determinants and metabolic abilities with environmental relevance. *PLOS ONE* 10:e0139467. doi: 10.1371/journal.pone. 0139467

Ren, C. G., Bai, Y. J., Kong, C. C., Bian, B., and Xie, Z. H. (2016). Synergistic interactions between salt-tolerant rhizobia and arbuscular mycorrhizal fungi on salinity tolerance of *Sesbania cannabina* plants. *J. Plant Growth Regul.* 35, 1098–1107. doi: 10.1007/s00344-016-9607-0

Rodrigues, C. D., Ramírez-Guadiana, F. H., Meeske, A. J., Wang, X., and Rudner, D. Z. (2016). GerM is required to assemble the basal platform of the SpoIIIA–SpoIIQ transenvelope complex during sporulation in *Bacillus subtilis*. *Mol. Microbiol.* 102, 260–273. doi: 10.1111/mmi. 13457

Rosenthal, A. Z., Matson, E. G., Eldar, A., and Leadbetter, J. R. (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J.* 5, 1133–1142. doi: 10.1038/ismej. 2011.3

Shtratnikova, V. Y., Schelkunov, M. I., Fokina, V. V., Pekov, Y. A., Ivashina, T., and Donova, M. V. (2016). Genome-wide bioinformatics analysis of steroid metabolism-associated genes in Nocardioides simplex VKM Ac-2033D. *Curr. Genet.* 62, 643–656. doi: 10.1007/s00294-016-0568-4

Sorensen, J. G., and Loeschcke, V. (2007). Studying stress responses in the post-genomic era: its ecological and evolutionary role. *J. Biosci.* 32, 447–456. doi: 10.1007/s12038-007-0044-x

Tajima, T., Hayashida, N., Matsumura, R., Omura, A., Nakashimada, Y., and Kato, J. (2012). Isolation and characterization of tetrahydrofuran-degrading *Rhodococcus aetherivorans* strain M8. *Process. Biochem.* 47, 1665–1669. doi: 10.1016/j.procbio.2011.08.009

Urlacher, V. B., and Girhard, M. (2012). Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends Biotechnol.* 30, 26–36. doi: 10.1016/j.tibtech.2011.06.012

Vilchez-Vargas, R., Junca, H., and Pieper, D. H. (2010). Metabolic networks, microbial ecology and 'omics' technologies: towards understanding *in situ* biodegradation processes. *Environ. Microbiol.* 12, 3089–3104. doi: 10.1111/j. 1462-2920.2010.02340.x

Wong, Y. H., Sun, J., He, L. S., Chen, L. G., Qiu, J. W., and Qian, P. Y. (2015). High-throughput transcriptome sequencing of the cold seep mussel *Bathymodiolus platifrons*. *Sci. Rep.* 5:16597. doi: 10.1038/srep16597

Wu, X. L., Liang, R. X., Dai, Q. Y., Jin, D. C., Wang, Y. Y., and Chao, W. L. (2010). Complete degradation of di-n-octyl phthalate by biochemical cooperation between *Gordonia* sp. strain JDC-2 and *Arthrobacter* sp. strain JDC-32 isolated from activated sludge. *J. Hazard. Mater.* 176, 262–268. doi: 10.1016/j.jhazmat. 2009.11.022

Yam, K. C., Okamoto, S., Roberts, J. N., and Eltis, L. D. (2011). Adventures in *Rhodococcus* - from steroids to explosives. *Can. J. Microbiol.* 57, 155–168. doi: 10.1139/W10-115

Yamasaki, S., and Anderson, P. (2008). Reprogramming mRNA translation during stress. *Curr. Opin. Cell. Biol.* 20, 222–226. doi: 10.1016/j.ceb.2008.01.013

Yao, Y., Lu, Z., Zhu, F., Min, H., and Bian, C. (2013). Successful bioaugmentation of an activated sludge reactor with *Rhodococcus* sp. YYL for efficient tetrahydrofuran degradation. *J. Hazard. Mater.* 261, 550–558. doi: 10.1016/j. jhazmat.2013.08.007

Yao, Y. L., Lv, Z. M., Min, H., Lv, Z. H., and Jiao, H. P. (2009). Isolation, identification and characterization of a novel *Rhodococcus* sp. strain in biodegradation of tetrahydrofuran and its medium optimization using sequential statistics-based experimental designs. *Bioresour. Technol.* 100, 2762–2769. doi: 10.1016/j.biortech.2009.01.006

Zhang, S. D., and Gant, T. W. (2005). Effect of pooling samples on the efficiency of comparative studies using microarrays. *Bioinformatics* 21, 4378–4383. doi: 10.1093/bioinformatics/bti717

Zhao, Y., Bi, K., Gao, Z., Chen, T., Liu, H. Q., Xie, J. T., et al. (2017). Transcriptome analysis of *Arabidopsis thaliana* in response to *Plasmodiophora brassicae* during early infection. *Front. Microbiol.* 2017:673. doi: 10.3389/fmicb.2017. 00673

Zhao, Z., and Wong, J. W. (2009). Biosurfactants from *Acinetobacter calcoaceticus* BU03 enhance the solubility and biodegradation of phenanthrene. *Environ. Technol.* 30, 291–299. doi: 10.1080/09593330802630801

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Liu, He, Huang, Ran, Oluwafunmilayo and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 March 2017
doi: 10.3389/fmicb.2017.00462

The Role of Enriched Microbial Consortium on Iron-Reducing Bioaugmentation in Sediments

Yuanyuan Pan[1,2,3,4], Xunan Yang[1,3,4], Meiying Xu[1,3,4] and Guoping Sun[1,3,4]**

[1] Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, Guangdong Institute of Microbiology, Guangzhou, China, [2] School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China, [3] State Key Laboratory of Applied Microbiology Southern China, Guangzhou, China, [4] Guangdong Open Laboratory of Applied Microbiology, Guangzhou, China

Microbial iron reduction is an important biogeochemical process and involved in various engineered processes, including the traditional clay dyeing processes. Bioaugmentation with iron reducing bacteria (IRB) is generally considered as an effective method to enhance the activity of iron reduction. However, limited information is available about the role of IRB on bioaugmentation. To reveal the roles of introduced IRB on bioaugmentation, an IRB consortium enriched with ferric citrate was inoculated into three Fe(II)-poor sediments which served as the pigments for Gambiered Guangdong silk dyeing. After bioaugmentation, the dyeabilities of all sediments met the demands of Gambiered Guangdong silk through increasing the concentration of key agent [precipitated Fe(II)] by 35, 27, and 61%, respectively. The microbial community analysis revealed that it was the minor species but not the dominant ones in the IRB consortium that promoted the activity of iron reduction. Meanwhile, some indigenous bacteria with the potential of iron reduction, such as *Clostridium*, *Anaeromyxobacter*, *Bacillus*, *Pseudomonas*, *Geothrix*, and *Acinetobacter*, were also stimulated to form mutualistic interaction with introduced consortium. Interestingly, the same initial IRB consortium led to the different community successions among the three sediments and there was even no common genus increasing or decreasing synchronously among the potential IRB of all bioaugmented sediments. The Mantel and canonical correspondence analysis showed that different physiochemical properties of sediments influenced the microbial community structures. This study not only provides a novel bioremediation method for obtaining usable sediments for dyeing Gambiered Guangdong silk, but also contributes to understanding the microbial response to IRB bioaugmentation.

Keywords: iron reducing bacteria, bioaugmentation, consortium, river sediments, high-throughput sequencing, microbial response

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Seung Gu Shin, Pohang University of Science and Technology, South Korea
Jun-Jie Zhang, Indiana University School of Medicine, USA

****Correspondence:***
Guoping Sun
sgpgim@163.com
Xunan Yang
yangxn@gdim.cn

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *06 December 2016*
Accepted: *06 March 2017*
Published: *20 March 2017*

Citation:
Pan Y, Yang X, Xu M and Sun G (2017) The Role of Enriched Microbial Consortium on Iron-Reducing Bioaugmentation in Sediments. Front. Microbiol. 8:462. doi: 10.3389/fmicb.2017.00462

INTRODUCTION

Microbial iron reduction as a fundamental biogeochemical process widely exists in the freshwater sediments. It is regarded as the crucial mediator in the carbon, nitrogen, sulfur, and phosphorus cycles (Li et al., 2012). Furthermore, iron reduction plays an important role on degradation of organic contaminants (Zhang et al., 2013; Baek et al., 2016) and bioremediation of toxic metal

compounds (Hassan et al., 2015; Si et al., 2015), and particularly, it is also vital for sediments used in environmental-friendly and traditional clay dyeing processes such as mud-tannic dyeing techniques (Pan et al., 2016). Therefore, investigating iron reduction is crucial to understand biogeochemical dynamics in natural sediment environments.

Microbial iron reduction in the sediments depends on quantity and activity of iron reducing bacteria (IRB) and the amounts of reducible Fe(III) such as amorphous Fe(III) oxides (Lovley, 2006). Therefore, in order to increase the activity of iron reduction, bioaugmentation could be an effective, economical and environmental-friendly approach to provide sufficient microbes with special functions. Since the pure specialized strains often failed to compete with indigenous bacteria in the sediment environments (Thompson et al., 2005), consortia with higher diversity were considered as a better choice to enhance the activity of iron reduction. Although a few studies have reported the possibility and efficiency of bioaugmentation with IRB to enhance iron reduction and then affect the performance of anaerobic digestion (Baek et al., 2016), the roles of microbial communities behind the effects remain unknown, such as the survival of introduced consortium and the shift of microbial communities. There are two reasons as follows. Firstly, due to insufficient sequences, the traditional molecular techniques such as denaturing gradient gel electrophoresis and terminal restriction fragment length polymorphism could not exactly monitor the change of microbial communities (Gao and Tao, 2012; Baek et al., 2016). Secondly, the low resolution of traditional techniques leads to difficulty tracking the colonization and succession of the introduced microorganisms (Lentini et al., 2012), especially for IRB consortium with no specific functional gene marker. Nowadays, as the developments of the next generation sequencing, high-throughput microbial community analysis has been applied to study the dynamic succession of environmental microbial communities (Zhou et al., 2014; Zhao et al., 2016). With this method, researchers could track the abundance of introduced consortium without any specific gene marker, and link the succession of microbial communities including indigenous bacteria to the change of physicochemical properties in IRB bioaugmentation system.

Currently, two kinds of viewpoints were proposed about the key factor that influenced the bioaugmentation performance with selected strains or consortia. In general, many studies found that the survival and function of introduced strains were the vital factors influencing the bioaugmentation performance (Herrero and Stuckey, 2015; Baek et al., 2016). In contrast, several studies suggested that there was no direct relation between the abundance of introduced strains and bioaugmentation performance. They speculated that the shift of indigenous microbial community enhanced the performance (Qu et al., 2015; Xun et al., 2015). In terms of the bioaugmentation of IRB consortium, it still needs further investigation to ensure which factor affects the iron respiration activity.

In this study, amending the sediment for Gambiered Guangdong silk was served as a case to study microbial effects during IRB bioaugmentation. Generally, traditional dyeing crafts were obtained through the reaction between the ferric iron clay minerals (e.g., goethite, hematite, palygorskite, and akaganeite) and natural organic dyes, for example, the well-known Maya Blue (Van Olphen, 1966), Bogolan cloth in Mali (Hilu and Hersey, 2005), and Amami Oshima Tsumugi (Wakimoto et al., 2004). Nevertheless, the sediment with abundant precipitated Fe(II) was vital for Gambiered Guangdong silk and it reacted with tannins pre-adsorbed on the silk to form shiny black color (Pan et al., 2016). Usually, the available sediment could be obtained from natural environments. However, with the increasing water pollution, clean sediment is decreasing dramatically. Bioaugmentation with IRB could be a potential technology for exploring new sources of ideal sediment and increasing the recycle and re-use rate of sediment. Unfortunately, very limited information was available on this sustainable method. Therefore, the aim of the present study was to (1) evaluate the feasibility of bioaugmenting three Fe(II)-poor sediments with the enriched IRB consortium; (2) explore microbial response including indigenous and exogenous microbes to IRB bioaugmentation via high-throughput sequencing; and (3) identify the dominant microorganisms which promoted sediment dyeability.

MATERIALS AND METHODS

Materials

Three original river sediments (Orig-SD1, Orig-SD2, and Orig-CH) were collected from creeks in Pearl River Delta, China. Sediments were passed through a 100-mesh standard sieve (0.15 mm opening size) and kept in a cold room (4°C) in darkness for later use. Orig-SD1 and Orig-SD2 were identified as useable sediments because fabrics coated with them met the color demand of Gambiered Guangdong silk (one side: shiny black; the other side: brown; Pan et al., 2016), while Orig-CH was unusable. The physicochemical characteristics and performance details of these three original sediments were listed in Supplementary Table S1. In order to test the bioaugmentation performance, three modified sediments with little ferrous were considered, including high temperature (121°C) and pressure (0.1 MPa) oxidized Orig-SD1 (SD1), air-dried Orig-SD2 (SD2), and unusable Orig-CH (CH).

The silk textiles used in this study were prepared in the ChengYi factory (Foshan, China), which has been dyed with the tannin extract of Ju-liang roots to form a brown color. The prepared textiles (Orig-textile, the color characteristics seen in **Table 1**) were cut into 2 cm × 2 cm square pieces for sediment coating.

Enrichment Culture of IRB

Enrichment cultures were prepared in 20-mL serum bottles containing 15 mL ferric citrate medium (g L^{-1}: ferric citrate, 3.4; NH_4Cl, 1.0; $KHPO_4$, 0.25; $K_2HPO_4.3H_2O$, 0.72; $CaCl_2.2H_2O$, 0.07; $MgSO_4.7H_2O$, 0.6; and glucose, 10) and 1 g Orig-SD1 as the inoculants which were sealed with Teflon-coated butyl rubber septa and aluminum crimp caps (Wang et al., 2008). All culture bottles were incubated at 30°C in the dark. The consortium was subcultured in a new serum bottle

TABLE 1 | Color characteristics of mud inoculated with or without the enriched IRB consortium.

Sediments	Sediment-coated side			Back side		
	L_1*	a*$_1$	b*$_1$	L_2*	a*$_2$	b*$_2$
SD1W	29.51	12.07	10.33	–	–	–
SD1S	23.11	0.29	0.98	34.36	8.78	13.6
SD2W	26.11	4.62	2.91	–	–	–
SD2S	24.91	1.32	0.95	35.43	11.13	13.17
CHW	28.52	10.73	6.88	–	–	–
CHS	24.86	0.9	1.43	37.15	11.04	14.25
Orig-SD1	24.39	0.32	0.91	36.42	10.86	14.12
Orig-textile	28.17	12.38	9.75	39.90	10.81	15.42

with 15 mL fresh culture medium once the enrichment cultures turned light green or colorless from yellow color (2–7 days, given by ferric citrate redox indicator). The subculture was consecutively repeated five times. The IRB culture was washed twice and then suspended with sterilized deionized water (OD600 = 1.0) before it was used as bioaugmentation inoculants.

Bioaugmentation Process

The bioaugmentation experiment procedures are follows: an aliquot (3 g) of the unusable sediments (SD1, SD2, and CH) transferred into the 10-mL vial. Three milliliters of enriched consortium were added into the three sediments, uniformly mixed and cultured at 30°C in the dark for 7 days (SD1S, SD2S, and CHS). Sediments added with sterilized water were used as controls (SD1W, SD2W, and CHW). Each treatment was with three replicates. After 7 days, all sediments with and without bioaugmentation were used to coat the textiles prepared as mentioned above. After reacting for 1 h, the rest sediment on the textile was washed away and the textiles were dried in the sun.

Characteristics of Inoculated Sediments and Sediment-Coated Textiles

After incubation, a thin layer of sediment was used to coat the Orig-textile for 1 h according to our previous method (Pan et al., 2016) and then the rest sediment was washed away. The color characteristics of both sides of sediment-coated textiles were evaluated with the Commission International d'Eclairage (CIE) Lab coordinates by USPRO Colorimeter (Datacolor 110TM, USA). Optical source was D65, viewing angle 10° and measure diameter 10 mm. The CIELAB color system is organized with three axes in a spherical form: L*, a*, and b*. L* is associated with the lightness of the color and moves from top (100, white) to bottom (0, black), whereas a* and b* are associated with changes in redness–greenness (positive a* is red and negative a* is green) and in yellowness–blueness (positive b* is yellow and negative b* is blue). The L*, a*, and b* were calculated from three repetitive measurements for every sample. And color difference (ΔE) between textile samples was used as the indicator on judging the dyeability of sediments and calculated according to the Eq. 1. From the point of technical dyeing using natural raw dyes, a somewhat wider color difference of $\Delta E \leq 3$ could be permitted with the use of natural resources (Martínez et al., 2001).

$$\Delta E = \sqrt{(L_i^* - L_j^*)^2 + (a_i^* - a_j^*)^2 + (b_i^* - b_j^*)^2} \quad (1)$$

L_i*, a_i*, b_i*: color values of the *i* textile;
L_j*, a_j*, b_j*: color values of the *j* textile.

The sediment pH and oxidation-reduction potential (ORP) were monitored with a S20 K pH meter (Mettler Toledo, Switzerland). The organic matter content was estimated according to the previous method (Haller et al., 2011) except extended to 4 h. HCl-extractable Fe(II) and total Fe (TFe) in the sediment were extracted with 0.5 mol/L HCl for 1 h (Lovley and Phillips, 1987). After centrifuged at 8000 rpm for 5 min, the supernatant was determined using the 1,10-phenanthroline colorimetric method at 510 nm on a full wavelength scanner (Thermo Scientific, MULTISKAN GO). TFe including Fe(III) and Fe(II) in the sediment was extracted with hydroxylamine hydrochloride and determined as the Fe(II) determination (Lovley and Phillips, 1987).

DNA Extraction and Sequencing

DNA was extracted from 250 mg of sediment samples using the PowerSoil^{TM} DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. The bacterial 16S rRNA genes were amplified using the PCR primers 515f/806r targeting the V4 region (Pylro et al., 2014). To distinguish the different samples, a Barcoded-tag with six nucleotide bases was randomly added to the upstream of the universal primer. The primers which were added with Barcoded-tag sequences were Barcoded-tag fusion primers. After quantification and quality control, PCR products were gradually diluted and quantified. The V4 tag PCR products were pooled with the other samples and sequenced using 300 bp paired-end model with the Illumina MiSeq platform at Chengdu Institute of Biology (Chengdu, China).

Sequence Data Analysis

After sequencing, the final V4 tag sequences were assembled through finding the overlap between paired-end reads by the FLASH software. Chimeras were identified via UCHIME

algorithm on mothur platform. Low quality fragments were filtered out using QIIME software. Sequences were clustered to operational taxonomic units (OTUs) at 97% sequence similarity by using UCLUST software (Edgar, 2010). Singletons were removed from the whole sequence data set and each sample was randomly sampled and normalized at 11,000 sequences. The numbers of original reads and final OTUs are listed in Supplementary Table S2. The dissimilarity test [non-metric multidimensional scaling (NMDS)] based on Bray–Curtis similarity distance matrices were performed by the Vegan package in R 3.1.3. The dominant OTUs in each group were depicted in a heat map conducted with R 3.1.3, and canonical correspondence analysis (CCA) was used to analyze the relationship between these OTUs and sediment properties with Mantel test. The Illumina sequence raw data reported here was submitted to the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP083001.

RESULTS

Performance of Bioaugmented Sediments

After 7-day incubation, all of the three bioaugmented sediments (SD1S, SD2S, and CHS) revealed higher dyeabilities than their control groups (**Table 1**). The bioaugmented sediments performed shiny black color on the sediment-coated side ($L^*_{SD1S} = 23.11$; $L^*_{SD2S} = 24.91$; $L^*_{CHS} = 24.86$), comparable with the standard cloth dyed with Orig-SD1 ($\Delta E < 3$). Furthermore, the bioaugmented sediments did not penetrate to the back and the backside still kept brown ($\Delta E_{SD1S} = 2.97$; $\Delta E_{SD2S} = 1.41$; $\Delta E_{CHS} = 0.84$), which met the demand of two tone colors for Gambiered Guangdong silk. In contrast, the control sediments (SD1W, SD2W, and CHW) still kept the original brown color ($L^*_{SD1W} = 29.51$; $L^*_{SD2W} = 26.11$; $L^*_{CHW} = 28.52$), near to the color of Orig-textile ($L^*_{Orig-textile} = 28.17$).

Sediment Properties

In order to clarify the influences of IRB inoculation on sediment dyeability, characteristics of sediments were investigated (**Table 2**). The pH and LOI of the bioaugmented groups showed a slight increase and decrease, respectively, compared to the control group, except for the pH of SD1S. As expected, the concentration of HCl-extractable Fe(II) in bioaugmented sediments were significantly higher than those in controls ($p < 0.01$), increasing by 3.5-, 12-, and 83-fold, while there was no significant change in HCl-extractable TFe (**Table 2**). In addition, the ORP results were well consistent with the Fe(II) results: the controls had positive ORP with oxidizing conditions while the IRB bioaugmented groups had negative ORP with reducing conditions for iron reducing.

Microbial Community Analysis in Enriched IRB Consortium and Sediments

Eleven thousand effective sequences were re-sampled from each sediment sample, resulting in 1,440–3,759 OTUs at 97% sequence identify cutoff (Supplementary Table S2) and being assigned to different taxa (**Figure 1**). In the enriched iron reducing consortium, Firmicutes (77.4%) was the overwhelmingly dominant phylum, followed by Proteobacteria with 11.0%. Among the Firmicutes phylum, nearly 44.3% sequences was assigned to genus *Clostridium*. At the OTU level, one-third sequences were assigned to *Clostridium* sp. (denovo 173227).

The NMDS analysis was applied to analyze sediments microbial communities with and without bioaugmentation. Although the data points of sediments with the same source were adjacent, there was a distinction between bioaugmented and non-bioaugmentation samples. Moreover, the succession of microbial communities proceeded in different directions of community succession (**Figure 2**).

In details, in terms of the phylum level (**Figure 1**), Firmicutes and Proteobacteria dominated in SD1 and SD2, while the CH community was mainly dispersed by Firmicutes, Proteobacteria, and Bacteroidetes. After bioaugmentation, the average abundance of Firmicutes increased from 55.3% (SD1W), 4.6% (SD2W), and 6.2% (CHW) to 59.7% (SD1S), 35.6% (SD2S), and 8.3% (CHS), respectively, but total percentages of Proteobacteria in all sediments decreased, even by 8.2% for SD1S. At the genus level (Supplementary Figure S1), bioaugmentation with IRB consortium led to a significant increase in the relative abundance of genera *Symbiobacterium* (by 24.0%), *Planctomyces* (by 1.1%), *Geobacter* (by 0.2%), *Novosphingobium* (by 5.8%), and *Nevskia* (by 1.7%) in SD1; *Clostridium* (by 3.0%), *Bacillus* (by 13.9%), *Brevibacillus* (by 2.8%), and *Janthinobacterium* (by 1.6%) in SD2; *Clostridium* (by

TABLE 2 | Physiochemical characteristics and dyeability evaluation of sediments inoculated with or without iron reducing consortium.

Sediment	Total Fe (μmol g^{-1})[a]	Fe(II) (μmol g^{-1})[b]	pH	ORP	LOI (%)	$\Delta E_1/\Delta E_2$
SD1W	224.38 ± 14.97	25.56 ± 17.22	6.69 ± 0.03	194.6 ± 15.70	7.29 ± 1.26	15.91/–
SD1S	247.05 ± 26.13	114.35 ± 34.24**	6.63 ± 0.06	−102.13 ± 48.31**	6.54 ± 0.48	1.29/2.97
SD2W	234.24 ± 45.36	12.21 ± 0.36	6.98 ± 0.21	20.68 ± 1.03	8.08 ± 0.84	5.06/–
SD2S	271.46 ± 11.02	86.69 ± 6.66**	7.08 ± 0.11	−64.28 ± 18.44**	7.51 ± 0.36	1.14/1.41
CHW	140.14 ± 6.92	0 ± 3.52	5.09 ± 0.07	243.13 ± 28.81	9.23 ± 0.62	12.70/–
CHS	136.91 ± 23.41	83.98 ± 16.46**	5.61 ± 0.14**	−56.05 ± 25.51**	8.11 ± 0.67	1.02/0.84

*LOI, organic matter; ORP, redox potential; total Fe, 0.5 mol L^{-1} hydroxylamine hydrochloride extraction; Fe(II), 0.5 mol L^{-1} HCl extraction; ΔE_1 and ΔE_2 represents the color difference of sediment-coated side and back side compared with Orig-SD1, respectively. $\Delta E < 3$ indicated the good dyeability. **$p < 0.01$ in t-test.*

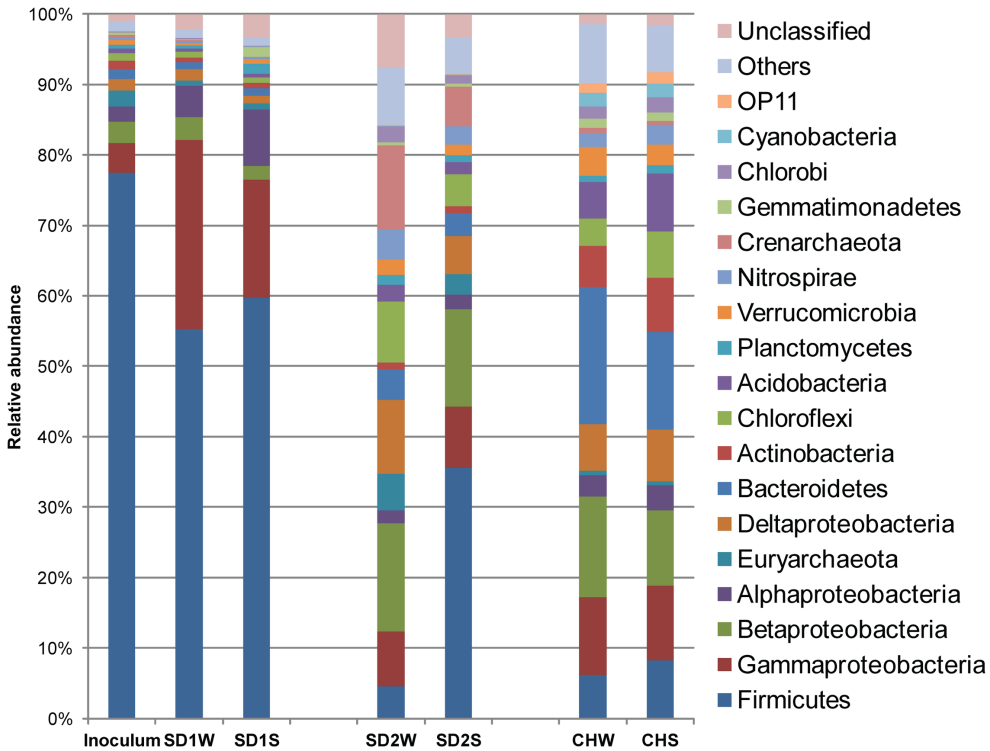


FIGURE 1 | Diversity in the bacteria communities at the phylum level in response to iron reducing consortium treatments for the three sediments, identified by 16S rRNA gene sequences. The bacterial phyla represented by >1% of total sequences are presented here. Phylum making up less than 1% of the total sequences were classified as others.

0.9%), *Anaeromyxobacter* (by 0.6%), and *Geothrix* (by 0.5%) in CH. In addition, the dominant 38 OTUs (>1% total sequences for each sample) of all the samples were analyzed (**Figure 3**), which were mainly composed of Firmicutes (17 OTUs) and Proteobacteria (12 OTUs). The major species (e.g., denovo 173227) in the IRB consortium showed no dominant position in bioaugmented sediments. The number of the significantly increased OTUs in bioaugmented sediments were 8 (SD1S), 5 (SD2S), and 6 (CHS), respectively, and some of these OTUs were different among the three sediments. Moreover, some sediment-specific (indigenous) IRB increased and be only found in their corresponding sediment (Supplementary Table S3).

Potential IRB

The relative abundance of potential iron reducers were listed in **Table 3**. The genus *Clostridium* was the most abundant genus in the IRB consortium, but just occupied 1.11, 3.55, and 1.20% in SD1S, SD2S, and CHS, respectively. *Pseudomonas* (11.64%) was the most abundant genus of IRB in SD1S. *Bacillus* (14.19%) and *Brevibacillus* (2.84%) were the predominant iron-reducing genus in SD2S and increased by 50- and 46-fold, respectively, compared with the SD2W. Differently, the iron reducing genera for CHS were dispersed and composed of *Anaeromyxobacter* (1.37%) and *Geothrix* (1.07%) both increased by about twofold. The significantly

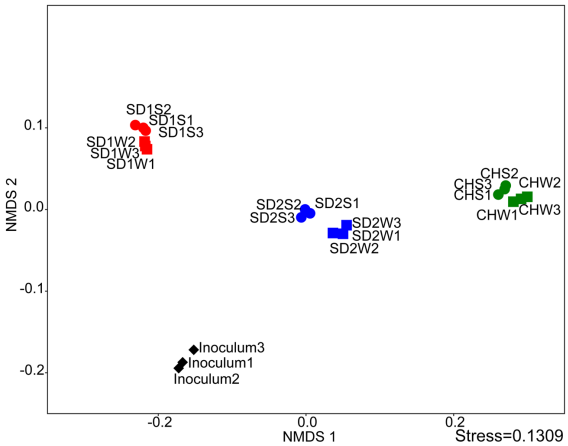


FIGURE 2 | Comparing the microbial communities between inoculated groups and the controls using NMDS analysis at the genus level.

increased OTUs in the sediments but not in the IRB consortium (Supplementary Table S3), were classified to several genera which were reported to own the ability of iron reduction such as *Anaeromyxobacter* (Breidenbach et al., 2015), *Bacillus* (Alabbas et al., 2013), *Clostridium* (Hassan et al., 2015), *Azospira* (Peng et al., 2016) *Paenibacillus* (Petrie et al., 2003),

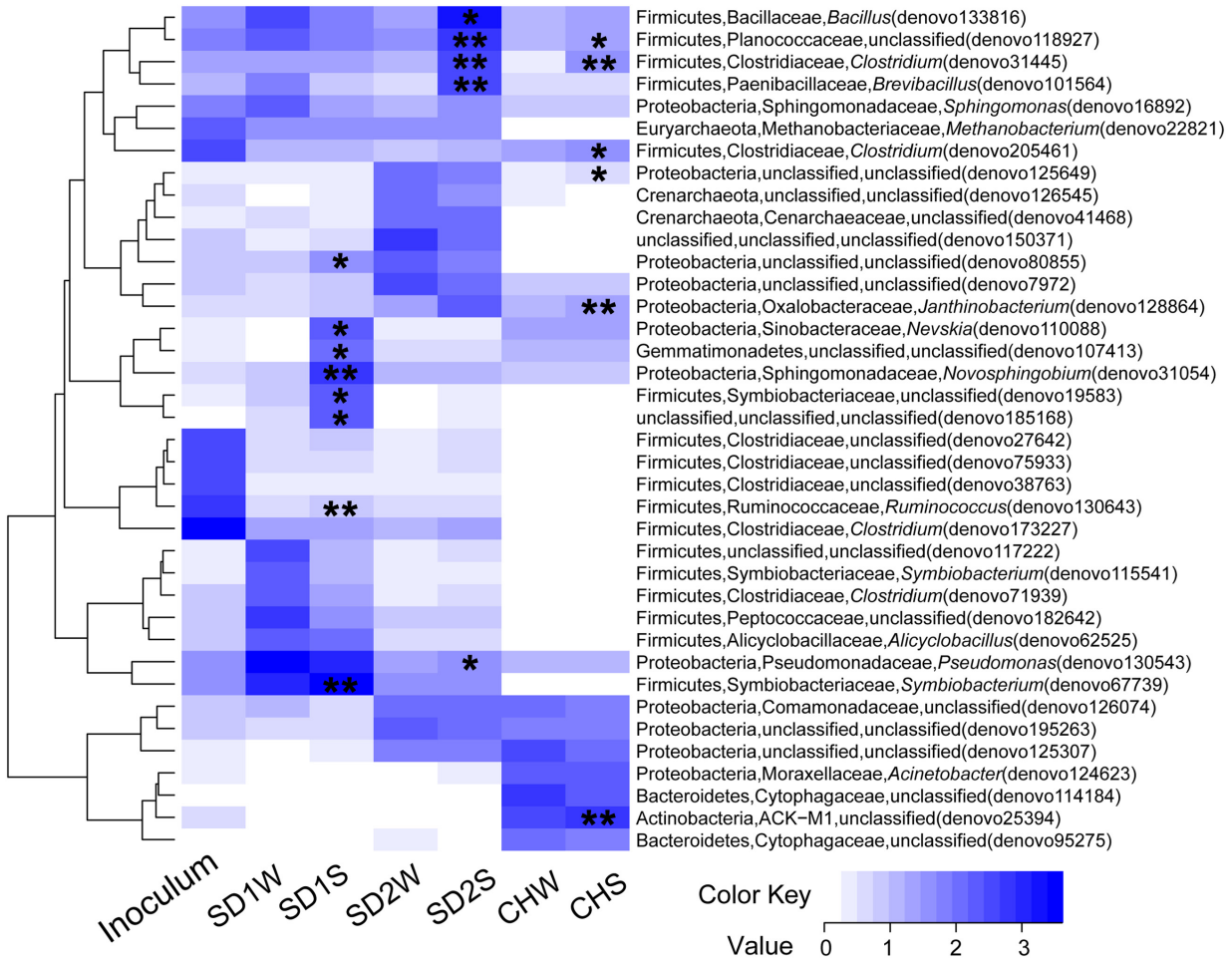


FIGURE 3 | Heatmap of the enriched iron reducing consortium and sediments with and without inoculation based on the $\log_{10}$ (sequence numbers +1) of the 38 OTUs whose reads occupied above 1% of total sequences. The phylum, family, and genus for each OTU were showed in the figure. * and ** indicated this OTU with a significant increase in the inoculated sediments (SD1S, SD2S, and CHS) compared to the relative controls (SD1W, SD2W, and CHW). *$0.01 < p < 0.05$, **$p < 0.01$.

Desulfosporosinus (Bertel et al., 2012), and *Treponema* (Baek et al., 2016).

Relationship between Microbial Community and Sediment Properties

Significant correlation was observed between the composition of the bacterial community (OTU level) and physiochemical properties [pH, ORP, Fe(III), LOI] of sediments with different sources through the Mantel test ($p < 0.01$) and canonical correspondence analysis (CCA) (**Figure 4**). CCA showed the first two components (CCA1 and CCA2) together explained 76.68% of the total variation of sediment microbial community. Although some parameters [e.g., Fe(II) and ORP] had significant changes after bioaugmentation (**Table 2**), greater sediment-to-sediment difference than treatment-to-control in CCA profile (**Figure 4**) implied that the whole physiochemical properties were the substantive influence on the bacterial community constitution.

DISCUSSION

While microbial iron reduction in the sediment environments has been concerned for several decades, little has been done to characterize bioaugmented iron reduction with IRB consortium. In this study, we successfully enhanced iron reduction and then achieved the goal of improving the sediment dyeability for Gambiered Guangdong Silk by inoculating enriched IRB consortium (**Table 1**). As expected, the concentrations of HCl-extractable ferrous in all the bioaugmented sediments increased by 35% (SD1), 26% (SD2), and 61% (CH) comparing with the control groups, respectively (**Table 2**). Meanwhile, the negative ORP with reducing conditions favored the precipitation of ferrous (Bongoua-Devisme et al., 2013). These results were coincident with our previous study which suggested that precipitated ferrous iron in the sediment was the key factor for the success of dyeing technique of Gambiered Guangdong silk (Pan et al., 2016).

TABLE 3 | The potential iron reducing bacteria in different phylogenetic OTUs taxa obtained by pyrosequencing of 16S rRNA genes using Miseq platforms.

Taxonomic description	Percent of total sequence (%)						
	Consortium	SD1W	SD1S	SD2W	SD2S	CHW	CHS
Firmicutes							
Clostridia							
Clostridiaceae							
Clostridium	**44.25**	**4.08**	**1.11**	0.53	**3.55****	0.34	**1.20****
Peptococcaceae							
Desulfosporosinus	0.12	**2.02**	**0.99**	0.05	0.14	0.02	0.06
Bacilli							
Bacillaceae							
Bacillus	0.54	**2.62**	0.61	0.28	**14.19****	0.32	0.34
Paenibacillaceae							
Brevibacillus	0.17	**1.46**	0.34	0.06	**2.84****	0.03	0.03
Proteobacteria							
Betaproteobacteria							
Pseudomonadaceae							
Pseudomonas	0.44	**23.37**	**11.64**	0.25	**1.31**	0.42	0.40
Deltaproteobacteria							
Myxococcaceae							
Anaeromyxobacter	0.13	0.05	0.05	0.29	0.28	0.76	**1.37****
Geobacteraceae							
Geobacter	0.17	0.10	0.27**	0.45	0.27	0.83	**1.06**
Gammaproteobacteria							
Moraxellaceae							
Acinetobacter	0.05	0.02	0.01	0.01	0.02	**3.52**	**3.35**
Acidobacteria							
Holophagae							
Holophagaceae							
Geothrix	0.00	0.00	0.00	0.05	0.04	0.54	**1.07****

*Consortium represented for the enriched iron reducing bacteria. **Indicated that the relative abundances in inoculated samples were higher than those in their controls significantly ($p < 0.01$). Bold numbers indicate the most important percent of total sequence. The potential iron reducing genera were classified according to the published references, Clostridium (Hassan et al., 2015), Desulfosporosinus (Bertel et al., 2012), Bacillus (Alabbas et al., 2013), Brevibacillus (Ding et al., 2015), Pseudomonas (Naganuma et al., 2006), Anaeromyxobacter (Breidenbach et al., 2015), Geobacter (Lentini et al., 2012), Acinetobacter (Hassan et al., 2015), and Geothrix (Nevin and Lovley, 2002).*

To explain that the enhanced iron reduction was ascribed to either the survival of introduced IRB consortium or the shift of indigenous microorganisms, the microbial communities were taken insight through high-throughput technique. In general, the dominant bacteria would be responsible for the enhanced performance. Unexpectedly, the dominant specie (denovo 173227) in enriched IRB consortium, which accounted for 42.95% of total sequences and assigned to potential iron-reducing genus *Clostridium* (Li et al., 2011; Peng et al., 2016), did not proliferate significantly in bioaugmented sediments and only accounted for little percentages (0.22, 0.21, and 0%, respectively, in SD1S, SD2S, and CHS). Moreover, the other major OTUs in the consortium also lost their dominant position after augmented into sediments (**Figure 3**), which implied the uncertain risk in using the well-growing strains as the augmentation agent. In this study, however, the bioaugmentation was success, which might be attributed to the minor species in the inoculums. Minor species are considered as the seed bank in a microbial community (Pedrós-Alió, 2006; Campbell et al., 2011) and serve as the keystone within complex consortia with the potential to become dominant in response to shifts in environmental conditions (Sogin et al., 2006). These minor species survived in new environment and were responsible for iron reducing function (**Figure 3** and **Tables 2**, **3**). As **Figure 3** illustrated, the minor OTUs once in consortium substantially showed higher abundances in bioaugmented sediments than those in relative controls, and most of them were assigned to several potential iron-reducing genera, such as *Bacillus*, *Brevibacillus*, *Clostridium*, and *Pseudomonas* (**Table 3**). In addition, some indigenous OTUs (only observed in sediments) also increased after bioaugmentation (Supplementary Table S3), which might be attributed to the mutualism that the exogenous species collaborate with the indigenous IRB and perform higher exploitability of Fe(III). These stimulated bacteria and survival species in the consortium formed a multispecies interactive

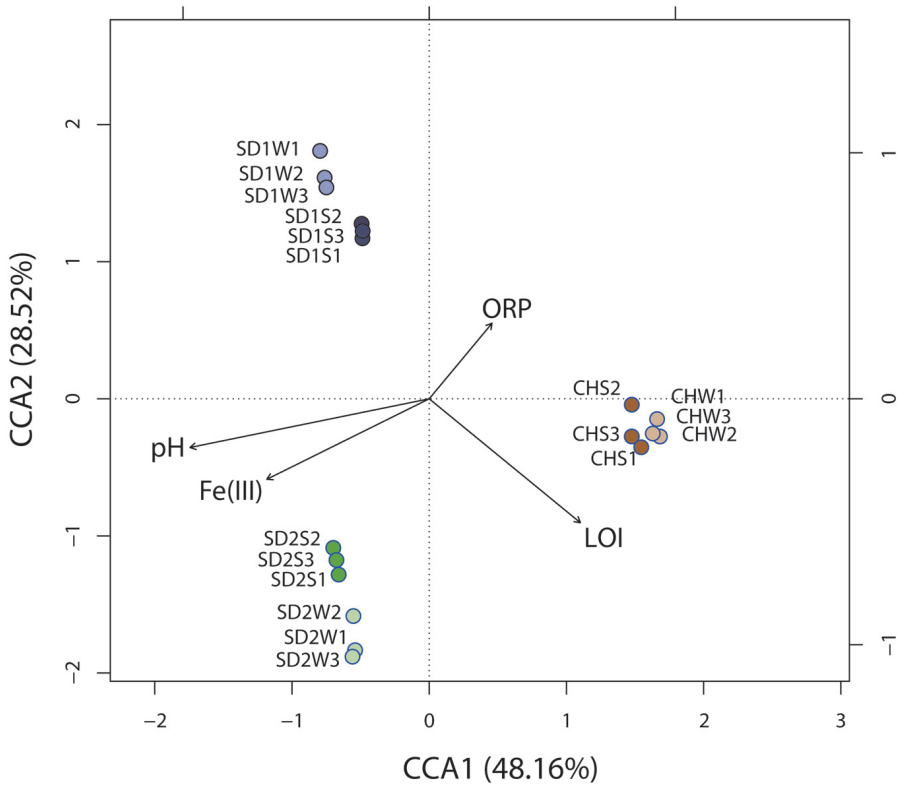


FIGURE 4 | Canonical correspondence analysis (CCA) of microbial community structures (OTUs) from six sediment samples with respect to the four environmental variables. Arrows indicate the direction and magnitude of measurable variables associated with community structures.

network. Therefore, iron reduction could be functioned not only through proliferating the exogenous minor taxa but also collaborators with the indigenous IRB.

Interestingly, as the **Table 3** and **Figures 2**, **3** shown, the same consortium had triggered the succession in different directions. These might be attributed to the differences in the sediment characteristics (Böer et al., 2009). Distinct OTUs had a differential niche adaptation and tended to adapt to changes in their environments (Storey et al., 2015). Some members might extinct due to poor ability to adapt to changing environments, while others could proliferate in the new environments (Faust and Raes, 2012). Eventually, the relative balanced state of community structure was formed based on the physicochemical characteristics (Zhou et al., 2014; Qu et al., 2015; Sanders, 2016). In the present study, significant correlation was definitely observed between the composition of the bacterial community and such properties of sediments with different sources through the Mantel test ($p < 0.01$) and CCA (**Figure 4**). Furthermore, the bacterial community structure analysis showed reproducible inter-groups and significantly different intra-groups (**Figures 2**, **4**). The results indicated that the environment conditions influenced the microbial communities, which also explained the reason that specific IRB played the roles in different sediments (**Table 3**). That is, because of the allopatric speciation, the same consortium formed different functional assemblies. This implied that inoculating a consortium was equivalent to providing a function library (seed bank); the key contributors, who might used to be minor species, would function in compatible environment. Therefore, it is also suggested that bioaugmentation with microbial consortia might be a better choice than with specialized strains, due to their adaptation to a wider environmental conditions as well as their synergistic interactions.

CONCLUSION

The findings demonstrated that sediments bioaugmented with enriched IRB consortium obtained good dyeability for Gambiered Guangdong silk due to the increased Fe(II) concentration. The bioaugmentation process facilitated iron reduction through the mutualism interaction between survived minor species from the IRB consortium and some stimulated indigenous bacteria including *Clostridium*, *Anaeromyxobacter*, *Bacillus*, *Pseudomonas*, *Geothrix*, and *Acinetobacter*. Meanwhile, due to the different physiochemical properties, the same IRB consortium led to the community successions on different directions in the three sediments. This study suggested that consortium might be a better choice than pure strains because of a lower requirement for environmental conditions.

ACKNOWLEDGMENT

The authors acknowledge the kind help of Zhu Liang (one of the two inheritors for the dyeing technique of Gambiered Guangdong silk, China) for providing textiles and introducing us to dyeing processes.

AUTHOR CONTRIBUTIONS

YP designed the study, performed experiments, analyzed the data, and wrote the manuscript; XY analyzed the data, interpreted the results, and revised the manuscript; MX and GS revised the manuscript and approved the final version.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No. 51508111), Science and Technology Planning Project of Guangdong Province, China (Nos. 2014A020220006, 2016B070701017, and 2013B091500081), and the Natural Science Foundation of Guangdong Province, China (No. 2014A030310140).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00462/full#supplementary-material

REFERENCES

Alabbas, F. M., Bhola, S. M., Spear, J. R., Olson, D. L., and Mishra, B. (2013). The shielding effect of wild type iron reducing bacteria flora in the corrosion of linepipe steel. *Eng. Fail. Anal.* 33, 222–225. doi: 10.1016/j.engfailanal.2013.05.020

Baek, G., Kim, J., Shin, S. G., and Lee, C. (2016). Bioaugmentation of anaerobic sludge digestion with iron-reducing bacteria: process and microbial responses to variations in hydraulic retention time. *Appl. Microbiol. Biotechnol.* 100, 927–937. doi: 10.1007/s00253-015-7018-y

Bertel, D., Peck, J., Quick, T. J., and Senko, J. M. (2012). Iron transformations induced by an acid-tolerant *Desulfosporosinus* species. *Appl. Environ. Microbiol.* 78, 81–88. doi: 10.1128/AEM.06337-11

Böer, S. I., Hedtkamp, S. I. C., van Beusekom, J. E. E., and Ramette, A. (2009). Time- and sediment depth-related variations in bacterial diversity and community structure in subtidal sands. *ISME J.* 3, 780–791. doi: 10.1038/ismej.2009.29

Bongoua-Devisme, A. J., Cébron, A., Kassin, K. E., and Berthelin, J. (2013). Microbial communities involved in Fe reduction and mobility during soil organic matter (SOM) mineralization in two contrasted paddy soils. *Geomicrobiol. J.* 30, 347–361. doi: 10.1080/01490451.2012.688928

Breidenbach, B., Pump, J., and Dumont, M. G. (2015). Microbial community structure in the Rhizosphere of rice plants. *Front. Microbiol.* 6:1537. doi: 10.3389/fmicb.2015.01537

Campbell, B. J., Yu, L., Heidelberg, J. F., and Kirchman, D. L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12776–12781. doi: 10.1073/pnas.1101405108

Ding, L., Su, J., Xu, H., Jia, Z., and Zhu, Y. (2015). Long-term nitrogen fertilization of paddy soil shifts iron-reducing microbial community revealed by RNA-13C-acetate probing coupled with pyrosequencing. *ISME J.* 9, 721–734. doi: 10.1038/ismej.2014.159

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832

Gao, D. W., and Tao, Y. (2012). Current molecular biologic techniques for characterizing environmental microbial community. *Front. Environ. Sci. Eng.* 1, 82–97. doi: 10.1007/s11783-011-0306-6

Haller, L., Tonolla, M., Zopfi, J., Peduzzi, R., Wildi, W., and Poté, J. (2011). Composition of bacterial and archaeal communities in freshwater sediments with different contamination levels (Lake Geneva, Switzerland). *Water Res.* 45, 1213–1228. doi: 10.1016/j.watres.2010.11.018

Hassan, Z., Sultana, M., Westerhoff, H. V., Khan, S. I., and Röling, W. F. M. (2015). Iron cycling potentials of arsenic contaminated groundwater in Bangladesh as revealed by enrichment cultivation. *Geomicrobiol. J.* 33, 779–792. doi: 10.1080/01490451.2015.1111471

Herrero, M., and Stuckey, D. C. (2015). Bioaugmentation and its application in wastewater treatment: a review. *Chemosphere* 140, 119–128. doi: 10.1016/j.chemosphere.2014.10.033

Hilu, S., and Hersey, I. (2005). *Bogolanfini Mud Cloth*. Atglen, PA: Schiffer Publishing Ltd.

Lentini, C. J., Wankel, S. D., and Hansel, C. M. (2012). Enriched iron(III)-reducing bacterial communities are shaped by carbon substrate and iron oxide mineralogy. *Front. Microbiol.* 3:404. doi: 10.3389/fmib.2012.00404

Li, H., Peng, J., Weber, K. A., and Zhu, Y. (2011). Phylogenetic diversity of Fe(III)-reducing microorganisms in rice paddy soil: enrichment cultures with different short-chain fatty acids as electron donors. *J. Soils Sediments* 11, 1234–1242. doi: 10.1007/s11368-011-0371-2

Li, Y., Yu, S., Strong, J., and Wang, H. (2012). Are the biogeochemical cycles of carbon, nitrogen, sulfur, and phosphorus driven by the "Fe^{III}-Fe^{II} redox wheel" in dynamic redox environments? *J. Soils Sed.* 12, 683–693. doi: 10.1007/s11368-012-0507-z

Lovley, D. R. (2006). *Dissimilatory Fe(III)- and Mn(IV)-Reducing Prokaryotes.* (New York, NY: Springer), 635–658. doi: 10.1007/0-387-30742-7_21

Lovley, D. R., and Phillips, E. J. P. (1987). Rapid assay for microbially reducible ferric iron in aquatic sediments. *Appl. Environ. Microbiol.* 53, 1536–1540.

Martínez, J. A., Melgosa, M., Pérez, M., and Hita, E. (2001). Visual and instrumental color evaluation in red wines. *Food Sci. Technol. Int.* 7, 439–444. doi: 10.1106/VFAT-5REN-1WK2-5JGQ

Naganuma, T., Sato, M., Hoshii, D., Amano-Murakami, Y., Iwatsuki, T., and Mandemack, K. W. (2006). Isolation and characterization of *Pseudomonas* strains capable of Fe(III) reduction with reference to redox response regulator genes. *Geomicrobiol. J.* 23, 145–155. doi: 10.1080/01490450600596565

Nevin, K. P., and Lovley, D. R. (2002). Mechanisms for accessing insoluble Fe(III) oxide during dissimilatory Fe(III) reduction by *Geothrix fermentans*. *Appl. Environ. Microbiol.* 68, 2294–2299. doi: 10.1128/AEM.68.5.2294-2299.2002

Pan, Y., Yang, X., Chen, X., Xu, M., and Sun, G. (2016). The right mud: studies in the mud-coating technique of Gambiered Guangdong silk. *Appl. Clay Sci.* 135, 516–520. doi: 10.1016/j.clay.2016.09.024

Pedrós-Alió, C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol.* 14, 257–263. doi: 10.1016/j.tim.2006.04.007

Peng, Q., Shaaban, M., Wu, Y., Hu, R., Wang, B., and Wang, J. (2016). The diversity of iron reducing bacteria communities in subtropical paddy soils of China. *Appl. Soil Ecol.* 101, 20–27. doi: 10.1016/j.apsoil.2016.01.012

Petrie, L., North, N. N., Dollhopf, S. L., Balkwill, D. L., and Kostka, J. E. (2003). Enumeration and characterization of iron (III)-reducing microbial communities from acidic subsurface sediments contaminated with uranium (VI). *Appl. Environ. Microbiol.* 69, 7467–7479. doi: 10.1128/AEM.69.12.7467-7479.2003

Pylro, V. S., Roesch, L. F. W., Morais, D. K., Clark, L. M., Hirsch, P. R., and Tótola, M. R. (2014). Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *J. Microbiol. Methods* 107, 30–37. doi: 10.1016/j.mimet.2014.08.018

Qu, Y., Zhang, X., Ma, Q., Deng, J., Deng, Y., Van Nostrand, J. D., et al. (2015). Microbial community dynamics and activity link to indigo production from indole in bioaugmented activated sludge systems. *PLoS ONE* 10:e0138455. doi: 10.1371/journal.pone.0138455

Sanders, M. E. (2016). Probiotics and microbiota composition. *BMC Med.* 14:82. doi: 10.1186/s12916-016-0629-z

Si, Y., Zou, Y., Liu, X., Si, X., and Mao, J. (2015). Mercury methylation coupled to iron reduction by dissimilatory iron-reducing bacteria. *Chemosphere* 122, 206–212. doi: 10.1016/j.chemosphere.2014.11.054

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U.S.A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103

Storey, S., Chualain, D. N., Doyle, O., Clipson, N., and Doyle, E. (2015). Comparison of bacterial succession in green waste composts amended with inorganic fertilizer and wastewater treatment plant sludge. *Bioresour. Technol.* 179, 71–77. doi: 10.1016/j.biortech.2014.11.107

Thompson, I. P., van der Gast, C. J., Ciric, I., and Singer, A. C. (2005). Bioaugmentation for bioremediation: the challenge of strain selection. *Environ. Microbiol.* 7, 909–915. doi: 10.1111/j.1462-2920.2005.00804.x

Van Olphen, H. (1966). Maya blue: a clay-organic pigment? *Science* 154, 645–646. doi: 10.1126/science.154.3749.645

Wakimoto, R., Tazaki, K., Nawatani, N., Ikeda, Y., Imai, S., Sato, K., et al. (2004). Characterization of muddy clays for "Oshima Tsumugi" using traditional dye method. *Earth Sci.* 58, 199–214.

Wang, W., Qu, D., and Xu, J. (2008). Isolation of iron-reducing bacteria in paddy soil and its Fe(III) reduction potential analysis. *Nat. Sci. Educ.* 36, 103–107.

Xun, W., Huang, T., Zhao, J., Ran, W., Wang, B., Shen, Q., et al. (2015). Environmental conditions rather than microbial consortium composition determine the bacterial composition, microbial biomass and enzymatic activity of reconstructed soil microbial communities. *Soil Biol. Biochem.* 90, 10–18. doi: 10.1016/j.soilbio.2015.07.018

Zhang, T., Tremblay, P. L., Chaurasia, A. K., Smith, J. A., Bain, T. S., and Lovely, D. R. (2013). Anaerobic benzene oxidation via phenol in *Geobacter metallireducens*. *Appl. Environ. Microbiol.* 79, 7800–7806. doi: 10.1128/AEM.03134-13

Zhao, X., Yang, J., Bai, S., Ma, F., and Wang, L. (2016). Microbial population dynamics in response to bioaugmentation in a constructed wetland system under 10°C. *Bioresour. Technol.* 205, 166–173. doi: 10.1016/j.biortech.2016.01.043

Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J. D., et al. (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* 111, 836–845. doi: 10.1073/pnas.1324044111

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Pan, Yang, Xu and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 September 2017
doi: 10.3389/fmicb.2017.01712

Enhancing Nitrate Removal from Freshwater Pond by Regulating Carbon/Nitrogen Ratio

Rong Chen[1,2], Min Deng[1], Xugang He[1,3,4] and Jie Hou[1]**

[1] College of Fisheries, Huazhong Agricultural University, Wuhan, China, [2] School of Environmental Studies, China University of Geosciences, Wuhan, China, [3] Freshwater Aquaculture Collaborative Innovation Center of Hubei Province, Wuhan, China, [4] Hubei Provincial Engineering Laboratory for Pond Aquaculture, Wuhan, China

Nitrogen accumulation is a serious environmental problem in freshwater ponds, which can lead to massive death of fish and shrimps as well as the eutrophication. The removal of nitrate by regulating the carbon to nitrogen (C/N) ratio and the underlying mechanisms were investigated. The nitrate removal system comprised 530-mL medium containing 5 mg/L NO_3^--N and 0–66.6 mg/L COD (i.e., C/N ratio of 0–13.3) and 20 g ponds sediments. When the C/N ratio was higher than 8, the nitrate removal efficiency nearly reached 100% during the incubation period and the accumulation of nitrite was negligible. When the C/N ratio was below 8, the nitrate removal efficiency was lower and significant nitrite accumulation occurred. The nitrate removal rate increased with the C/N ratio increased, which was ascribed to the increase in the absolute abundance of denitrifiers (*nirS*, *nirK*, and *nosZ*). Although both *nirS*-type and *nirK*-type denitrifiers were found in the sediments of freshwater pond, *nirS*-type denitrifiers were predominant. *Dechloromonas* was the major *nirS*-type denitrifier for nitrate removal in *nirS*-type with the C/N ratios above 5.33, while the majority of the *nirK*-type denitrifiers were unclassified. Thus, this study implied that the appropriate C/N ratio played an important role on the removal of excess nitrate from freshwater ponds.

Keywords: C/N ratio, denitrification, functional gene, community structure, freshwater ponds

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Yiyong Zhou, Institute of Hydrobiology (CAS), China
Ivan Iliev, Plovdiv University "Paisii Hilendarski", Bulgaria

****Correspondence:***
Xugang He
xgh@mail.hzau.edu.cn
Jie Hou
254550384@qq.com

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *13 April 2017*
Accepted: *24 August 2017*
Published: *08 September 2017*

Citation:
Chen R, Deng M, He X and Hou J (2017) Enhancing Nitrate Removal from Freshwater Pond by Regulating Carbon/Nitrogen Ratio. Front. Microbiol. 8:1712. doi: 10.3389/fmicb.2017.01712

INTRODUCTION

Freshwater ponds provide abundant food resources to humans, especially in China, supplying nearly 20 million tons of fishery production every year. Recently, high density and intensive farming models with superfluous feeding and fertilization have been used to achieve higher economic efficiency, resulting in nitrogen accumulation owing to excessive residual feed and excrement (Avnimelech and Ritvo, 2003). Although nitrogen is an important nutrient for aquatic organisms such as fish, excess nitrogen in freshwater ponds leads to the massive death of fish and shrimps, and the eutrophication (Camargo and Alonso, 2006). For instance, nitrite (NO_2^-), a reductive product of nitrate (NO_3^-), is toxic to aquatic organisms due to the damage to hemoglobin (Camargo and Alonso, 2006). In addition, contaminated pond water also affects the water quality of the surrounding larger water bodies such as lakes (Cao et al., 2007). In ponds, nitrate, organic nitrogen, and ammonium nitrogen (NH_4^+-N) are the predominant nitrogen species. While nitrification couples the conversion of NH_4^+-N to nitrate (Wu et al., 2009), denitrification reduces nitrate to nitrogen (Kraft et al., 2014), further releasing nitrogen into the

atmosphere. Aquatic animals require feed with a high protein concentration (Hari et al., 2006) of up to 50% (Wei et al., 2016). About 75% of nitrogen in feed ends up in water through ammonification of uneaten feed and excretion (Crab et al., 2012). The high concentration of nitrogen in water limits the transformation capacity of denitrification in natural ponds because denitrifiers are mainly heterotrophic microorganisms requiring high carbon content for their growth (Crab et al., 2012). The excessive accumulation of nitrogen will spoil the living environment of the aquatic animals (Hari et al., 2006). Hence, it is a major environmental concern to develop cost-effective processes for controlling nitrogen in contaminated freshwater ponds.

Previous studies have shown that ion exchange, reverse osmosis, electrochemical processes, and biological treatment can effectively remove nitrate from wastewater (Shrimali and Singh, 2001; Koparal and Ogutveren, 2002; Mohan et al., 2016). However, the regeneration of anion exchange resin requires a high amount of regenerant for ion exchange, the discharge of brine and high concentrations of nitrate during reverse osmosis may lead to secondary pollution, and the cost of power for electrochemical processes is very high. Consequently, biological treatment has attracted increasing attention in recent years owing to its higher efficiency and lower cost (Shrimali and Singh, 2001; Mohan et al., 2016). Constructed wetlands and sequencing batch reactors (SBRs) are commonly used for nitrate removal from wastewater (Wu et al., 2009; Mohan et al., 2016) through denitrification to achieve the conversion of nitrate to nitrogen. The redox environment often occurs in anoxic conditions at the bottom of freshwater ponds, which is beneficial for the removal of nitrate by enhancing the denitrification efficiency. Although wastewater treatment plants (WWTPs) have exhibited ideal nitrate removal efficiency through denitrification and their underlying mechanisms have been extensively explored (Zhao et al., 2013; Mannina et al., 2016), the microbial communities in activated sludge applied to WWTPs are different from those in freshwater ponds. As a result, the denitrification mechanism in freshwater ponds may differ from that in WWTPs. In addition, WWTPs just require the absence of nitrite in the effluent water, and nitrite accumulation during the intermediate process can be ignored. As nitrite is toxic to aquatic organisms (Camargo and Alonso, 2006), its accumulation cannot occur during the entire process of denitrification in freshwater ponds.

Denitrification, an important nitrogen removal mechanism that can ameliorate the effects of nitrogen pollution via conversion of nitrate to nitrogenous gas (Kraft et al., 2014; Ward and Jensen, 2014; Morrissey and Franklin, 2015), is the major pathway for the removal of nitrogen from water bodies (Altabet et al., 1995; Hargreaves, 1998; Laverman et al., 2010). Previous studies have shown that several factors, including temperature, pH, dissolved oxygen (DO) level, organic carbon species, and ratio of organic carbon to nitrogen (C/N), can affect the denitrification efficiency (Strong et al., 2011; Kraft et al., 2014). In freshwater ponds, the temperature varies along with ambient environment and the pH was maintained in the range of 6–8, and these factors are difficult to adjust by manual operation. In contrast, the type of organic matter and C/N ratio can be easily regulated by daily feed. The C/N ratio has been identified as a key environmental factor that determines the products of nitrate reduction (Kraft et al., 2014). Therefore, it is rational to hypothesize that we can obtain higher denitrification efficiency by regulating the C/N ratio through the addition of extra organic carbon.

In the present study, the effect and extent of the impact of different C/N ratios on the nitrate removal efficiency in ponds sediments were investigated. For this purpose, denitrifiers from the sediments of freshwater ponds were cultured in nutrient medium with different C/N ratios for 30 days, and the variation in nitrate content during the incubation period was explored. Three functional genes, *nirS*, *nirK*, and *nosZ*, were employed to measure the abundance and community structure (without *nosZ*) of the denitrifiers by using real-time quantitative polymerase chain reaction (qPCR) technology and high-throughput sequencing technology, respectively, at different time points during the incubation period. The results obtained could help in understanding the role of C/N ratio in nitrate removal and the variations in the abundance and community structure of denitrifiers, facilitating the development of efficient technology for nitrogen removal from freshwater ponds.

MATERIALS AND METHODS

Chemicals and Sediments

KCl, NaH_2PO_4, Na_2HPO_4, CH_3COONa (NaAc), and KNO_3 were above analytical grade and purchased from Sinopharm Chemical Reagent Co., Ltd, China. The sediment samples were collected from a fishpond located in Gong'an, Hubei province, China (112°15′44.63″E, 29°55′14.62″N). About 5 kg of sediment was collected at 0.2 m below the surface of the pond using a UWITEC Sediment corer 60 (Mondsee, Austria) on April 12, 2014, and transported to the laboratory at 4°C. Then, the sediment was washed with 1 mol/L KCl solution to decrease the ammonia concentration to below 0.5 mg/L, washed with 1 mol/L phosphate-buffered saline (PBS, pH = 7.0) to remove potassium ions, and centrifuged at 6000 *g* for 10 min.

Batch Experiments

All the experiments were conducted at 25 ± 2°C in 550-mL glass bottle reactors wrapped by aluminum foils to avoid the light. Every glass reactor was sealed by using a suitable bottle cap with a seal ring to assure good sealing. Prior to the experiment, 20 g of pretreated sediment samples were suspended in 530 mL of medium containing 20 mM PBS (pH 7.5) and 0.2 mg/L trace elements, including Cu, Fe, Mn, Ca, Mg, and Co (Bi et al., 2015). The suspension was purged with nitrogen gas (99.999%) for 30 min to remove oxygen. To start the experiment, medium containing 5 mg N/(L·d) with different concentrations of organic carbon [0, 13.35, 26.65, 40, and 66.65 mg COD/(L·d) supplied by NaAc] were added to the suspension. The hydraulic retention time (HRT) was 1 day (24 h) during the entire operation period (30 days). After addition of the influent, the

reactors were purged with nitrogen gas (99.999%) for 30 min. Each experiment was conducted in triplicate, and the control experiments were performed with sterile water and sediment samples.

To obtain an insight into the process of nitrate removal, the kinetics of nitrate reduction was determined. A total of 300 g of pretreated sediment were respectively added to two reactors (1100 mL) containing 950 mL of the medium [5 mg N/(L·d) and 40 mg COD/(L·d)]. On day 15, the sediment samples were collected from the two reactors and centrifuged at 6000 *g* for 10 min. Subsequently, 20 g of the centrifuged sediment were added to the reactor containing 530 mL of the medium (5 mg/L NO_3^--N with different concentrations of organic carbon 0, 13.35, 26.65, 40, and 66.65 mg/L COD, supplied by NaAc). The medium in each reactor was bubbled with nitrogen gas (99.999%) for 30 min after sampling. On day 30, the sediment samples were collected from the two reactors and conducted with the same treatment as on day 15. Each experiment was performed in triplicate, and the control comprised sterile water and sediment samples. At predetermined time points (days 0, 15, and 30), about 8 mL of the medium were filtered through a 0.45-μm membrane, centrifuged at 6000 *g* for 10 min, and about 2 g of the sediment were collected and stored at −80°C for DNA extraction.

DNA Extraction and qPCR

Soil DNA kits D5625-01 (Omega, United States) were used to extract and purify the total genomic DNA from the samples. The extracted genomic DNA was detected by 1% agarose gel electrophoresis and stored at −20°C until further use. The target fragments of *nirK*, *nirS*, and *nosZ* were subjected to PCR, and all the primers used were synthesized by TSINGKE Biotechnology Co. (Wuhan, China) and diluted to a concentration of 10 mmol/L (Supplementary Table S1). The PCR products were cloned into the pMD18-T Easy Vector (Takara, Dalian, China). The plasmids containing specific functional genes (i.e., *nirK*, *nirS*, and *nosZ*) were obtained from TSINGKE Biotechnology Co. (Beijing, China). The standard samples were diluted to yield a series of 10-fold concentrations and subsequently used for constructing qPCR standard curves. The R^2 value for each standard curve exceeded 0.99, indicating good linear relationships over the concentration ranges used in this study. The amplification efficiency for each standard curve was between 98 and 102%.

qPCR was performed on a Qiagen Q thermocycler (Qiagen, Germany) with 20-μL reaction mixture containing 10 μL of SYBR Green II PCR master mix (Takara), 1 μL of template DNA (sample DNA or plasmid DNA for standard curves), forward and reverse primers (Supplementary Table S2), and sterile water (Millipore, United States). The reaction was performed using a three-step thermal cycling procedure, and the protocol and parameters for each target gene are presented in Supplementary Table S2. Each qPCR comprised 40 cycles, followed by a melting curve analysis. All the measurements were performed in triplicate. Sterile water was used as a negative control, and the qPCR data were normalized to copies/g dry sediment.

Illumina MiSeq Sequencing and Data Analysis

PCR and sequencing were conducted as described in a previous study (Caporaso et al., 2012) with primers nirScd3aF–nirSR3cd and nirKFlaCu–nirKR3Cu for amplifying *nirS* and *nirK*, respectively. Sequencing was performed using MiSeq Benchtop Sequencer (Illumina, United States) by Shanghai Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China), and the sequencing data were analyzed using Mothur software (Schloss et al., 2009).

Analysis

Every 5 days, the DO level and pH of the influent and effluent were measured by using DO 2000 LDOTM (Thermo Eberline Trading GmbH, Wermelskirchen, Germany) and HI-9025 pH meter (Hanna, Padova, Italy), respectively. The DO level and pH of the influent and effluent ranged between 0.32 and 0.48 mg/L and between 7.18 and 7.46, respectively. The NH_4^+-N concentration in the influent and effluent ranged from 0.28 to 0.47 mg/L.

The concentrations of NH_4^+-N, nitrite-nitrogen (NO_2^--N), and nitrate-nitrogen (NO_3^--N) were determined by using NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific, New York, United States) according to standard analytical procedures (Water Environment Federation, 2005). Specifically, the concentrations of NH_4^+-N, NO_2^--N, and NO_3^--N were measured spectrophotometry using Nessler's reagent, *N*-(1-naphthyl)ethylenediamine dihydrochloride, and UV spectrophotometry using hydrochloric acid, respectively. The concentration of nitrogen species in the influent and effluent and HRT (24 h) were used to calculate the NO_2^--N removal efficiency and accumulation rate. Correlation coefficients were calculated to evaluate the associations between C/N ratio and nitrate removal efficiency, accumulation of NO_2^--N, and nitrogen transformation genes. The C/N ratio was referred to the adding COD to the nitrate.

The denitrification kinetics was determined with the first-order model as follows: $S = S_0 \cdot e^{-k.t}$ (Simkins and Alexander, 1984; Tiemeyer et al., 2007), where k is the denitrification rate constant [mg/(L·h)], t is the reaction time (h), and S_0 and S (mg/L) are the concentrations of NO_3^--N at reaction time 0 and t, respectively.

Nucleotide Sequence Accession Numbers

The nucleotide sequences obtained in this study were deposited in the GenBank database under accession nos. KR232567–KR232570 for *nirS* and KP262401 for *nirK*.

RESULTS AND DISCUSSION

Effect of C/N Ratio on Nitrate Removal

In the absence of extra organic carbon (C/N ratio = 0), the removal efficiency of nitrate increased from 18.1 to 42.6% in 30 days (**Figure 1A**). The decrease in nitrate during the incubation

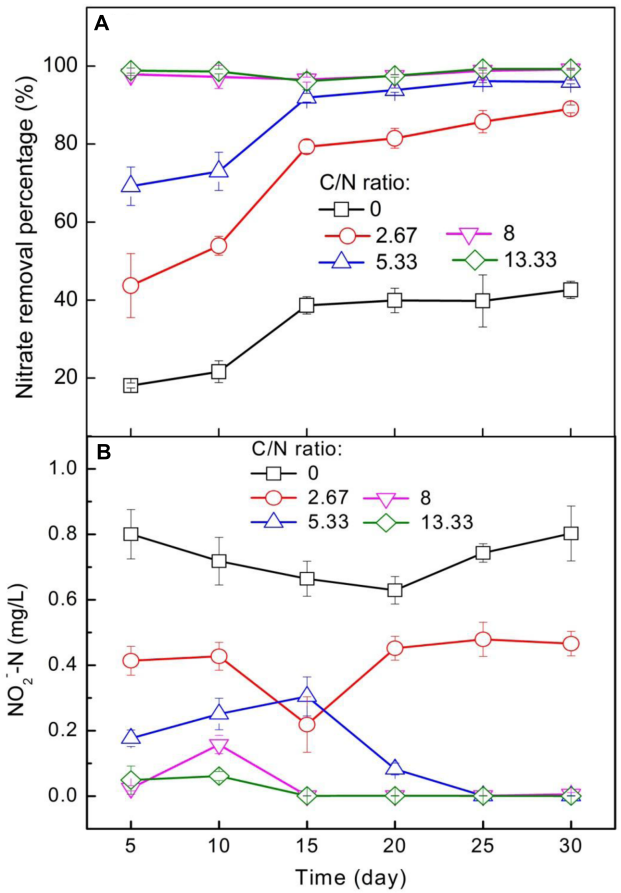


FIGURE 1 | Effects of C/N ratio on **(A)** nitrate removal efficiency and **(B)** accumulation of nitrite.

period may be due to the denitrification using pristine organic carbon in the sediments. Interestingly, a remarkable increase in nitrate removal efficiency was observed with the addition of extra organic carbon (**Figure 1A**). When the C/N ratio was less than 8, the nitrate removal efficiency increased from 43.7 to 69.2% on day 5 and reached 89 and 96% on day 30 with the increase in C/N ratio from 2.67 to 5.33. However, nitrate could not be efficiently removed when the C/N ratio was 0–2.67, which was reasonably lower than the theoretically as well as experimentally determined value of 2.86 and 3.5–4.5 for complete denitrification, respectively (Henze et al., 1994). When the C/N ratio was higher than 8, the nitrate removal efficiency nearly reached 100% during the entire incubation period, which may be owing to the presence of adequate organic carbon for denitrification (Cervantes et al., 2001). Thus, the increase in nitrate removal efficiency with the increase in C/N ratio confirmed the hypothesis that nitrate removal is mediated by C/N ratio. Furthermore, correlation analysis showed a positive correlation between C/N ratio and average nitrate removal efficiency ($r = 0.919$, $P < 0.01$) on day 15 (Supplementary Table S3), and similar results were also found a tother time points with correlation coefficients above 0.7 and a significance level below 0.01 (Supplementary Table S3). These results further support the association between nitrate removal and C/N ratio.

As NO_2^- is the initial product of denitrification and is toxic to aquatic organisms, the variation of NO_2^- was measured during the incubation period (**Figure 1B**). In the absence of extra organic carbon (C/N ratio = 0), significant accumulation of NO_2^--N was observed within 30 days. The concentration of NO_2^--N was nearly stabilized at 0.8 mg/L, amounting to 16% of the initial NO_3^--N concentration (5 mg/L). The nitrate removal efficiency ranged from 18.1 to 42.6% at the C/N ratio of 0 (**Figure 1A**), suggesting that some nitrate was reduced to nitrite (**Figure 1B**). However, nitrite is toxic to denitrifiers (Tiemeyer et al., 2007; Wang Y. et al., 2014), and thus may further inhibit the denitrification step ($NO_2^- \rightarrow$ NO or N_2O) by decreasing the activity of related enzymes. With the increase in C/N ratio from 2.67 to 13.33, the concentration of NO_2^--N decreased from 0.41 to 0.05 on day 5 and from 0.47 to undetectable levels (the detection limit for NO_2^--N was 0.003 mg/L) on day 30. The notable decrease in NO_2^--N concentration at higher C/N ratios suggested that the addition of extra organic carbon is necessary for reducing the risk of nitrite poisoning in freshwater ponds. Correlation analysis showed a negative correlation between NO_2^--N and C/N ratio ($r > 0.8$, $P < 0.01$) during the incubation period (Supplementary Table S4). Previous studies had demonstrated that the formation and accumulation of nitrite were controlled by the C/N ratio (Hari et al., 2006; Zhi and Ji, 2014). For example, for the low-carbon-concentration treatment (37.7% carbohydrates), the concentrations of nitrite and nitrate were 6.5- and 6.9-fold higher than those in the high-carbon-concentration treatment (about 47.6% carbohydrates) (Wei et al., 2016). The decrease in the concentrations of nitrate and nitrate in the presence of high concentrations of carbon may be attributed to denitrification. Hari et al. (2006) and Avnimelech (1999) observed a similar effect of carbon concentration on the concentrations of nitrite and nitrate. The decrease in the concentrations of nitrite and nitrate is beneficial to aquatic animals because toxic nitrite can lead to low survival rate or decreased growth (Wei et al., 2016).

To determine the optimal C/N ratio, the nitrate removal efficiency and cost should be comprehensively considered. A high C/N ratio could cause energy waste and lead to ammonification (Kraft et al., 2014), whereas a low C/N ratio could result in deficiency in nitrate removal and accumulation of nitrite. In the present study, the complete removal of nitrate and negligible accumulation of nitrite were achieved at a C/N ratio of 8–13.33, revealing that a C/N ratio above 8 is required to avoid poisoning of aquatic organisms by achieving complete denitrification (without NO_2^--N accumulation). The results of this study are similar to those reported in previous studies, which achieved optimal nitrogen removal (in SBR system) at a C/N ratio of 11.1 and 11.2 (Münch et al., 1996; Chiu et al., 2007).

Kinetics of Nitrate Removal at Different C/N Ratios

To get further insight into the influence of C/N ratio on nitrate removal, the kinetics of nitrate removal on days 15 and 30 was

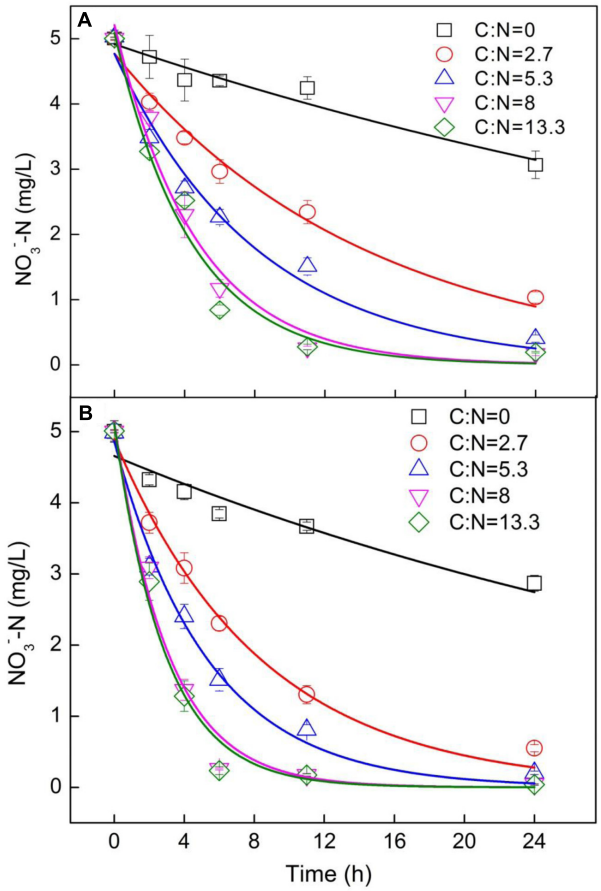


FIGURE 2 | Kinetic models of nitrate removal with time on **(A)** day 15 and **(B)** day 30. A first-order kinetic model was used to fit the decrease in nitrate with time.

explored (**Figure 2**). On day 15, the concentration of NO_3^--N decreased from 5 to 0.2 mg/L in 24 h with the increase in C/N ratio from 0 to 13 (**Figure 2A**). The decrease in NO_3^--N followed first-order kinetics, with rate constants of 0.02, 0.07, 0.12, 0.21, and 0.23 mg/(L·h) for systems with C/N ratios of 0, 2.67, 5.33, 8, and 13.3, respectively (**Table 1**). The positive correlation of denitrification rate constants with C/N ratio ($r = 0.96$, $P = 0.01$, Supplementary Figure S1a; $r = 0.94$, $P = 0.02$, Supplementary Figure S1b) further confirmed that denitrification is dependent on C/N ratio. On day 30, the rate constants were 0.022, 0.12, 0.19, 0.32, and 0.34 mg/(L·h) for systems with C/N ratios of 0, 2.67, 5.33, 8, and 13.3, respectively (**Table 1**), which were higher than those noted on day 15. The increase in the denitrification rate with the incubation was presumably due to the increase in the population of denitrifiers with the incubation time. Previous studies have shown that the increase in C/N ratio can stimulate bacterial growth (Inwood et al., 2007), subsequently improving the denitrification efficiency per cell. Accordingly, in the present study, the variations in the abundance of functional genes, diversity, and community of denitrifiers were examined.

TABLE 1 | Kinetic parameters of the different nitrate concentration.

Group	15th day		30th day	
	k [mg/(L·h)]	R^2	*k* [mg/(L·h)]	R^2
C:N = 0	0.019 ± 0.0024	0.94	0.022 ± 0.0037	0.89
C:N = 2.67	0.070 ± 0.0062	0.98	0.12 ± 0.0084	0.99
C:N = 5.33	0.12 ± 0.014	0.97	0.19 ± 0.014	0.99
C:N = 8	0.21 ± 0.024	0.98	0.32 ± 0.040	0.97
C:N = 13.33	0.23 ± 0.030	0.97	0.34 ± 0.036	0.98

Effect of C/N Ratio on the Absolute Abundance of Functional Genes

Nitrite reduction, the symbolic and key step of denitrification, is catalyzed by nitrite reductases, including copper-containing and multiheme enzymes encoded by the genes *nirK* and *nirS*, respectively (Stouthamer, 1992; Fan et al., 2015). As nitrous oxide reduction is the last step of denitrification, the related functional gene *nosZ* is usually utilized to estimate complete denitrification (Kandeler et al., 2006; Bowles et al., 2012). In the present study, to illustrate the influence of C/N ratio on growth of denitrifiers, the abundances of *nirK*, *nirS*, and *nosZ* were measured during the incubation period (**Figure 3**). In the absence of extra organic carbon (C/N ratio = 0), the negligible increase in the abundances of *nirK*, *nirS*, and *nosZ* implied slow growth of denitrifiers. In contrast, with the addition of extra organic carbon, the abundances of these functional genes observably increased with the incubation period. The abundances of *nirS*, *nirK*, and *nosZ* increased from 1.98×10^6 to 5.96×10^7 copies/g (**Figure 3A**), from 1.23×10^5 to 3.43×10^6 copies/g (**Figure 3B**), and from 3.30×10^6 to 5.12×10^7 copies/g (**Figure 3C**), respectively, indicating the growth of microbial population. The positive correlation between the abundances of *nirS*, *nirK*, and *nosZ* and C/N ratio increased on days 15 and 30 (**Table 2**), further affirming that the C/N ratio plays a critical role in the growth of denitrifiers.

At C/N ratios below 8, the absolute abundances of *nirS*, *nirK*, and *nosZ* significantly increased significantly with C/N ratio increased, suggesting that the supply of organic carbon is insufficient for the reproduction of denitrifiers. At C/N ratios above 8, negligible increase in the abundances of the three genes was noted, which indicated that the denitrifier population size may be limited by the electron acceptor (NO_3^--N). Besides, excessive organic carbon may be beneficial for other kinds of heterotrophic microorganisms (Kraft et al., 2014), inhibiting the growth of denitrifiers. Therefore, a C/N ratio of 8 was recommended.

During the incubation period, the absolute abundance of *nirK* was always less than that of *nirS*, which may be ascribed to the insufficient organic carbon or unsuitable NaAc for the growth of denitrifiers containing *nirK*. As the functions of *nirS* and *nirK* are similar (Stouthamer, 1992), denitrifiers containing *nirS* predominated in nitrate removal. The abundance of *nosZ* positively increased with the C/N ratio (**Table 2**), suggesting that organic carbon is important for denitrifiers containing *nosZ*. Inwood et al. (2007) confirmed that the increase in organic

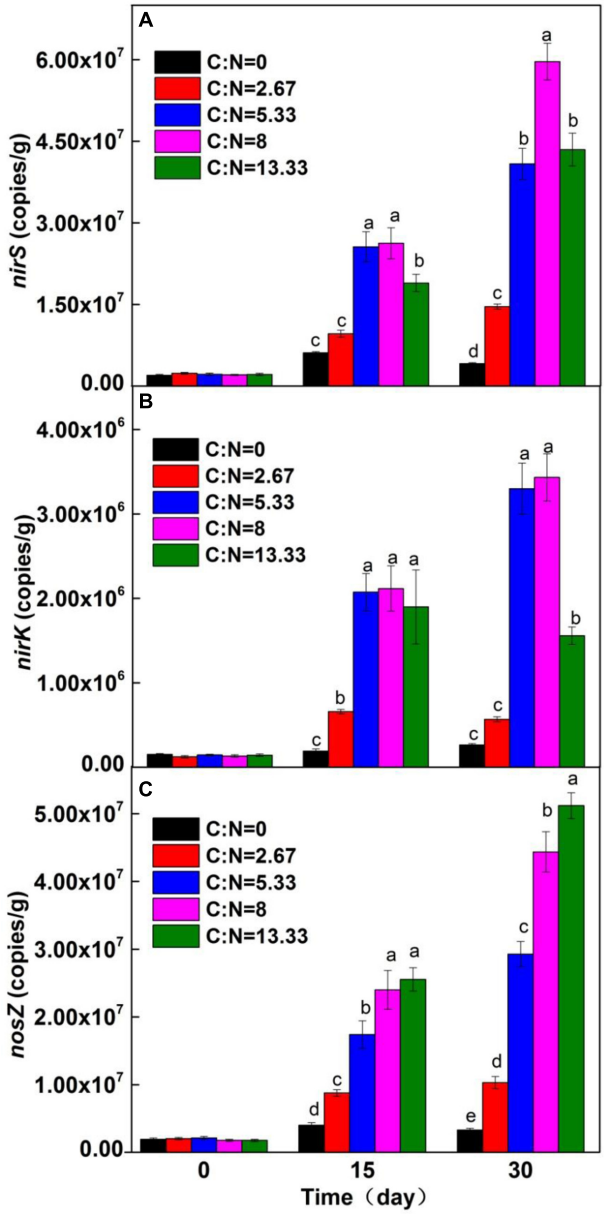


FIGURE 3 | Absolute abundance of functional genes: **(A)** *nirS*, **(B)** *nirK*, and **(C)** *nosZ*. Different letters (a–e) represent samples with a significant difference at the level of P=0.05 (one-way ANOVA).

carbon reduced the production of N_2O. In the present study, an obvious increasing trend of *nirS*, *nirK*, and *nosZ* abundances (**Figure 3** and **Table 2**) provided direct evidence of an ecological association and symbiotic relationship among the denitrifier communities at the molecular level (functional genes).

Effect of C/N Ratio on the Diversity and Composition of Denitrifiers

To investigate the effect of C/N ratio on the diversity and composition of denitrifiers, the Shannon index, Simpson index,

TABLE 2 | Correlation between gene abundance and C/N ratio.

Gene	15th day		30th day	
	Correlation coefficient	Significance	Correlation coefficient	Significance
nirS	0.61	0.02	0.77	0.00
nirK	0.76	0.00	0.46	0.08
nosZ	0.93	0.00	0.96	0.00

and sequence were examined. The results showed that the C/N ratio was not significantly correlated with the diversity index (including Shannon and Simpson indices) with respect to *nirS* (r = −0.26, 0.038; P = 0.672, 0.956, respectively), but a slight correlation with the diversity index with respect to *nirK* (r = 0.75, −0.80; P = 0.11, 0.10, respectively). This could possibly be owing to the higher C/N ratio being unfavorable for the growth of denitrifiers containing *nirS*, but favorable for the growth of denitrifiers containing *nirK*. Chen et al. (2010) noted that denitrifiers containing *nirK* were more sensitive to the C/N ratio than denitrifiers containing *nirS*. Furthermore, Yoshida et al. (2009) found that *nirS*-containing denitrifiers in rice field were not sensitive to C/N ratio. Besides, the diversity of denitrifiers containing *nirK* may be unimportant for the whole system owing to their low absolute abundance. Therefore, we presumed that the diversity of denitrifiers was not important for the apparent nitrate removal, but it is important for the stability of ecosystem.

The community composition, relative abundance, and molecular phylogenetic relationship of denitrifiers containing *nirS* were determined by hierarchical clustering of *nirS* (**Figure 4A**). The results revealed that Proteobacteria (46.00–90.28%) and β-Proteobacteria were dominant at the phylum and class level, respectively, similar to those observed in a previous study on groundwater (Chu and Wang, 2013). Eight known genera (i.e., *Dechloromonas*, *Azoarcus*, *Azospira*, *Rubrivivax*, *Thiobacillus*, *Vogesella*, and *Zoogloea*) were detected, among which *Dechloromonas* was predominant. At C/N ratios below 5.33, bacteria-unclassified (35.72–48.33%) and *Dechloromonas* (18.81–38.89%) were predominant, while at C/N ratios above 5.33, *Dechloromonas* (56.18–70.60%) was dominant. With further increases in the C/N ratio, the percentage of Rhodocyclaceae-unclassified increased. It has been reported that the majority of *Dechloromonas* could reduce nitrate by utilizing ferrous iron or organic carbon as the electron donor and without nitrite accumulation (Coates and Achenbach, 2004). Moreover, studies on denitrifier community in activated sludge indicated that Rhodocyclaceae could easily utilize organic carbon for efficient denitrification (Khan et al., 2002; Ginige et al., 2005). In addition, the presence of *Rubrivivax*, which lacks nitrate reductase and needs external nitrite for denitrification, could mitigate the accumulation of nitrite (Saarenheimo et al., 2015).

By using hierarchical clustering of *nirK*, the community compositions, relative abundance, and molecular phylogenetic relationship of denitrifiers containing *nirK* were explored (**Figure 4B**). At the phylum level, bacteria-unclassified (67.96–93.82%) was dominant, similar to that reported in a previous

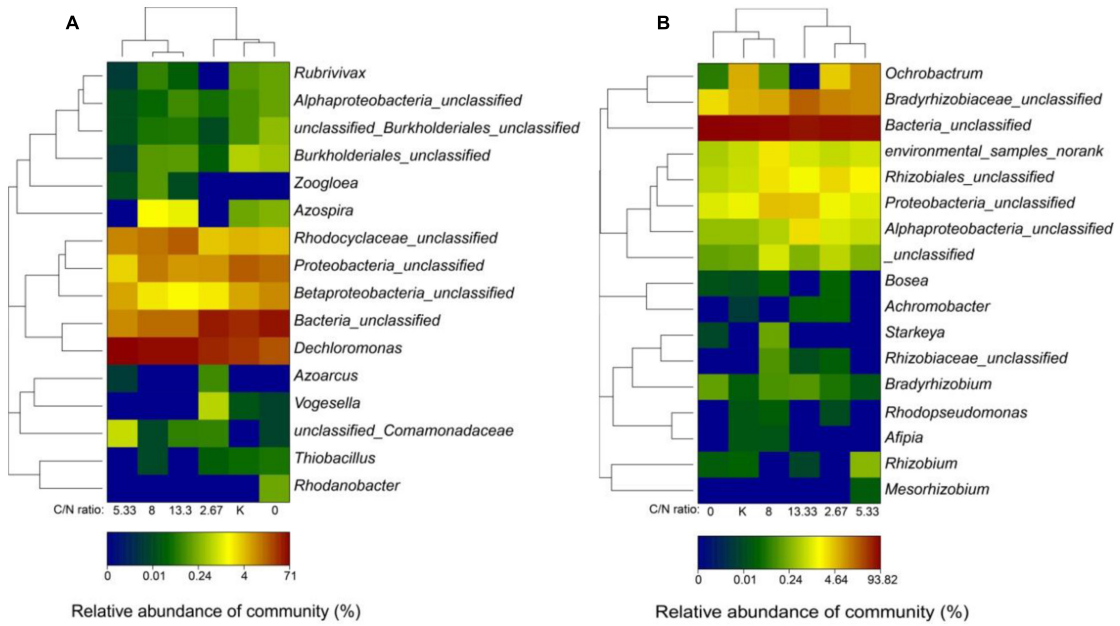


FIGURE 4 | Hierarchical cluster analysis of the communities of **(A)** *nirS*-type and **(B)** *nirK*-type denitrifiers. The *y*-axis indicates clustering of 300 most abundant OTUs (3% distance) in the reads. The OTUs were arranged based on genus. Sample communities were clustered based on the complete linkage method. The color intensity of the scale indicates the relative abundance of each OTU read. Relative abundance was defined as the number of sequences affiliated with that OTU divided by the total number of sequences per sample. *K* represents control samples (incubated with distilled water).

research on soil (Bremer et al., 2007). At the genus level, nine known genera (namely, *Achromobacter*, *Afipia*, *Bosea*, *Bradyrhizobium*, *Mesorhizobium*, *Ochrobactrum*, *Rhizobium*, *Rhodopseudomonas*, and *Starkeya*) were detected. The proportion of genera affiliated with Bradyrhizobiaceae-unclassified was obviously higher in the system with a C/N ratio of 13.33 (21.45%) than that in the other groups (3.19–12.50%), which was owing to sufficient electron donor (Guo et al., 2011; Wang Z. et al., 2014). At C/N ratios below 5.33, the percentage of *Ochrobactrum* increased with increasing C/N ratio ($r = 0.988$, $P = 0.099$); however, at C/N ratios above 5.33, the percentage of *Ochrobactrum* decreased with the increase in C/N ratio ($r = -0.762$, $P = 0.448$). Summarily, Proteobacteria and *Dechloromonas* were predominant at the phylum and genus level, respectively, and *Dechloromonas* may play a dominant role in nitrate removal.

CONCLUSION

This study investigated the influence of C/N ratio on nitrate removal from the sediments of freshwater ponds. The nitrate removal efficiency increased with increases in the C/N ratio. When the C/N ratio was higher than 8, adequate denitrification without NO_2^--N accumulation was achieved. When the C/N ratio was less than 8, excessive nitrite accumulation occurred. Hence, a C/N ratio of 8 is recommended for nitrate removal from pond sediments. When the C/N ratio was increased from 0 to 13.3, the absolute abundance of *nirS*, *nirK*, and *nosZ* genes correspondingly increased. The positive correction between nitrate removal rate and nitrogen functional genes further confirmed that different nitrogen transformation processes were coupled to synergistically contribute to nitrogen removal at the molecular level (functional genes). *Dechloromonas* dominated the overall process of nitrate removal in systems with a C/N ratio above 5.33, while the majority of *nirK*-type denitrifiers were unclassified.

AUTHOR CONTRIBUTIONS

RC designed and conducted the experiment, accomplished the first draft, and corrected the article. MD and JH provided some valuable suggestion for this article. XH proposed the topic and corrected the manuscript.

FUNDING

The authors want to thank the State Science and Technology Support Program of China (2012BAD25B06), the earmarked fund for China Agriculture Research System (CARS-46), and Agro-scientific Research in the Public Interest Project (No. 201203083) for financially supporting this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01712/full#supplementary-material

REFERENCES

Altabet, M. A., Francois, R., Murray, D. W., and Prell, W. L. (1995). Climate-related variations in denitrification in the Arabian Sea from sediment 15N/14N ratios. *Nature* 373, 506–509. doi: 10.1038/373506a0

Avnimelech, Y. (1999). Carbon nitrogen ratio as a control element in aquaculture systems. *Aquaculture* 176, 227–235. doi: 10.1016/s0044-8486(99)00085-x

Avnimelech, Y., and Ritvo, G. (2003). Shrimp and fish pond soils: processes and management. *Aquaculture* 220, 549–567. doi: 10.1016/S0044-8486(02)00641-5

Bi, Z., Takekawa, M., Park, G., Soda, S., Zhou, J., Qiao, S., et al. (2015). Effects of the C/N ratio and bacterial populations on nitrogen removal in the simultaneous anammox and heterotrophic denitrification process: mathematic modeling and batch experiments. *Chem. Eng. J.* 280, 606–613. doi: 10.1016/j.cej.2015.06.028

Bowles, M. W., Nigro, L. M., Teske, A. P., and Joye, S. B. (2012). Denitrification and environmental factors influencing nitrate removal in Guaymas Basin hydrothermally altered sediments. *Front. Microbiol.* 3:377. doi: 10.3389/fmicb.2012.00377

Bremer, C., Braker, G., Matthies, D., Reuter, A., Engels, C., and Conrad, R. (2007). Impact of plant functional group, plant species, and sampling time on the composition of *nirK*-type denitrifier communities in soil. *Appl. Environ. Microbiol.* 73, 6876–6884. doi: 10.1128/aem.01536-07

Camargo, J. A., and Alonso, A. (2006). Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: a global assessment. *Environ. Int.* 32, 831–849. doi: 10.1016/j.envint.2006.05.002

Cao, L., Wang, W., Yang, Y., Yang, C., Yuan, Z., Xiong, S., et al. (2007). Environmental impact of aquaculture and countermeasures to aquaculture pollution in China. *Environ. Sci. Pollut. Res.* 14, 452–462. doi: 10.1065/espr2007.05.426

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Cervantes, F. J., David, A., and Gómez, J. (2001). Nitrogen removal from wastewaters at low C/N ratios with ammonium and acetate as electron donors. *Bioresour. Technol.* 79, 165–170. doi: 10.1016/S0960-8524(01)00046-3

Chen, Z., Luo, X., Hu, R., Wu, M., Wu, J., and Wei, W. (2010). Impact of long-term fertilization on the composition of denitrifier communities based on nitrite reductase analyses in a paddy soil. *Microb. Ecol.* 60, 850–861. doi: 10.1007/s00248-010-9700-z

Chiu, Y. C., Lee, L. L., Chang, C. N., and Chao, A. C. (2007). Control of carbon and ammonium ratio for simultaneous nitrification and denitrification in a sequencing batch bioreactor. *Int. Biodeterior. Biodegrad.* 59, 1–7. doi: 10.1016/j.ibiod.2006.08.001

Chu, L., and Wang, J. (2013). Denitrification performance and biofilm characteristics using biodegradable polymers PCL as carriers and carbon source. *Chemosphere* 91, 1310–1316. doi: 10.1016/j.chemosphere.2013.02.064

Coates, J. D., and Achenbach, L. A. (2004). Microbial perchlorate reduction: rocket-fuelled metabolism. *Nat. Rev. Microbiol.* 2, 569–580. doi: 10.1038/nrmicro926

Crab, R., Defoirdt, T., Bossier, P., and Verstraete, W. (2012). Biofloc technology in aquaculture: Beneficial effects and future challenges. *Aquaculture* 356, 351–356. doi: 10.1016/j.aquaculture.2012.04.046

Fan, H., Bolhuis, H., and Stal, L. J. (2015). Drivers of the dynamics of diazotrophs and denitrifiers in North Sea bottom waters and sediments. *Front. Microbiol.* 6:738. doi: 10.3389/fmicb.2015.00738

Ginige, M. P., Keller, J., and Blackall, L. L. (2005). Investigation of an acetate-fed denitrifying microbial community by stable isotope probing, full-cycle rRNA analysis, and fluorescent in situ hybridization-microautoradiography. *Appl. Environ. Microbiol.* 71, 8683–8691. doi: 10.1128/AEM.71.12.8683-8691.2005

Guo, G. X., Deng, H., Qiao, M., Mu, Y. J., and Zhu, Y. G. (2011). Effect of pyrene on denitrification activity and abundance and composition of denitrifying community in an agricultural soil. *Environ. Pollut.* 159, 1886–1895. doi: 10.1016/j.envpol.2011.03.035

Hargreaves, J. A. (1998). Nitrogen biogeochemistry of aquaculture ponds. *Aquaculture* 166, 181–212. doi: 10.1016/s0044-8486(98)00298-1

Hari, B., Kurup, B. M., Varghese, J. T., Schrama, J. W., and Verdegem, M. C. J. (2006). The effect of carbohydrate addition on water quality and the nitrogen budget in extensive shrimp culture systems. *Aquaculture* 252, 248–263. doi: 10.1016/j.aquaculture.2005.06.044

Henze, M., Kristensen, G. H., and Strube, R. (1994). Rate-capacity characterization of wastewater for nutrient removal processes. *Water Sci. Technol.* 29, 101–107.

Inwood, S. E., Tank, J. L., and Bernot, M. J. (2007). Factors controlling sediment denitrification in midwestern streams of varying land use. *Microb. Ecol.* 53, 247–258. doi: 10.1007/s00248-006-9104-2

Kandeler, E., Deiglmayr, K., Tscherko, D., Bru, D., and Philippot, L. (2006). Abundance of *narG*, *nirS*, *nirK*, and *nosZ* genes of denitrifying bacteria during primary successions of a glacier foreland. *Appl. Environ. Microbiol.* 72, 5957–5962. doi: 10.1128/AEM.00439-06

Khan, S. T., Horiba, Y., Yamamoto, M., and Hiraishi, A. (2002). Members of the family Comamonadaceae as primary poly (3-hydroxybutyrate-co-3-hydroxyvalerate)-degrading denitrifiers in activated sludge as revealed by a polyphasic approach. *Appl. Environ. Microbiol.* 68, 3206–3214. doi: 10.1128/aem.68.7.3206-3214.2002

Koparal, A. S., and Ogutveren, U. B. (2002). Removal of nitrate from water by electroreduction and electrocoagulation. *J. Hazard. Mater.* 89, 83–94. doi: 10.1016/S0304-3894(01)00301-6

Kraft, B., Tegetmeyer, H. E., Sharma, R., Klotz, M. G., Ferdelman, T. G., Hettich, R. L., et al. (2014). The environmental controls that govern the end product of bacterial nitrate respiration. *Science* 345, 676–679. doi: 10.1126/science.1254070

Laverman, A. M., Garnier, J. A., Mounier, E. M., and Roose-Amsaleg, C. L. (2010). Nitrous oxide production kinetics during nitrate reduction in river sediments. *Water Res.* 44, 1753–1764. doi: 10.1016/j.watres.2009.11.050

Mannina, G., Capodici, M., Cosenza, A., and Di Trapani, D. (2016). Carbon and nutrient biological removal in a university of Cape town membrane bioreactor: analysis of a pilot plant operated under two different C/N ratios. *Chem. Eng. J.* 296, 289–299. doi: 10.1016/j.cej.2016.03.114

Mohan, T. V. K., Nancharaiah, Y. V., Venugopalan, V. P., and Sai, P. M. S. (2016). Effect of C/N ratio on denitrification of high-strength nitrate wastewater in anoxic granular sludge sequencing batch reactors. *Ecol. Eng.* 91, 441–448. doi: 10.1016/j.ecoleng.2016.02.033

Morrissey, E. M., and Franklin, R. B. (2015). Resource effects on denitrification are mediated by community composition in tidal freshwater wetlands soils. *Environ. Microbiol.* 17, 1520–1532. doi: 10.1111/1462-2920.12575

Münch, E. V., Lant, P., and Keller, J. (1996). Simultaneous nitrification and denitrification in bench-scale sequencing batch reactors. *Water Res.* 30, 277–284. doi: 10.1016/0043-1354(95)00174-3

Saarenheimo, J., Tiirola, M. A., and Rissanen, A. J. (2015). Functional gene pyrosequencing reveals core proteobacterial denitrifiers in boreal lakes. *Front. Microbiol.* 6:674. doi: 10.3389/fmicb.2015.00674

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/aem.01541-09

Shrimali, M., and Singh, K. P. (2001). New methods of nitrate removal from water. *Environ. Pollut.* 112, 351–359. doi: 10.1016/s0269-7491(00)00147-0

Simkins, S., and Alexander, M. (1984). Models for mineralization kinetics with the variables of substrate concentration and population density. *Appl. Environ. Microbiol.* 47, 1299–1306.

Stouthamer, A. H. (1992). Metabolic pathways in *Paracoccus denitrificans* and closely related bacteria in relation to the phylogeny of prokaryotes. *Antonie Van Leeuwenhoek* 61, 1–33. doi: 10.1007/BF00572119

Strong, P., McDonald, B., and Gapes, D. (2011). Enhancing denitrification using a carbon supplement generated from the wet oxidation of waste activated sludge. *Bioresour. Technol.* 102, 5533–5540. doi: 10.1016/j.biortech.2010.12.025

Tiemeyer, A., Link, H., and Weuster-Botz, D. (2007). Kinetic studies on autohydrogenotrophic growth of *Ralstonia eutropha* with nitrate as terminal electron acceptor. *Appl. Microbiol. Biotechnol.* 76, 75–81. doi: 10.1007/s00253-007-0983-z

Wang, Y., Zhou, S., Ye, L., Wang, H., Stephenson, T., and Jiang, X. (2014). Nitrite survival and nitrous oxide production of denitrifying phosphorus removal sludges in long-term nitrite/nitrate-fed sequencing batch reactors. *Water Res.* 67, 33–45. doi: 10.1016/j.watres.2014.08.052

Wang, Z., Guo, F., Liu, L., and Zhang, T. (2014). Evidence of carbon fixation pathway in a bacterium from candidate phylum SBR1093 revealed with genomic analysis. *PLOS ONE* 9:e109571. doi: 10.1371/journal.pone.0109571

Ward, B. B., and Jensen, M. M. (2014). The microbial nitrogen cycle. *Front. Microbiol.* 5:553. doi: 10.3389/fmicb.2014.00553

Water Environment Federation (2005). *Standard Methods for the Examination of Water and Wastewater*. Washington, DC: American Public Health Association (APHA).

Wei, Y., Liao, S. A., and Wang, A. L. (2016). The effect of different carbon sources on the nutritional composition, microbial community and structure of bioflocs. *Aquaculture* 465, 88–93. doi: 10.1016/j.aquaculture.2016.08.040

Wu, J., Zhang, J., Jia, W., Xie, H., Gu, R. R., Li, C., et al. (2009). Impact of COD/N ratio on nitrous oxide emission from microcosm wetlands and their performance in removing nitrogen from wastewater. *Bioresour. Technol.* 100, 2910–2917. doi: 10.1016/j.biortech.2009.01.056

Yoshida, M., Ishii, S., Otsuka, S., and Senoo, K. (2009). Temporal shifts in diversity and quantity of *nirS* and *nirK* in a rice paddy field soil. *Soil Biol. Biochem.* 41, 2044–2051. doi: 10.1016/j.soilbio.2009.07.012

Zhao, W., Wang, Y. Y., Liu, S. H., Pan, M. L., Yang, J., and Chen, S. W. (2013). Denitrification activities and N2O production under salt stress with varying COD/N ratios and terminal electron acceptors. *Chem. Eng. J.* 215, 252–260. doi: 10.1016/j.cej.2012.10.084

Zhi, W., and Ji, G. (2014). Quantitative response relationships between nitrogen transformation rates and nitrogen functional genes in a tidal flow constructed wetland under C/N ratio constraints. *Water Res.* 64, 32–41. doi: 10.1016/j.watres.2014.06.035

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chen, Deng, He and Hou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 June 2017
doi: 10.3389/fmicb.2017.01020

Predicting Species-Resolved Macronutrient Acquisition during Succession in a Model Phototrophic Biofilm Using an Integrated 'Omics Approach

Stephen R. Lindemann[1,2,3], Jennifer M. Mobberley[1†], Jessica K. Cole[1], L. M. Markillie[2], Ronald C. Taylor[1], Eric Huang[1], William B. Chrisler[1], H. S. Wiley[4], Mary S. Lipton[4], William C. Nelson[1], James K. Fredrickson[1] and Margaret F. Romine[1]*

[1] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, United States, [2] Whistler Center for Carbohydrate Research, Department of Food Science, Purdue University, West Lafayette, IN, United States, [3] Department of Nutrition Science, Purdue University, West Lafayette, IN, United States, [4] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, United States

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council, Argentina

Reviewed by:
Steven Singer, Lawrence Berkeley National Laboratory, United States
Scott Rice, Singapore Center on Environmental Life Sciences Engineering, Singapore

***Correspondence:**
Stephen R. Lindemann
lindemann@purdue.edu

†Present address:
Jennifer M. Mobberley, Program in Biomolecular Science and Engineering, Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara, CA, United States

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: 28 February 2017
Accepted: 22 May 2017
Published: 13 June 2017

Citation:
Lindemann SR, Mobberley JM, Cole JK, Markillie LM, Taylor RC, Huang E, Chrisler WB, Wiley HS, Lipton MS, Nelson WC, Fredrickson JK and Romine MF (2017) Predicting Species-Resolved Macronutrient Acquisition during Succession in a Model Phototrophic Biofilm Using an Integrated 'Omics Approach. Front. Microbiol. 8:1020. doi: 10.3389/fmicb.2017.01020

The principles governing acquisition and interspecies exchange of nutrients in microbial communities and how those exchanges impact community productivity are poorly understood. Here, we examine energy and macronutrient acquisition in unicyanobacterial consortia for which species-resolved genome information exists for all members, allowing us to use multi-omic approaches to predict species' abilities to acquire resources and examine expression of resource-acquisition genes during succession. Metabolic reconstruction indicated that a majority of heterotrophic community members lacked the genes required to directly acquire the inorganic nutrients provided in culture medium, suggesting high metabolic interdependency. The sole primary producer in consortium UCC-O, cyanobacterium *Phormidium* sp. OSCR, displayed declining expression of energy harvest, carbon fixation, and nitrate and sulfate reduction proteins but sharply increasing phosphate transporter expression over 28 days. Most heterotrophic members likewise exhibited signs of phosphorus starvation during succession. Though similar in their responses to phosphorus limitation, heterotrophs displayed species-specific expression of nitrogen acquisition genes. These results suggest niche partitioning around nitrogen sources may structure the community when organisms directly compete for limited phosphate. Such niche complementarity around nitrogen sources may increase community diversity and productivity in phosphate-limited phototrophic communities.

Keywords: carbon fixation, nitrate reduction, phosphate transport, sulfate reduction, metagenomics, metatranscriptomics, metaproteomics, periphyton

INTRODUCTION

Lack of mechanistic understanding of energy and element flow through microbial communities relegates them to black boxes in predictive models of ecosystem functioning (Nazaries et al., 2013; Graham et al., 2014; Rousk and Bengtson, 2014). Understanding the variables that control resource acquisition and partitioning in dynamic microbial communities is central to our ability

to predict how biogeochemical cycles will respond to environmental change (Konopka et al., 2015). This is especially true for benthic microbial communities, in which energy and element cycling are affected by both members' abundances and activities and physicochemical gradients on micron scales (Battin et al., 2003; Moran et al., 2014). In aquatic systems, phototrophic biofilms ("periphyton") frequently serve as an ecosystem entry point for energy, carbon, and other macronutrients (Vadeboncoeur and Steinman, 2002; Kautza et al., 2016) and exert significant impacts on nutrient fluxes and carbon cycling at fluvial scales (Flipo et al., 2004; Battin et al., 2008). In these multi-species biofilms, consisting of photoautotrophs and associated heterotrophic consorts encased in a biogenic matrix of extracellular polymers (Roeselers et al., 2008), exchange of macronutrients between autotrophs and heterotrophs drives community succession (Wagner et al., 2015).

Though much research has been directed at understanding nutrient uptake by phototrophic biofilms at whole-community scales (Battin et al., 2016), little is known about how individual members acquire and exchange nutrients. This has resulted, in part, from an inability to resolve metabolic function at the level of individual species. However, recent advances in the reconstruction of individual genomes from metagenomes now permit the assignment of potential function at the species level (Pope et al., 2011; Hugerth et al., 2015; Palomo et al., 2016). We have applied these approaches to generate species-resolved metagenomes for two unicyanobacterial consortia (Nelson et al., 2016). These consortia are each composed of one distinct cyanobacterium and a nearly identical suite of ∼18 heterotrophic species from *Alphaproteobacteria*, *Gammaproteobacteria*, and *Bacteroidetes*, which form a benthic biofilm that undergoes a reproducible succession in the laboratory (Cole et al., 2014). Using these model systems, we predicted entry points for energy and elements into phototrophic biofilm communities by identifying each species' functional potential for light energy capture and acquisition of the macronutrients carbon, nitrogen, phosphorus, and sulfur. Metatranscriptomic and metaproteomic analyses enabled attribution of energy and macronutrient acquisition processes at the level of individual species during succession. We observed common responses across the community to some nutrients (e.g., phosphorus) as well as highly individual strategies for others (e.g., nitrogen). These data allow a mechanistic understanding of community nutrient flow and suggest that niche complementarity or plasticity centered around nitrogen may minimize competition, maintaining diversity in phototrophic biofilms when organisms directly compete for limited phosphorus resources.

MATERIALS AND METHODS

Consortia and Succession Experiments

Unicyanobacterial consortia UCC-A and UCC-O were isolated from a phototrophic microbial mat in Hot Lake, Washington (Lindemann et al., 2013) and were cultivated and sampled for succession experiments as previously described (Cole et al., 2014). Briefly, sequentially passaged enrichment cultures were inoculated at a 1:50 dilution into HLA medium (essentially, BG-11 medium supplemented with 400 mM $MgSO_4$, 100 mM Na_2SO_4, and 25 mM KCl at pH 8.0) in T75 tissue culture flasks with vented caps (Costar, Corning, Inc., Corning, NY, United States) under 35 $\mu E\ m^{-2}\ s^{-1}$ (General Electric PL/AQ, Fairfield, CT, United States). Cultures were incubated for 28 days and sterile deionized water was added weekly to maintain constant volume and salinity, with triplicate biological replicates (parallel T-75 flasks) harvested and split for proteomic and transcriptomic analysis at weekly intervals as previously described (on days 7, 14, 21, and 28 of cultivation; Cole et al., 2014). Briefly, biofilms were dislodged from the bottom of an entire T-75 flask using a cell scraper (Costar; Corning, Inc., Corning, NY, United States), placed into a 50 mL conical vial, and homogenized using sterile 3 mm glass beads and hand shaking, divided into separate samples for transcriptomics and proteomics, and centrifuged at 4°C and 5000 × *g* for 10 min prior to decanting and plunge freezing in liquid N_2.

Species-Resolved Genome Information

Metagenomes for UCC-A and UCC-O and genomes of isolates therefrom (*Algoriphagus marincola* HL-49, *Aliidiomarina calidilacus* HL-53 (Morton et al., in review), *Erythrobacteraceae* sp. HL-111, *Halomonas* sp. HL-48, *Halomonas* sp. HL-93, *Marinobacter excellens* HL-55, *Marinobacter* sp. HL-58, and *Porphyrobacter* sp. HL-46, *Roseibaca calidilacus* HL-91 (Maezato et al., in review), and *Salinivirga fredricksonii* HL-109 (Cole et al., in review) were generated by the DOE Joint Genome Institute, and species-resolved genome reconstructions were recently described (Nelson et al., 2016). These were evaluated for the presence of genes involved in light energy, carbon, nitrogen, phosphorus, or sulfur acquisition under experimental conditions. Genome completeness estimates from single-copy-gene analysis (Nelson et al., 2016) are presented in **Figure 1**; *Rhodobacteraceae* sp. Bin24 was excluded from further analysis due to insufficient completeness. A list of the accession numbers for genome reconstructions from metagenomes (European Nucleotide Archive[1]) and isolate genomes (GenBank[2]) are provided in **Supplementary Table S1** and are also available through IMGer[3].

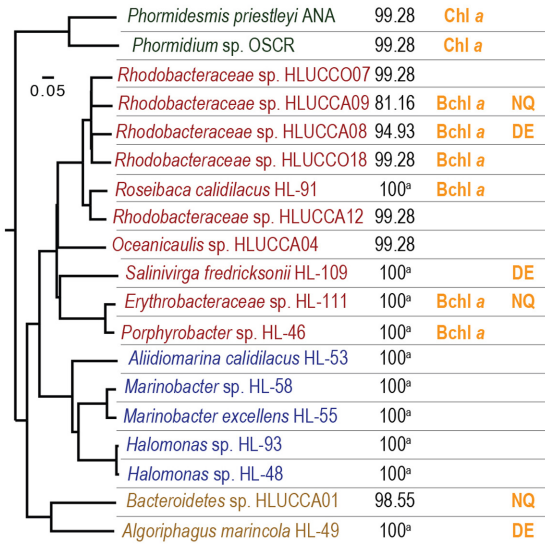
Protein Function Predictions

Protein-coding gene models and functional predictions were initially assigned by the DOE-JGI microbial genome annotation pipeline (Markowitz et al., 2014) and manually curated by evaluating additional evidence collected by submitting these gene models to RAST (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015) and BlastKoala (Kanehisa et al., 2016) as well as conducting local domain searches for TIGRfams (v. 14) and Pfams (v. 27) using HMMer v. 3.1 (Johnson et al., 2010). Identified genes involved in energy and macronutrient acquisition are provided in **Supplementary Table S2**. Note that, except for isolates where complete genomes are available,

[1]www.ebi.ac.uk/ena

[2]https://www.ncbi.nlm.nih.gov/genbank/

[3]https://img.jgi.doe.gov/cgi-bin/mer/main.cgi



Organism	% Comp.	Photo-system	PR[b]	Carbon fixation	NO_3^- red. to NO_2^-	NO_2^- red. to NH_4^+	SO_4^{2-} activation	SO_4^{2-} red. to SO_3^{2-}	SO_3^{2-} red. to S^{2-}	PO_4^{3-} uptake[f]	Alkaline phosphatase
Phormidesmis priestleyi ANA	99.28	Chl *a*		*rbcLXS*	*narB*	*nirA*	*sat, cysC*	*cysH*	*sir*	*pstSABC*	*phoAX*
Phormidium sp. OSCR	99.28	Chl *a*		*rbcLXS*	*narB*	*nirA*	*sat, cysC*	*cysH*	*sir*	*pstSABC*	*phoX*
Rhodobacteraceae sp. HLUCCO07	99.28						*sat-cysC*	*cysH*		*pst, yjbB*	
Rhodobacteraceae sp. HLUCCA09	81.16	Bchl *a*	NQ				*sat-cysC*			*pstSABC*	*phoAX*
Rhodobacteraceae sp. HLUCCA08	94.93	Bchl *a*	DE		*narGHIJ*		*sat-cysC*	*cysH*	*cysIJ*	*pstAB*[g]	*phoX*
Rhodobacteraceae sp. HLUCCO18	99.28	Bchl *a*					*sat-cysC*			*pst, yjbB*	*phoX*
Roseibaca calidilacus HL-91	100[a]	Bchl *a*			*narGHIJ*		*cysDNC*	*cysH*[e]	*cysIJ*	*pst, pitA*	
Rhodobacteraceae sp. HLUCCA12	99.28				*narGHIJ*		*sat-cysC*	*cysH*	*cysIJ*	*pst, yjbB, pitA*	
Oceanicaulis sp. HLUCCA04	99.28						*cysDNC*			*pstSABC*	*phoDX*
Salinivirga fredricksonii HL-109	100[a]		DE				*cysDNC*			*pst, yjbB, pitA*	
Erythrobacteraceae sp. HL-111	100[a]	Bchl *a*	NQ				*cysDNC*	*cysH*	*cysIJ*	*pstSABC*	*phoADX*
Porphyrobacter sp. HL-46	100[a]	Bchl *a*					*cysDNC*	*cysH*	*cysIJ*	*pstSABC*	*phoADX*
Aliidiomarina calidilacus HL-53	100[a]						*cysDNC*			*pst, yjbB, pitA*	*phoA*
Marinobacter sp. HL-58	100[a]			*nasA, narG, napA*[d]	*nirB*		*cysDNC*	*cysH*[e]	*cysIJ*	*pst, pitA*	*phoX*
Marinobacter excellens HL-55	100[a]			*napAB*			*cysDN*	*cysH*[e]	*cysIJ*	*pst, yjbB, pitA*	*phoDX*
Halomonas sp. HL-93	100[a]			*nasA, narG*[d]	*nirB*		*cysDN*	*cysH*	*cysIJ*	*pst, pitA*	*phoX*
Halomonas sp. HL-48	100[a]			*nasA*	*nirB*		*cysDN*	*cysH*	*cysI*	*pst, pitA*	*phoX*
Bacteroidetes sp. HLUCCA01	98.55		NQ	*rbcL*[c]		*nrfAH*	*cysDNC*			*yjbB*	*phoA*
Algoriphagus marincola HL-49	100[a]		DE				*cysDNC*	*cysH*[e]	*sir*	*yjbB*	*phoD*

FIGURE 1 | Macronutrient acquisition functions detected in phototrophic consortium member genomes. Isolates belonging to the *Cyanobacteria* are green, *Alphaproteobacteria* are red, *Gammaproteobacteria* are blue, and *Bacteroidetes* are yellow. [a]Genomic information derived from sequenced isolates. [b]For description of the types of proteorhodopsins, see Supplementary Notes 2. [c]Similar to form IV RuBisCo, and thus unlikely to permit autotrophy (Supplementary Notes 3). [d]These organisms contain multiple complete nitrate reductases; only one gene of each type of multi-subunit nitrate reductase is denoted to conserve space. [e]Where organisms have multiple phosphate-acquisition systems, *pst* denotes the presence of *pstSABC*. [f]The *pstAB* genes are contiguous to a contig edge in the *Rhodobacteraceae* sp. HLUCCA08 reconstruction; *pstSC* are assumed to be within the gap between contigs.

missing functions could reside in the gaps between contigs despite the near-completeness of genome reconstructions. Consequently, our predictions of organismal function are necessarily conservative, especially since some functions could require only a small number of genes that are frequently found together in operons. For details on genome annotation, please see the Supplementary Notes.

RNA Extraction

RNA was extracted using Invitrogen TRIzol Reagent (cat. #15596018), followed by genomic DNA removal and cleanup using Qiagen RNase-Free DNase Set kit (cat. #79254) and Qiagen Mini RNeasy kit (cat. #74104). An Agilent 2100 Bioanalyzer was used to assess the integrity of the RNA samples; only RNA samples having RNA Integrity Number score of 8–10 were sequenced.

RNA Sequencing

The Applied Biosystems SOLiD Total RNA-Seq kit (cat. #4445374) was used to generate the cDNA template library according to manufacturer's instructions. The SOLiD EZ Bead system was used to perform emulsion clonal bead amplification to generate bead templates for sequencing on the 5500XL SOLiD platform. The 50-base read sequences produced by the 5500XL SOLiD sequencer were mapped in color space using SOLiD LifeScope software version 2.5 using the default parameters against an artificial chromosome, as previously described (Hess et al., 2013). LifeScope has been previously determined to be the optimal method for mapping sequence data obtained via the SOLiD 5500 system (Pranckevičiene et al., 2015), but it is optimized for use with eukaryotic genomes with multiple linear chromosomes. Consequently, we constructed an artificial chromosome as a reference by concatenating genomes of isolated UCC members and all contigs of remaining metagenome bins. These were separated by ten ambiguous nucleotides to prevent edge effects that might otherwise disturb mapping near contig boundaries. This reference contained species-resolved genome information for all consortium members. Transcription of each organism's genes was individually normalized to reads per kilobase per million reads (RPKM), using the total reads mapped to all the organism's genes as the normalization basis. It should be noted that though this normalization approach facilitates gene expression comparisons within a single species across time points, normalized gene expression cannot be compared across species. The transcriptomics data associated with this study has been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus under accession number GSE99220.

Global Proteomics: Extraction, Digestion, and 2D-LC-MS/MS Analysis

UCC-O cell pellets (typically ∼100–500 mg, dry weight) were suspended in 100 mM NH_4HCO_3 buffer (pH 8.0) and then subjected to bead beating in a Bullet Blender homogenizer (Next Advance Inc., Averill Park, NY, United States) for 3 min with 0.1 mm zirconia/silica beads (BioSpec Products, Inc.,

Bartlesville, OK, United States). Proteins were separated into global, soluble, and insoluble fractions and processed into peptides for subsequent LC-MS/MS analysis using previously described methods (Callister et al., 2006). Briefly, proteins designated for global analysis were denatured and reduced under following conditions: 7M Urea, 5 mM DTT, 100 mM ammonium bicarbonate buffer at 60°C for 45 min. After denaturing, samples were diluted eightfold with 100 mM ammonium bicarbonate and a sufficient amount of calcium chloride was added to achieve 1 mM. Tryptic digestion was performed for 3 h at 37°C with 1:50 (w/w) trypsin-to-protein ratio. The digested sample was desalted and cleaned via solid phase extraction (SPE) C18 (Supelco, Bellefonte, PA, United States). Sample was concentrated in Speed-Vac (Thermo Savant, Holbrook, NY, United States) before performing a BCA Assay to determine final peptide concentration. A portion of the lysate for soluble/insoluble analysis was ultracentrifuged and the supernatant was treated as above, whereas the pellet was resuspended by sonication in denaturing buffer containing 1% CHAPS in 50 mM ammonium bicarbonate (pH 7.8) before enzymatic digestion. On line 2D-LS-MS/MS analysis of the peptides was achieved by using previously described methods (Smith et al., 2014). MS analysis was performed using a LTQ Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA, United States) operated as described by Callister et al. (2006). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD006440 and doi: 10.6019/PXD006440.

MS/MS Data Analysis

MS/MS data were searched against protein sets for all organisms present in UCC-O (Nelson et al., 2016), a total of 60,759 proteins. Protein sets were augmented by common contaminant protein sequences (human keratin, bovine trypsin, and serum albumin precursor) to detect residual peptides derived from processing. MS/MS spectra were preprocessed using DeconMSn and DtaRefinery tools to deconvolute, deisotope, and remove systematic error in mass measurement accuracy (Mayampurath et al., 2008; Petyuk et al., 2010). The MS-GF+ search algorithm was used to match MS/MS spectra to peptide sequences (Kim et al., 2008), including partial tryptic cleavage peptides, dynamic modification of methionine oxidations, and maximum 10 ppm parent ion mass tolerance in the search. IDPicker 3.0 was used to filter peptide-spectrum matches to 2% FDR and apply parsimony filtering to derive a minimum protein list, with each protein supported by at least two distinct peptides (Ma et al., 2009; Tabb et al., 2010). For *Phormidium* sp. OSCR, proteomic coverage was sufficient to holistically describe gene expression; consequently, spectral counts of all peptides of a protein were summed and are presented as percentage of this species' share of peptides observed at each time point.

Fluorescence *In Situ* Hybridization (FISH)

Species-specific fluorescence probes were designed using the DECIPHER R package (Wright et al., 2014). Briefly, a list of the targeted 16S sequences and a FASTA file of potential off-target 16S sequences was generated from the UCC-O genome sequence (Nelson et al., 2016). Optimal specific probes were then generated with a length between 18 and 21 nucleotides, and a 46°C hybridization temperature in 0.9M NaCl and 20% formamide. This yielded probe GATACCCGAAAGCATCTCT for HL-49. Specificity of the probe was experimentally validated using mixtures of axenic cultures derived from the UCC-O biofilm. FISH of biofilms grown for 7 days was performed following the method described by Amann and Fuchs (2008). After FISH hybridization, samples were stained with Hoechst 33342 (Sigma, St. Louis, MO, United States) to stain DNA. Microscopic images were acquired on a Zeiss LSM 710 Scanning Confocal Laser Microscope (Carl Zeiss MicroImaging GmbH, Jena, Germany) for both fluorescence and differential interference contrast imaging. Z-stack confocal images were acquired using an EC Plan-Neofluar 10x/0.30 M27 objective. DNA fluorescence was excited at 405 nm visualized at 410–495 nm. FISH fluorescence was excited at 561 nm visualized at 566–680 nm. Phycocyanin/chlorophyll auto-fluorescence was excited at 633 nm and visualized at 647–721 nm. The images were further processed with Volocity (Perkin Elmer, Waltham, MA, United States).

RESULTS AND DISCUSSION

Genome Predictions for Energy and Macronutrient Acquisition Potential of Species

We utilized our previously generated, species-resolved metagenome analysis (Nelson et al., 2016) to predict each consortium member's genetic potential to harvest light energy and to directly acquire the required macronutrients carbon, nitrogen, phosphorus, and sulfur from the inorganic sources present in the culture medium (HLA; Cole et al., 2014). An overview of functions predicted within each member's genome is shown in **Figure 1** and indicated by its type (e.g., Chl *a* vs. Bchl *a*-based phototrophy) or its keystone genes; the full list of genes important for these functions is detailed in **Supplementary Table S2**.

Energy

The consortia are routinely grown with light as the only energy source. Although oxygenic phototrophs are the main sources of energy capture by the consortia (Cole et al., 2014), several heterotrophic members encode bacteriochlorophyll *a* (Bchl *a*)-based photosystems. Such systems may operate under anoxic conditions, using reduced substrates (e.g., sulfide, fumarate) as electron donors for carbon fixation, or aerobically, driving cyclic transport of electrons to generate ATP. However, no Bchl *a*-containing photosystem has yet been shown to be capable of both anaerobic and aerobic phototrophy (Yurkov and Beatty, 1998; Rathgeber et al., 2012). Sequence analysis alone is incapable of differentiating anaerobic from aerobic photosystems, as both contain structurally and phylogenetically similar reaction centers and light-harvesting antenna complexes (Yurkov and Csotonyi, 2009). Multiple *Rhodobacteraceae* spp. (*Roseibaca calidilacus* HL-91 and HLUCCA08, HLUCCA09, and HLUCCO18) and

Erythrobacteraceae spp. (HL-111 and *Porphyrobacter* sp. HL-46) in the consortia possess all the genes required for Bchl *a* synthesis and photosystem assembly. Under continuously illuminated oxic growth conditions, we postulate that only organisms with aerobic photosystems (aerobic anoxygenic phototrophs, or AAPs) are likely to supplement cyanobacterial light energy acquisition (Cole et al., 2014). However, because AAP photosystems do not generate the reductant required to fix inorganic carbon (Yurkov and Beatty, 1998), they are constrained to photoheterotrophy or chemoheterotrophy.

One additional means by which light energy could be captured by the consortia is through bacterial proteorhodopsins (PRs), which are transmembrane proteins that use photons to drive proton or other ion gradients that generate ATP and energize membrane transporters (McCarren and DeLong, 2007). PR-like proteins are encoded by heterotrophic members of *Bacteroidetes* (HLUCCA01 and *A. marincola* HL-49) and *Alphaproteobacteria* (*Rhodobacteraceae* spp. HLUCCA08 and HLUCCA09, *Salinivirga fredricksonii* HL-109, and *Erythrobacteraceae* sp. HL-111). In *Bacteroidetes* sp. HLUCCA01, *Rhodobacteraceae* sp. HLUCCA09, and *Erythrobacteraceae* sp. HL-111, the putative rhodopsin contains the RYXN(X_{10})Q transport motif characteristic of the NQ family of rhodopsins rather than the RYXD(X_{10})E proton-transport motif; NQ rhodopsins are common in hypersaline environments (Kwon et al., 2013) and have recently been shown to transport sodium ions (Balashov et al., 2014). This suggests that rhodopsins may perform other functions besides maintaining proton-motive force for ATP generation in these organisms, such as regulating osmotic pressure or driving efflux pumps via cation antiport (Fuhrman et al., 2008). Consequently, it is possible that these PR-containing heterotrophs could harvest light energy although recent theoretical work suggests the net energetic advantage of PR-containing bacteria is significantly smaller than for AAPs (Kirchman and Hanson, 2013).

Carbon

Genomic evidence supports our previous hypothesis that the cyanobacteria were the sole autotrophs within the consortia (Cole et al., 2014). Both cyanobacteria contained the genes (*rbcL, rbcS*) required to construct the canonical, hexadecameric ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo). RuBisCo catalyzes the addition of carbon dioxide to ribulose-1,5-bisphosphate and is required for the Calvin-Benson-Bassham reductive pentose phosphate cycle of carbon fixation (Tabita et al., 2007; Erb et al., 2012). Although *Bacteroidetes* sp. HLUCCA01 also contains an *rbcL* homolog, its catalytic motif is similar to the form IV RbcLs of *Rhodopseudomonas palustris* (gi: 77687805) and *Rhodospirillum rubrum* (gi: 48764419) as it is ∼100 residues shorter than form I-III RbcLs and His replaces the canonical Glu^{204} residue in the catalytic motif (Carré-Mlouka et al., 2006). Form IV RuBisCos, also termed RuBisCo-like proteins, have been found to lack the ability to fix carbon dioxide but instead catalyze an enolization reaction important for the salvage of methionine from methylthioadenosine (Tabita et al., 2007; Erb et al., 2012). Consequently, this gene is unlikely to enable autotrophy in *Bacteroidetes* sp. Bin01. HLUCCA01.

Although it is likely that the vast majority of the inorganic carbon entering the consortia comes through the cyanobacterial primary producers, a modest amount of carbon could enter the community through the anapleurotic reactions of heterotrophs and, especially, of photoheterotrophs. For example, *Roseobacter denitrificans* has been shown to acquire 10–15% of its protein carbon through the activities of pyruvate carboxylase and/or phosphoenolpyruvate (PEP) carboxylase (Tang et al., 2009). All members of the consortia encoded at least one mechanism to convert PEP or pyruvate to oxaloacetate through the incorporation of carbon dioxide or bicarbonate (i.e., either pyruvate or PEP carboxylase) except *A. calidilacus* HL-53.

Nitrogen

Because the consortia are routinely cultivated with abundant nitrate (17.6 mM), the primary organisms through which nitrogen enters the consortia will be those capable of reducing nitrate to nitrite to bioavailable ammonium (Cole et al., 2014). Although both *Phormidium* sp. OSCR and *Phormidesmis priestleyi* ANA also contain *nif* genes encoding the MoFe nitrogenases that catalyze the reduction of N_2 to ammonium, the expression of nitrogenases is typically strongly repressed in cyanobacteria when nitrate or ammonium are available (e.g., Bottomley et al., 1979; Mackerras and Smith, 1986), as under these growth conditions. Both cyanobacterial genomes possess ferredoxin-dependent, cytoplasmic nitrate and nitrite reductases, allowing them to serve as primary producers with respect to nitrogen acquisition. However, nitrate assimilation is not uniformly distributed among bacteria. Thus, though we originally hypothesized that our enrichment conditions would select for nitrate-assimilating heterotrophs, genomic analysis suggested that the majority of heterotrophic species lacked the ability to reduce nitrate, nitrite, or both. Only the gammaproteobacterial genomes, except for *A. calidilacus* HL-53 and *M. excellens* HL-55, contained cytoplasmic NO_3^- nitrate (*nasA*) and nitrite reductases (*nirB*) required for assimilatory reduction. In addition to assimilatory reductases, some gammaproteobacterial members also contained dissimilatory nitrate reductase genes, which have been shown to substitute for *nasA* in some species (Moreno-Vivian et al., 1999); *Halomonas* sp. HL-93 contained the membrane-bound respiratory reductase *narGHIJI*, *M. excellens* HL-55 possesses the periplasmic *napAB* system, and *Marinobacter* sp. HL-58 encoded both types. Alphaproteobacterial members *Rhodobacteraceae* spp. HL-91, HLUCCA08, and HLUCCA12 also encoded *narGHIJ* for dissimilatory reduction of nitrate to nitrite but no predicted nitrite reductases. Conversely, *Bacteroidetes* sp. HLUCCA01 contained the dissimilatory, ammonifying nitrite reductase *nrfAH* but appeared to lack a nitrate reductase. All members of the consortia contained Amt-like ammonium transporters and glutamine synthetases required to incorporate ammonium.

Sulfur

Although all consortium members were predicted to transport and activate sulfate, only a subset appeared to be able to reduce sulfate to sulfide for the biosynthesis of cysteine and methionine, thereby serving as entry points for sulfur into the

community. All members were capable of activating sulfate to 3′-phosphoadenylyl-sulfate (PAPS), which generates a high-energy phosphoric-sulfuric acid anhydride bond and allows transfer or reduction of the sulfurylyl group (Leyh et al., 1988). *Roseibaca calidilacus* HL-91, *Marinobacter* sp. HL-58, *Marinobacter excellens* HL-55, and *Algoriphagus marincola* HL-49 encode thioredoxin-dependent adenosine 5′-phosphosulfate (APS) reductases (TIGR2055 family); these enzymes are capable of reducing both APS and PAPS to sulfite. However, *Bacteroidetes* sp. HLUCCA01, *A. calidilacus* HL-53, *Oceanicaulis* sp. HLUCCA04, *S. fredricksonii* HL-109, and *Rhodobacteraceae* spp. HLUCCA09 and HLUCCO18 all were predicted to lack both PAPS and sulfite reductases, which are required to produce sulfide for cysteine biosynthesis. Consequently, these members are predicted to depend upon sulfate assimilators for acquisition of bioavailable sulfur under routine culture conditions. Since these organisms are derived from a phototrophic mat, regions of which are transiently or permanently sulfidic (Lindemann et al., 2013), it is possible that organisms without the ability to reduce PAPS are able to directly acquire sulfide from the native environment for cysteine and methionine biosynthesis.

Phosphorus

Due to the extremely high concentrations of Mg^{2+} in Hot Lake and in culture medium formulated to cultivate mat-associated microorganisms, inorganic phosphorus is sparingly soluble (Lindemann et al., 2013; Cole et al., 2014; Zachara et al., 2016). The types of phosphate transporters (Willsky and Malamy, 1980; Rao and Torriani, 1990) possessed by members of the consortia presented a genomic signature of competition for phosphate; although high-affinity, ATP-utilizing PstSABC-like transporters were found in all members except both of the *Bacteroidetes*, low-affinity, high-rate PiT-like proton:phosphate symporters are encoded only in seven of the nineteen examined genomes. Both members from *Bacteroidetes* relied upon phosphate:sodium symporters of the PNaS family (termed *yjbB*; Lebens et al., 2002) for phosphate uptake, which were also present in six other genomes. The insolubility of magnesium and calcium phosphates suggests that the majority of community phosphorus exchange may occur via salvage of phosphorus-containing organic compounds or phosphatase-mediated removal of orthophosphate from inorganic phases. All but four alphaproteobacterial member species (*S. fredricksonii* HL-109, *Roseibaca calidilacus* HL-91, and *Rhodobacteraceae* spp. HLUCC07 and HLUCC12) encoded at least one alkaline phosphatase.

Species-Resolved Macronutrient Acquisition during Succession

Species-resolved genome information enabled reconstruction of the dynamics of member abundance and resource acquisition in UCC-O over a 28-day succession period (Cole et al., 2014) via metatranscriptomic and metaproteomic analysis (**Figure 2**). *Phormidium* sp. OSCR dominated both metatranscriptomic and metaproteomic analyses, averaging ∼50–60% of total mRNA reads (**Figure 2A**) and ∼90% of total peptide spectral counts (**Figure 2B**). This large share of cyanobacterial peptides allowed us to comprehensively evaluate the activity of *Phormidium* sp. OSCR via proteomics. Peptide spectral counts were insufficient to do this for any heterotroph. In general, the low RNA and protein abundances of some heterotrophic organisms greatly obscured their activity. Only six of the 17 heterotrophs were sufficiently represented throughout the succession cycle to comprehensively describe patterns in their gene transcription. The remaining eleven either could only be examined at a subset of time points or with respect to certain highly expressed functions, some of which were also represented in the metaproteome. We focus here on the six most abundant heterotrophs for which sufficient transcriptomics data exist to describe expression patterns over time (expression data for the referenced genes of all organisms are provided in **Supplementary Table S4**).

Metatranscriptomics and metaproteomics results were in agreement on the trends in member abundance between time points, but displayed large differences in their estimation of the relative gene expression activity (i.e., the fraction of the community's total mRNA or protein attributable to a given species) across the consortium members (**Figure 2C**; also see **Supplementary Table S3**). In general, the trends in relative gene expression activity (**Figures 2A,B**) matched our previous examination of relative genome abundance through succession cycles, notably the replacement of *A. calidilacus* HL-53 and other *Gammaproteobacteria* with members from *Alphaproteobacteria* and *Bacteroidetes* (Cole et al., 2014). However, estimates of gene expression activity from metatranscriptomics and metaproteomics differed substantially for some members. The ratio of relative gene expression activity to corresponding proteins from paired samples were higher than unity (i.e., larger metatranscriptomic estimates of abundance) for *A. marincola* HL-49, *Bacteroidetes* sp. HLUCCA01, and *A. calidilacus* HL-53 and lower (i.e., larger metaproteomic estimates of abundance) for *Phormidium* sp. OSCR, *R. calidilacus* HL-91, *Rhodobacteraceae* sp. HLUCCO18, and *S. fredricksonii* HL-109. The transcript (read)/protein (peptide) relative abundance ratios ranged nearly 100-fold across the most abundant organisms (a maximum of ∼16.6 for HL-49 and minimum of ∼0.2 for HL-109) and were relatively stable throughout the succession (**Figure 2C**). Overall, the inter-sample transcript fold change (e.g., between days 7 and 14, 14 and 21, and 21 and 28 of each of the 18 organisms correlated relatively well to inter-sample protein fold change (**Figure 2D**), exhibiting a slope near unity (with an $R^2 \sim 0.46$). Where discrepancies existed, changes in an organism's share of the metaproteome tended to be larger than corresponding changes in the metatranscriptome (per-species transcript and protein abundance data are presented in **Supplementary Table S3**). To eliminate the influence of changes in member relative abundance (**Figures 2A,B**) that might otherwise obscure individual species' gene expression patterns, mRNA reads and peptide spectral counts for all genes were normalized to the total number reads/peptides attributed to that species on a per-sample basis (see Materials and Methods). The relatively constant expression level of the highly conserved housekeeping gene *rpoC* over time across multiple species supports our use of this normalization approach (**Figures 3–5** and **Supplementary Table S4**).

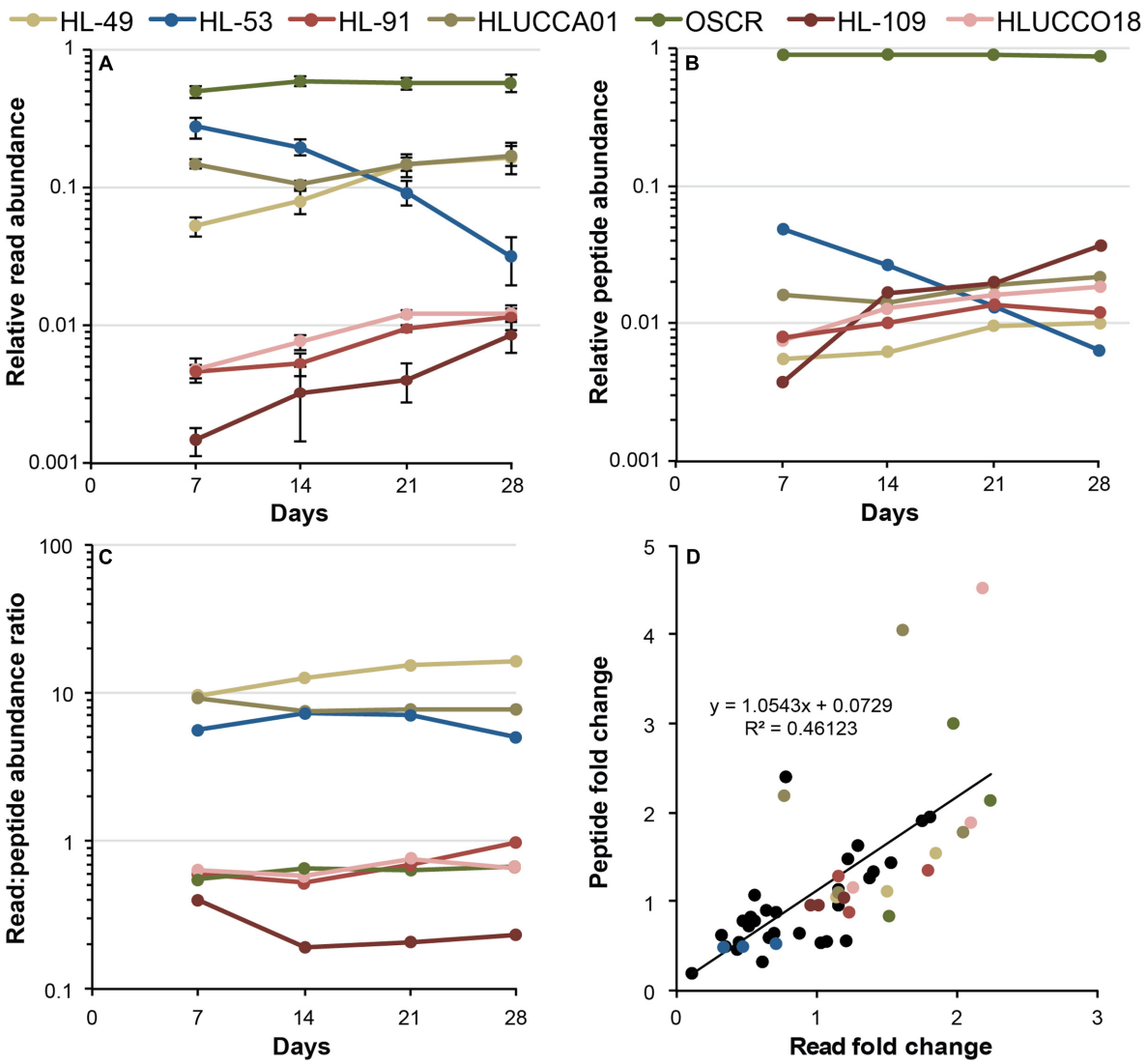


FIGURE 2 | Metatranscriptomic and metaproteomic measures of member activity dynamics. **(A)** Metatranscriptomic measurement of member activity during succession, represented as the fraction of the total reads per sample attributed to each organism. Values are plotted on a $\log_{10}$ scale, error bars denote the 95% confidence interval of the mean. **(B)** Metaproteomic evaluation of member activity during succession, represented as the share of peptide spectral counts attributed to each of the members. Values are plotted on a $\log_{10}$ scale. Error bars are omitted for clarity, but variance data are presented in **Supplementary Table S3**. **(C)** Ratio of member relative activity as assessed by metatranscriptomics to that observed by metaproteomics during succession. Flat lines ($m = 0$) indicate a monotonic relationship between metatranscriptomic and metaproteomic estimations. **(D)** Organism-centric changes in metatranscriptomic and metaproteomic estimates of relative activity across all members between sequential time points. The *y*-axis indicates the between-time point fold change in the relative abundance of an organism's peptide spectral counts, and the *x*-axis indicates the fold change in an organism's RNA read abundance. Colored dots represent the seven most dominant species according to the key, and black dots indicate fold changes for less-abundant species.

Metaproteomic analysis revealed large shifts in energy and macronutrient acquisition by the cyanobacterium *Phormidium* sp. OSCR during succession. As a share of its total proteome, abundance of both light-harvesting complexes (photosystems I and II) and antenna complexes (phycobiliproteins) declined steadily over time (**Figure 3A**), ending on day 28 at 66.5 and 63.6% of their respective day 7 totals and suggesting reduced per-cell light-energy capture. In contrast, production of chlorophyll *a* biosynthesis proteins remained steady. RuBisCo abundance declined over the 28-day succession, to 38.4% of the day 7 relative abundance. This indicates a substantially decreased per-cell fixation of inorganic carbon by the end of the experiment. Similarly, initially high expression of bacteriochlorophyll *a* synthesis and photosystem genes by photoheterotrophic species (*R. calidilacus* HL-91, *Porphyrobacter* sp. HL-46, and *Erythrobacter* sp. Bin15) declined ~four-fold during succession, suggesting reduced per-cell energy capture by these species as well (**Supplementary Table S4**). Nitrate assimilation by *Phormidium* sp. OSCR, as measured by NirA abundance, decreased concurrently with RuBisCo and was below

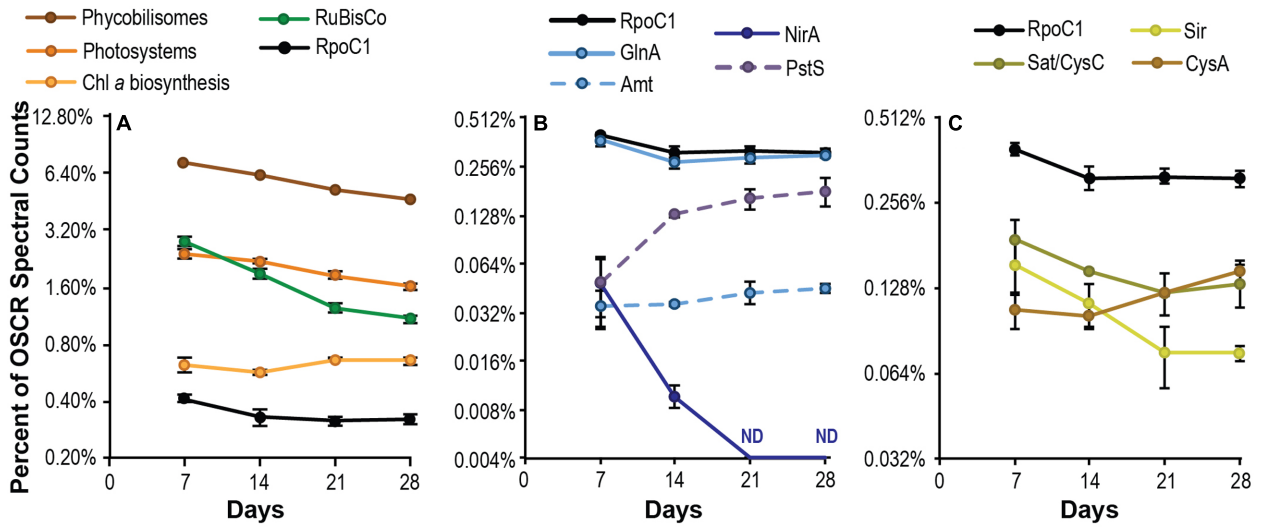


FIGURE 3 | Cyanobacterial energy capture and macronutrient acquisition during succession. Expression of *Phormidium* sp. OSCR proteins involved in light energy capture and carbon fixation **(A)**, nitrogen and phosphorus acquisition **(B)** and sulfur acquisition **(C)**, expressed as a percentage of total spectral counts per sample attributed to this species. The expression of the large subunit of the cyanobacterial beta' subunit of RNA polymerase, RpoC1, is provided for reference in all panels. Data are plotted on a $\log_2$ scale; error bars represent the 95% confidence interval of the mean. For energy and carbon capture systems **(A)**, data represent a sum of all the subsystem-specific cyanobacterial gene products notated in **Supplementary Table S2**.

detection by day 21. In contrast to the substantial reduction in nitrate assimilation proteins, ammonium uptake via Amt and incorporation via GlnA were relatively stable during succession, varying less than 1.5-fold (**Figure 3B**). We also observed substantial declines in sulfate activation (Sat/CysC, reduced 31.0%) and sulfite reduction (Sir, reduced 51.3%) between days 7 and 28 (**Figure 3C**), though Cys synthase (CysA) was stable or slightly increased. A large increase in phosphate transporter abundance, as assessed by expression of the substrate-binding subunit PstS, suggested phosphate scarcity.

Metatranscriptome analysis indicated that all but one of the dominant heterotrophic members also showed signs of increasing phosphate scarcity during succession. Although it is important to note that differences in translation efficiency (Taylor et al., 2013), post-transcriptional regulation, and protein and transcript turnover rates mean that mRNA abundance is an imperfect predictor of protein abundance (Waldbauer et al., 2012). However, transcript abundance reflects the intracellular and environmental signals to which an organism is responding. Transcripts for high-affinity *pst* phosphate transporters displayed approximately four-fold increases during succession in *A. calidilacus* HL-53, *Rhodobacteraceae* sp. HLUCCO18, *R. calidilacus* HL-91, and *S. fredricksonii* HL-109, and we observed similar increases in *yjbB* expression in *Bacteroidetes* sp. HLUCCA01 (**Figure 4**). PstS peptides were commonly observed, especially at later time points, and were among the few peptides observed from moderate-abundance members such as *M. excellens* HL-55, *Oceanicaulis* sp. HLUCCA04, and *Rhodobacteraceae* sp. HLUCCA09 (**Supplementary Table S4**), suggesting high expression within those organisms. Alkaline phosphatase genes (*phoA*, *phoD*) were also substantially upregulated in *A. calidilacus* HL-53 and *Rhodobacteraceae* sp. HLUCCO18 (**Figure 4** and **Supplementary Table S4**), though *phoA* expression in *A. calidilacus* declined after day 14, perhaps due to a reduced phosphate requirement late in the growth period. Peptides from alkaline phosphatases were observed for *Bacteroidetes* sp. HLUCCA01 and *A. calidilacus* HL-53 (PhoA) and for low-abundance member *Erythrobacteraceae* sp. HL-111 (PhoD). In organisms expressing a Pst transporter, expression of YjbB-family symporters and low-affinity PitA transporters was much lower. These data suggested that, like the cyanobacterium, most heterotrophic community members were responding to phosphate limitation by day 14. This is interesting in that UCC-O's biomass continued to increase nearly linearly throughout the 28-day experimental period and at day 14 is only approximately half its final value (Cole et al., 2014), suggesting these organisms respond transcriptionally to declining phosphate long before it becomes limiting for growth. Such a response may confer fitness in Hot Lake, where extreme competition for phosphate is likely, due to increases in magnesium concentrations throughout the seasonal cycle (Lindemann et al., 2013). In contrast, *A. marincola* HL-49 displayed stably low expression of both of its *yjbB* genes as well as *phoD*, suggesting this organism did not experience phosphorus limitation. In contrast to increasing phosphate transporter expression, heterotrophic expression of genes for sulfate activation, reduction of PAPS and sulfite, and cysteine synthesis was generally stable over the same period for all members (**Supplementary Table S4**).

Though unified (excepting *A. marincola* HL-49) in their expression patterns of phosphate and sulfate acquisition genes, the dominant heterotrophic members of UCC-O exhibited very divergent responses in expression of nitrogen acquisition genes. The abundance of gammaproteobacteria capable of assimilating

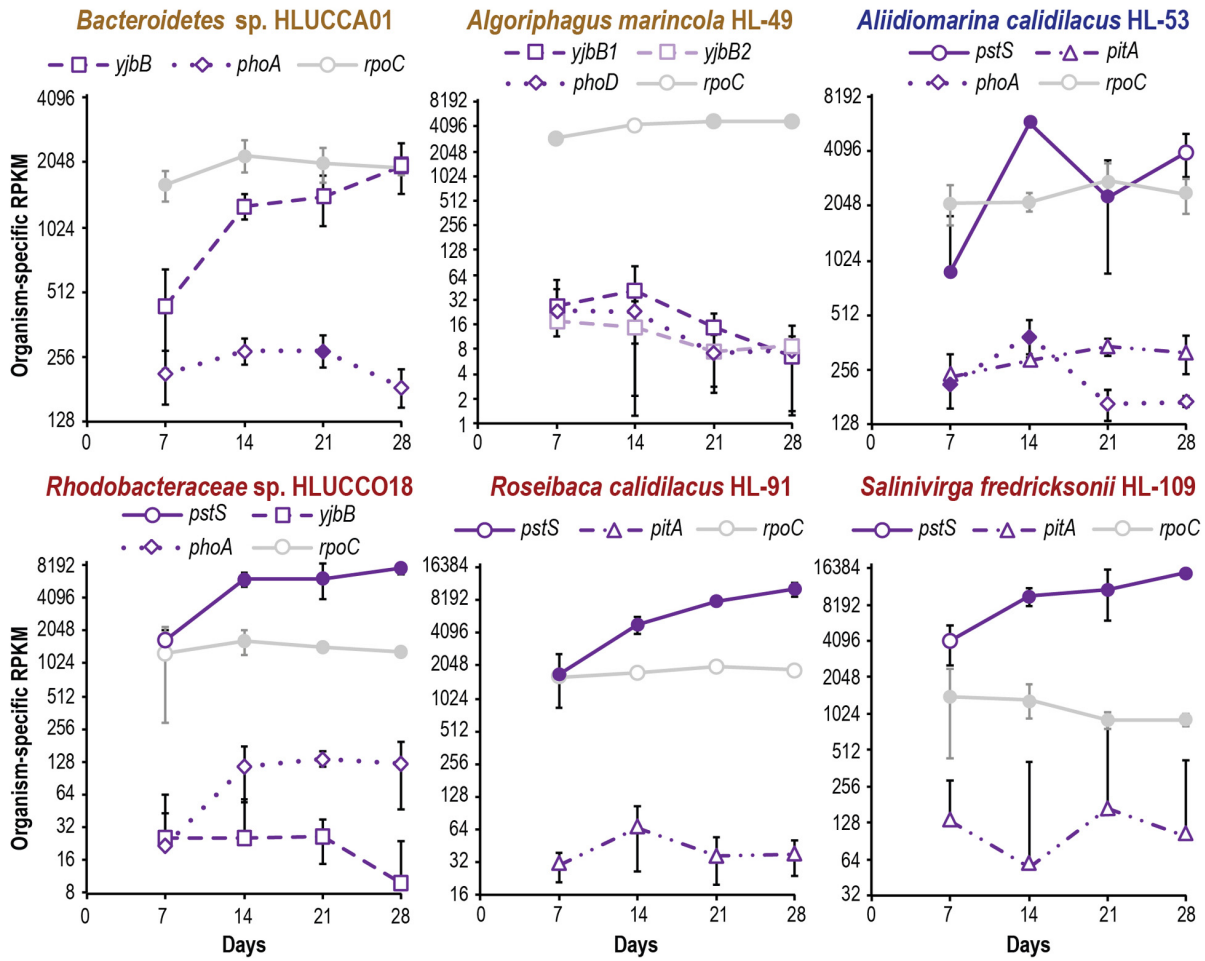


FIGURE 4 | Metatranscriptomic evaluation of phosphate acquisition gene expression in major heterotrophic species during succession. Expression of the substrate-binding subunit *pstS* is plotted as an indicative gene for *pstSABC*, as other transporters and phosphates are each encoded by a single gene. In organisms where they are present, the *pst* genes are operonic and expression of *pstABC* displayed similar trends as *pstS* (**Supplementary Table S4**). Organism-specific RPKM is plotted on a $\log_2$ scale; error bars indicate 95% confidence interval of the mean (where the 95% confidence interval is inclusive of zero or below, negative error bars are not shown). Filled symbols indicate that two or more peptides were observed for a gene's product at a given time point; open symbols denote no or inadequate proteomic evidence. Expression of *rpoC* is included as a reference.

nitrate were very low throughout the growth period and provided no conclusive evidence of *nasA* or *nirB* expression; consequently, the majority of community nitrate assimilation was assumed to be cyanobacterial. We therefore used expression of genes involved in ammonium uptake (*amt*-family transporters) and incorporation (glutamine synthetases *glnA* and *glnT*) as markers of heterotrophic nitrogen starvation, as these genes are known to be induced under nitrogen limitation in model organisms (Hervás et al., 2008; Zimmer et al., 2000). In *Rhodobacteraceae* spp. HLUCCO18 and *R. calidilacus* HL-91, *glnA* and *amt* were tightly co-expressed (**Figure 5**) but were decoupled in *Bacteroidetes* sp. HLUCCA01, *A. calidilacus* HL-53, and *S. fredricksonii* HL-109. Both *Bacteroidetes* sp. HLUCCA01 and *S. fredricksonii* HL-109 exhibited high and stable expression of *glnA* and high and declining transcription of *amt*, respectively, suggesting nitrogen limitation for these species was greatest early in succession and declined thereafter. Coexpression of *amt* and the major *glnA* genes in the *Rhodobacteraceae* spp. alternated between highs on days 7 and 21 and lows on days 14 and 28, suggestive of alternating nitrogen-deplete and -replete conditions for these two species. Alternate glutamine synthetase genes (*glnA2, glnA3,* and *glnT*) in *Rhodobacteraceae* sp. HLUCCO18 were poorly expressed and did not show significant variation over time. Similarly, ammonium transporters were poorly expressed in *A. marincola* HL-49 and *A. calidilacus* HL-53 despite consistently high expression of *glnA* in HL-49 and initially high expression in HL-53. Expression of the *A. calidilacus glnA* declined precipitously (~25-fold) by day 28, suggesting dramatically reduced ammonium incorporation late in the succession period. These data suggested that *A. marincola* and *A. calidilacus* may acquire their nitrogen from nitrogenous organic compounds provided by other members rather than direct incorporation of deamination-derived ammonium, and that *A. calidilacus* may switch its major nitrogen source

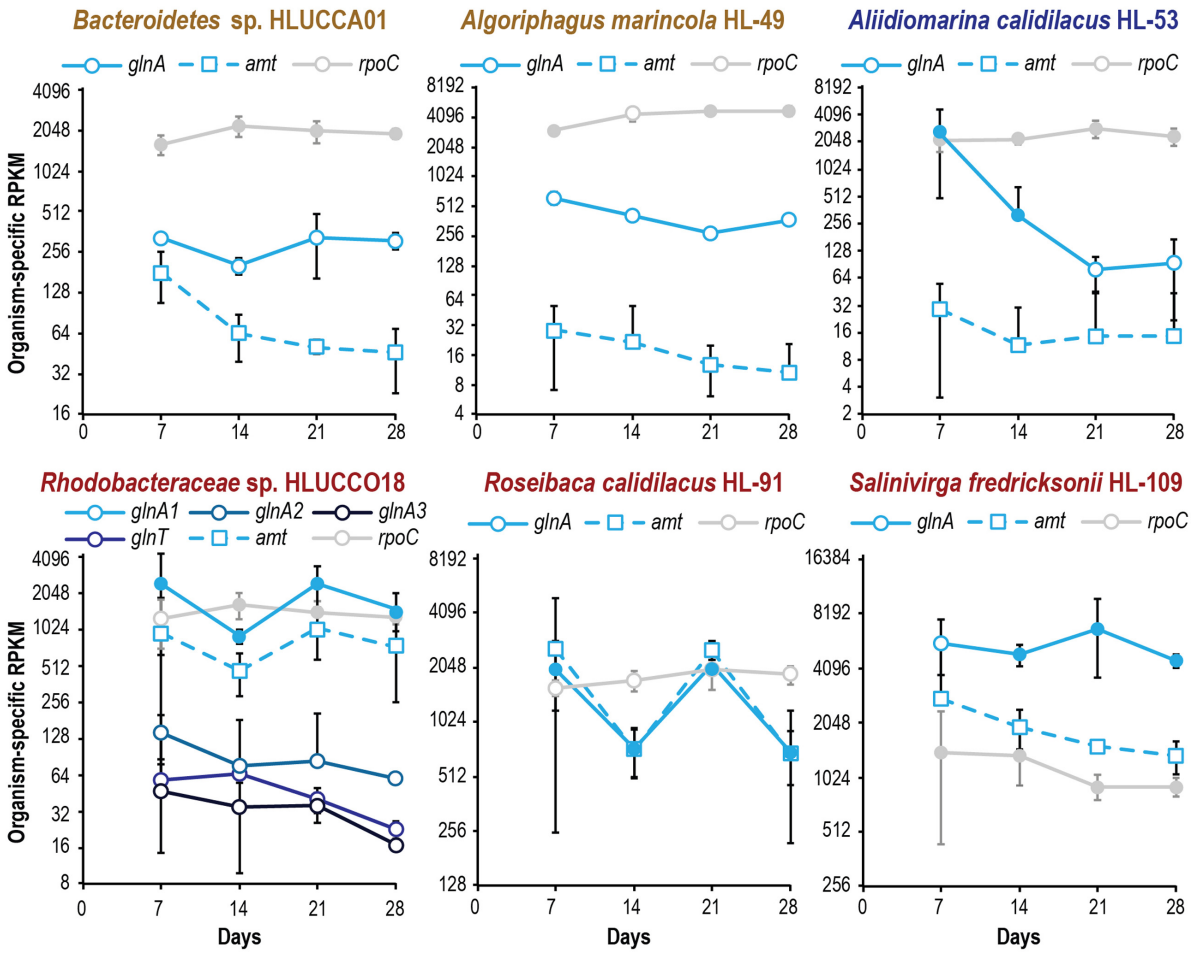


FIGURE 5 | Metatranscriptomic evaluation of ammonium incorporation gene expression in major heterotrophic species reveals species-specific patterns of nitrogen limitation during succession. Organism-specific RPKM is plotted on a $\log_2$ scale; error bars indicate 95% confidence interval of the mean (where the 95% confidence interval is inclusive of zero or below, negative error bars are not shown). Filled symbols indicate that two or more peptides were observed for a gene's product at a given time point; open symbols denote no or inadequate proteomic evidence. Expression of *rpoC* is included as a reference.

during succession. Taken together, these data suggest that, in contrast to response to phosphate limitation observed at the whole community level, nitrogen limitation is highly specific to individual species and impacts species differentially during succession.

Linking Expression Patterns to Energy and Nutrient Cycling during Succession

Although community-level functional potential and gene expression patterns in aquatic, phototrophic biofilms have been examined previously (Klatt et al., 2013; Leary et al., 2014; Graham et al., 2015; Sanli et al., 2015), combining species-resolved metagenomics with metatranscriptomics and metaproteomics provides a new approach for predicting the dynamics of energy and nutrient cycling at the level of individual species within a community context. This approach predicts the routes by which energy and elements enter communities, which is key to understanding how their interactions influence community dynamics and biogeochemical cycles. In this work, we used a tractable, phototrophic consortium, for which complete, species-resolved genomic information is available, to predict which species could serve as net "importers" of community energy and elemental resources and to examine species-specific energy and nutrient responses over biofilm succession. As the UCC biofilms are closed systems except for light energy input and gas exchange, each replicate can be treated as an individual microecosystem with respect to nutrient cycling (Gorden et al., 1969). Despite sequential passage in medium containing only inorganic macronutrients, we found a surprising lack of genomic potential in the dominant community members to assimilate these resources. This likely reflects extensive metabolic interdependency among species, especially heterotroph dependence upon the cyanobacterium for acquisition of all macronutrients (excepting phosphorus).

We evaluated individual-member functional responses during succession using metatranscriptomic and metaproteomic

analyses of identical samples. Although the two approaches agreed well on the direction and magnitude of within-organism changes in gene expression activity, they disagreed sharply in their estimation of relative activity across organisms. Interestingly, these patterns displayed taxon-specificity, with *Bacteroidetes* members exhibiting substantially greater relative abundances (∼10-fold) in the metatranscriptome than in the metaproteome. Transcript: protein abundance ratios were relatively stable over time, though member relative abundances in each measurement varied up to ∼10-fold, suggesting that these ratios are properties of the organisms and not solely the result of member abundances or growth rates (e.g., the transcript:protein ratio of *A. calidilacus* HL-53 remains stable despite a ∼10-fold reduction in relative abundance in each; **Supplementary Table S2**). Variation in this ratio potentially stems from differences in cell size and/or biomass, species-level variability in extraction efficiency for RNA versus protein, or organismal regulatory strategies; however, nearly 100-fold differences between transcript and protein abundances for some species suggests that extreme caution be exercised when using either approach to evaluate which microbial populations within a community are "active" (Siggins et al., 2012; Kolmeder and de Vos, 2014; Satinsky et al., 2015; Thureborn et al., 2016).

During succession, autotroph energy, carbon, nitrogen, and sulfur uptake decrease substantially, a phenomenon that has been long-observed in other phototrophic biofilms (Cooke, 1967). Phosphorus limitation in UCC-O by day 14, displayed by increased expression of phosphate acquisition genes across all members except *A. marincola* HL-49, likely caused a decline in the cyanobacterial growth rate and, consequently, the abundance of cyanobacterial inorganic carbon, nitrogen, and sulfur acquisition proteins. Despite this decline in acquisition of new, oxidized nitrogen (nitrate) and sulfur (sulfate) resources by *Phormidium* sp. OSCR late in succession, abundances of glutamine synthetase and cysteine synthase suggested stable incorporation of reduced nitrogen and sulfur into amino acids over the entire period. This suggests a shift late in succession from *de novo* synthesis toward cyanobacterial recycling of community reduced nitrogen and sulfur stores into amino acids. Notably, declines in cyanobacterial sulfate reduction were less extensive than for nitrate, suggesting the possibility of a loss process for bioavailable sulfur (i.e., biological or chemical sulfur oxidation) not present for nitrogen. Though transitions from open to closed biogeochemical cycles and reductions in energy capture as communities mature have been long-predicted (Odum, 1969), our data suggest that these transitions are mediated by divergent gene expression responses at the level of individual species. In this study, phosphate limitation appeared to usher in the transition from early successional phases with relatively high cellular growth rates and correspondingly high acquisition of oxidized resources (e.g., nitrate, sulfate) to more mature, slower growth phases where recycling of reduced forms appears to dominate. Though the mechanism for phosphate limitation in Hot Lake and its derived cultures (insolubility due to high Mg^{2+}) is unusual, it should be noted that phosphorus scarcity is common in periphyton biofilms (Rejmánková and Komárková, 2000; McCormick et al., 2001; Borovec et al., 2010; Hagerthey et al., 2011).

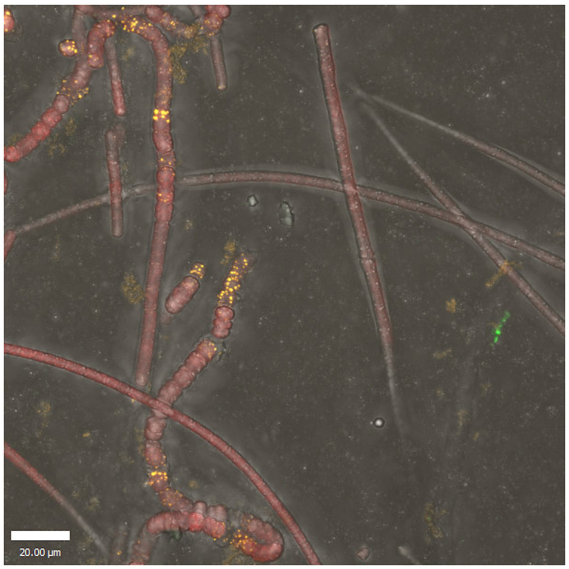


FIGURE 6 | Localization of *A. marincola* HL-49 within UCC-O biofilms via fluorescence *in situ* hybridization. Image represents a maximum-intensity Z-projection of a stack of confocal images, scale bar denotes 20 mm. Fluorescence intensity from FISH probes targeted against *A. marincola* HL-49 16S ribosomal RNA is depicted in yellow. Cyanobacterial chlorophyll *a* and phycocyanin auto-fluorescence intensity appear in red within filaments of *Phormidium* sp. OSCR. Fluorescence images are overlaid upon a differential interference contrast image of consortium biomass to display cell boundaries.

The gene expression patterns of heterotrophic and photoheterotrophic consortium members mirrored those of the cyanobacterium with respect to energy, sulfur, and phosphorus acquisition, but diverged with respect to nitrogen acquisition. Species-specific expression of macronutrient acquisition genes was largely consistent with prior taxonomy-based predictions of member functional roles (Cole et al., 2014). It is noteworthy that the only organism that does not show elevated transcription of phosphate-acquisition genes during succession is *A. marincola* HL-49, which displays high *glnA* but very low *amt* expression. This suggests a detritivorous role in which it acquires its phosphorus from detrital nucleic acids. This suggestion is somewhat supported by the preferential localization of *A. marincola* HL-49 in UCC biofilms to cyanobacterial cells that appear to have compromised membrane integrity, lacking significant photopigment fluorescence and phase contrast (**Figure 6**). If true, this may make *A. marincola* HL-49 a potential keystone species for liberation of otherwise-inaccessible phosphorus resources. Nucleic acid turnover may therefore be an ecological role important for maximizing sustainable biomass in periphyton biofilms through increases in the velocity of phosphorus cycling.

Similarly to *A. marincola* HL-49, the hypothesized protein degrader *A. calidilacus* HL-53 (Hou et al., 2004) also exhibits low *amt* but initially high *glnA* expression, suggesting consumption of nitrogen-containing molecules (e.g., amino acids) as carbon

and energy sources. As both *A. marincola* (Yoon et al., 2004) and *A. calidilacus* are likely to express extracellular proteases, either of these organisms may facilitate nitrogen availability to other members via deamination. A combination of declining population sizes and ammonium production of *A. calidilacus* HL-53 could also contribute to the day 21 nitrogen limitation of the *Rhodobacteraceae* spp. HLUCCO18 and HL-91. We hypothesize that differences in expression of ammonium incorporation genes across community members reflect species-specific consumption of different nitrogen sources that fluctuate in availability.

The predicted inability of many heterotrophs in the consortia to access oxidized forms of macronutrients, therefore relying upon metabolic exchange to supply these elements, concurs with our recent demonstration of a similar lack of acquisition systems in genomes reconstructed from Hot Lake-derived metagenomes (Mobberley et al., 2017). Of the 34 genome reconstructions reported in this study, 18 appeared to lack the ability to generate ammonium from more oxidized nitrogen sources and 23 lacked genes required for sulfate assimilation. That these genes are missing even within complete genomes from UCC heterotrophs (Nelson et al., 2016) further suggests that these functions are not lacking in mat metagenome-derived genome reconstructions solely due to incomplete genome information. Organisms lacking these functions were substantially more common at depth in the mat compared with overlying strata in closer communication with the water column (Mobberley et al., 2017). This is expected in that reduced forms of macronutrients (i.e., nitrogen, sulfur) that otherwise might be oxidized by other microbes as energy sources should be more stable in regions of the mat that are suboxic for at least some period. However, the data presented here suggest that stable metabolic interactions, in which heterotrophs are completely dependent upon other organisms for reduced macronutrients, occur reproducibly in ∼100 μm-thick phototrophic biofilms exposed to continuous light. We therefore submit that the ecological importance of such obligate interactions in phototrophic biofilms is not necessarily reflected by the fraction of reconstructed genomes predicted to lack acquisition systems. Furthermore, this study suggests that organismal specialization around different nitrogenous compounds may be a mechanism supporting maintenance of diversity as magnesium levels increase in Hot Lake and reduce phosphorous bioavailability.

Taken together, our results suggest the heterotrophs in UCC-O are generally phosphorus-replete but nitrogen-limited early in succession, after which phosphorus becomes limiting. They also suggest that the availability of distinct nitrogen sources might partition and structure diverse phototrophic microbial communities when phosphate, for which many organisms must directly compete, is sparingly available. In phosphorus-limited periphyton biofilms, niche partitioning around nitrogen sources may therefore help maintain the high diversity of these communities (Larson et al., 2016). As genome reconstruction from metagenome data becomes more accurate and omics measurements more sensitive and quantitative, it will enable similar studies in multiple phototrophic biofilm systems in the field to determine whether niche partitioning around nitrogen sources is a common community response to phosphate limitation in natural systems.

AUTHOR CONTRIBUTIONS

SL, JF, and MR designed the experiment, SL, JM, WN, and MR generated genome predictions, SL and JC performed succession experiments, LM, RT, and HW performed metatranscriptomic analyses, EH and ML performed metaproteomic analyses, SL, JM, and MR analyzed genome-resolved expression data, and all authors contributed to the manuscript.

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL) and from the Purdue University Departments of Food Science and Nutrition Science through support for SL. The work conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and Community Sequencing Project 701. Transcriptomics and MS-based measurements were performed in the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER at PNNL. The authors further thank Beau Morton and Karl Dana for their assistance in handling cultures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01020/full#supplementary-material

TABLE S1 | Accession numbers for unicyanobacterial consortium isolate genomes and genome reconstructions.

TABLE S2 | Member genes predicted to be involved in energy or macronutrient acquisition.

TABLE S3 | Per-sample RNA read and peptide abundances attributed to each consortium member genome.

TABLE S4 | Expression of energy and macronutrient acquisition genes by each consortium member during succession.

REFERENCES

Amann, R., and Fuchs, B. M. (2008). Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* 6, 339–348. doi: 10.1038/nrmicro1888

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Balashov, S. P., Imasheva, E. S., Dioumaev, A. K., Wang, J. M., Jung, K.-H., and Lanyi, J. K. (2014). Light-driven Na+ pump from *Gillisia limnaea*: a high-affinity Na+ binding site is formed transiently in the photocycle. *Biochemistry (Mosc)* 53, 7549–7561. doi: 10.1021/bi501064n

Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M., and Packmann, A. I. (2016). The ecology and biogeochemistry of stream biofilms. *Nat. Rev. Microbiol.* 14, 251–263. doi: 10.1038/nrmicro.2016.15

Battin, T. J., Kaplan, L. A., Denis Newbold, J., and Hansen, C. M. E. (2003). Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature* 426, 439–442. doi: 10.1038/nature02152

Battin, T. J., Kaplan, L. A., Findlay, S., Hopkinson, C. S., Marti, E., Packman, A. I., et al. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nat. Geosci.* 1, 95–100. doi: 10.1038/ngeo101

Borovec, J., Sirová, D., Mošnerová, P., Rejmánková, E., and Vrba, J. (2010). Spatial and temporal changes in phosphorus partitioning within a freshwater cyanobacterial mat community. *Biogeochemistry* 101, 323–333. doi: 10.1007/s10533-010-9488-4

Bottomley, P. J., Grillo, J. F., Baalen, C. V., and Tabita, F. R. (1979). Synthesis of nitrogenase and heterocysts by *Anabaena* sp. CA in the presence of high levels of ammonia. *J. Bacteriol.* 140, 938–943.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5:8365. doi: 10.1038/srep08365

Callister, S. J., Nicora, C. D., Zeng, X., Roh, J. H., Dominguez, M. A., Tavano, C. L., et al. (2006). Comparison of aerobic and photosynthetic *Rhodobacter sphaeroides* 2.4.1 proteomes. *J. Microbiol. Methods* 67, 424–436. doi: 10.1016/j.mimet.2006.04.021

Carré-Mlouka, A., Méjean, A., Quillardet, P., Ashida, H., Saito, Y., Yokota, A., et al. (2006). A new rubisco-like protein coexists with a photosynthetic rubisco in the planktonic cyanobacteria microcystis. *J. Biol. Chem.* 281, 24462–24471. doi: 10.1074/jbc.M602973200

Cole, J. K., Hutchison, J. R., Renslow, R. S., Kim, Y.-M., Chrisler, W. B., Engelmann, H. E., et al. (2014). Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Front. Syst. Microbiol.* 5:109. doi: 10.3389/fmicb.2014.00109

Cooke, G. D. (1967). The pattern of autotrophic succession in laboratory microcosms. *BioScience* 17, 717–721. doi: 10.2307/1294089

Erb, T. J., Evans, B. S., Cho, K., Warlick, B. P., Sriram, J., Wood, B. M., et al. (2012). A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nat. Chem. Biol.* 8, 926–932. doi: 10.1038/nchembio.1087

Flipo, N., Even, S., Poulin, M., Tusseau-Vuillemin, M.-H., Ameziane, T., and Dauta, A. (2004). Biogeochemical modelling at the river scale: plankton and periphyton dynamics: grand morin case study, France. *Ecol. Model* 176, 333–347. doi: 10.1016/j.ecolmodel.2004.01.012

Fuhrman, J. A., Schwalbach, M. S., and Stingl, U. (2008). Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.* 6, 488–494. doi: 10.1038/nrmicro1893

Gorden, R. W., Beyers, R. J., Odum, E. P., and Eagon, R. G. (1969). Studies of a simple laboratory microecosystem: bacterial activities in a heterotrophic succession. *Ecology* 50, 86–100. doi: 10.2307/1934666

Graham, E. B., Wieder, W. R., Leff, J. W., Weintraub, S. R., Townsend, A. R., Cleveland, C. C., et al. (2014). Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biol. Biochem.* 68, 279–282. doi: 10.1016/j.soilbio.2013.08.023

Graham, L. E., Knack, J. J., Graham, M. E., Graham, J. M., and Zulkifly, S. (2015). A metagenome for lacustrine *Cladophora* (Cladophorales) reveals remarkable diversity of eukaryotic epibionts and genes relevant to materials cycling. *J. Phycol.* 51, 408–418. doi: 10.1111/jpy.12296

Hagerthey, S. E., Bellinger, B. J., Wheeler, K., Gantar, M., and Gaiser, E. (2011). Everglades periphyton: a biogeochemical perspective. *Crit. Rev. Environ. Sci. Technol.* 41, 309–343. doi: 10.1080/10643389.2010.531218

Hervás, A. B., Canosa, I., and Santero, E. (2008). Transcriptome analysis of *Pseudomonas putida* in response to nitrogen availability. *J. Bacteriol.* 190, 416–420. doi: 10.1128/JB.01230-07

Hess, B. M., Xue, J., Markillie, L. M., Taylor, R. C., Wiley, H. S., Ahring, B. K., et al. (2013). Coregulation of terpenoid pathway genes and prediction of isoprene production in *Bacillus subtilis* using transcriptomics. *PLoS ONE* 8:e66104. doi: 10.1371/journal.pone.0066104

Hou, S., Saw, J. H., Lee, K. S., Freitas, T. A., Belisle, C., Kawarabayasi, Y., et al. (2004). Genome sequence of the deep-sea γ-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl. Acad. Sci. U.S.A.* 101, 18036–18041. doi: 10.1073/pnas.0407638102

Hugerth, L. W., Larsson, J., Alneberg, J., Lindh, M. V., Legrand, C., Pinhassi, J., et al. (2015). Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* 16:279. doi: 10.1186/s13059-015-0834-7

Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11:431. doi: 10.1186/1471-2105-11-431

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: kegg tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kautza, A., Mazeika, S., and Sullivan, P. (2016). The energetic contributions of aquatic primary producers to terrestrial food webs in a mid-size river system. *Ecology* 97, 694–705.

Kim, S., Gupta, N., and Pevzner, P. A. (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* 7, 3354–3363. doi: 10.1021/pr8001244

Kirchman, D. L., and Hanson, T. E. (2013). Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ. Microbiol. Rep.* 5, 188–199. doi: 10.1111/j.1758-2229.2012.00367.x

Klatt, C. G., Liu, Z., Ludwig, M., Kühl, M., Jensen, S. I., Bryant, D. A., et al. (2013). Temporal metatranscriptomic patterning in phototrophic *Chloroflexi* inhabiting a microbial mat in a geothermal spring. *ISME J.* 7, 1775–1789. doi: 10.1038/ismej.2013.52

Kolmeder, C. A., and de Vos, W. M. (2014). Metaproteomics of our microbiome — Developing insight in function and activity in man and model systems. *J. Proteomics* 97, 3–16. doi: 10.1016/j.jprot.2013.05.018

Konopka, A., Lindemann, S., and Fredrickson, J. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. *ISME J.* 9, 1488–1495. doi: 10.1038/ismej.2014.251

Kwon, S.-K., Kim, B. K., Song, J. Y., Kwak, M.-J., Lee, C. H., Yoon, J.-H., et al. (2013). Genomic makeup of the marine flavobacterium nonlabens (Donghaeana) dokdonensis and identification of a novel class of rhodopsins. *Genome Biol. Evol.* 5, 187–199. doi: 10.1093/gbe/evs134

Larson, C. A., Adumatioge, L., and Passy, S. I. (2016). The number of limiting resources in the environment controls the temporal diversity patterns in the algal benthos. *Microb. Ecol.* 72, 64–69. doi: 10.1007/s00248-016-0741-9

Leary, D. H., Li, R. W., Hamdan, L. J., Iv, W. J. H., Lebedev, N., Wang, Z., et al. (2014). Integrated metagenomic and metaproteomic analyses of marine biofilm communities. *Biofouling* 30, 1211–1223. doi: 10.1080/08927014.2014.977267

Lebens, M., Lundquist, P., Söderlund, L., Todorovic, M., and Carlin, N. I. A. (2002). The nptA gene of *Vibrio cholerae* encodes a functional sodium-dependent phosphate cotransporter homologous to the type II cotransporters of eukaryotes. *J. Bacteriol.* 184, 4466–4474. doi: 10.1128/JB.184.16.4466-4474.2002

Leyh, T. S., Taylor, J. C., and Markham, G. D. (1988). The sulfate activation locus of *Escherichia coli* K12: cloning, genetic, and enzymatic characterization. *J. Biol. Chem.* 263, 2409–2416.

Lindemann, S. R., Moran, J. J., Stegen, J. C., Renslow, R. S., Hutchison, J. R., Cole, J. K., et al. (2013). The epsomitic phototrophic microbial mat of Hot Lake, Washington: community structural responses to seasonal cycling. *Front. Microbiol.* 4:323. doi: 10.3389/fmicb.2013.00323

Ma, Z.-Q., Dasari, S., Chambers, M. C., Litton, M. D., Sobecki, S. M., Zimmerman, L. J., et al. (2009). IDPicker 2.0: improved protein assembly with high

discrimination peptide identification filtering. *J. Proteome Res.* 8, 3872–3881. doi: 10.1021/pr900360j

Mackerras, A. H., and Smith, G. D. (1986). Evidence for direct repression of nitrogenase by ammonia in the cyanobacterium *Anabaena cylindrica*. *Biochem. Biophys. Res. Commun.* 134, 835–844. doi: 10.1016/S0006-291X(86)80496-X

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., et al. (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 42, D560–D567. doi: 10.1093/nar/gkt963

Mayampurath, A. M., Jaitly, N., Purvine, S. O., Monroe, M. E., Auberry, K. J., Adkins, J. N., et al. (2008). DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* 24, 1021–1023. doi: 10.1093/bioinformatics/btn063

McCarren, J., and DeLong, E. F. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ. Microbiol.* 9, 846–858. doi: 10.1111/j.1462-2920.2006.01203.x

McCormick, P. V., O'Dell, M. B., Shuford, I. I. I. R. B. E., Backus, J. G., and Kennedy, W. C. (2001). Periphyton responses to experimental phosphorus enrichment in a subtropical wetland. *Aquat. Bot.* 71, 119–139. doi: 10.1016/S0304-3770(01)00175-9

Mobberley, J. M., Lindemann, S. R., Bernstein, H. C., Moran, J. J., Renslow, R. S., Babauta, J., et al. (2017). Organismal and spatial partitioning of energy and macronutrient transformations within a hypersaline mat. *FEMS Microbiol. Ecol.* doi: 10.1093/femsec/fix028 [Epub ahead of print].

Moran, J. J., Doll, C. G., Bernstein, H. C., Renslow, R. S., Cory, A. B., Hutchison, J. R., et al. (2014). Spatially tracking 13C-labelled substrate (bicarbonate) accumulation in microbial communities using laser ablation isotope ratio mass spectrometry. *Environ. Microbiol. Rep.* 6, 786–791. doi: 10.1111/1758-2229.12211

Moreno-Vivian, C., Cabello, P., Martinez-Luque, M., Blasco, R., and Castillo, F. (1999). Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J. Bacteriol.* 181, 6573–6584.

Nazaries, L., Pan, Y., Bodrossy, L., Baggs, E. M., Millard, P., Murrell, J. C., et al. (2013). Evidence of microbial regulation of biogeochemical cycles from a study on methane flux and land use change. *Appl. Environ. Microbiol.* 79, 4031–4040. doi: 10.1128/AEM.00095-13

Nelson, W. C., Maezato, Y., Wu, Y.-W., Romine, M. F., and Lindemann, S. R. (2016). Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. *Appl. Environ. Microbiol.* 82, 255–267. doi: 10.1128/AEM.02274-15

Odum, E. P. (1969). The strategy of ecosystem development. *Science* 164, 262–270. doi: 10.1126/science.164.3877.262

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Palomo, A., Jane Fowler, S., Gülay, A., Rasmussen, S., Sicheritz-Ponten, T., and Smets, B. F. (2016). Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of *Nitrospira* spp. *ISME J.* doi: 10.1038/ismej.2016.63 [Epub ahead of print]. doi: 10.1038/ismej.2016.63

Petyuk, V. A., Mayampurath, A. M., Monroe, M. E., Polpitiya, A. D., Purvine, S. O., Anderson, G. A., et al. (2010). DtaRefinery, a software tool for elimination of systematic errors from parent ion mass measurements in tandem mass spectra data sets. *Mol. Cell. Proteomics* 9, 486–496. doi: 10.1074/mcp.M900217-MCP200

Pope, P. B., Smith, W., Denman, S. E., Tringe, S. G., Barry, K., Hugenholtz, P., et al. (2011). Isolation of succinivibrionaceae implicated in low methane emissions from tammar wallabies. *Science* 333, 646–648. doi: 10.1126/science.1205760

Pranckevičiene, E., Rančelis, T., Pranculis, A., and Kučinskas, V. (2015). Challenges in exome analysis by LifeScope and its alternative computational pipelines. *BMC Res. Notes* 8:421. doi: 10.1186/s13104-015-1385-4

Rao, N. N., and Torriani, A. (1990). Molecular aspects of phosphate transport in *Escherichia coli*. *Mol. Microbiol.* 4, 1083–1090. doi: 10.1111/j.1365-2958.1990.tb00682.x

Rathgeber, C., Alric, J., Hughes, E., Verméglio, A., and Yurkov, V. (2012). The photosynthetic apparatus and photoinduced electron transfer in the aerobic phototrophic bacteria *Roseicyclus mahoneyensis* and *Porphyrobacter meromictius*. *Photosynth. Res.* 110, 193–203. doi: 10.1007/s11120-011-9718-1

Rejmánková, E., and Komárková, J. (2000). A function of cyanobacterial mats in phosphorus-limited tropical wetlands. *Hydrobiologia* 431, 135–153. doi: 10.1023/A:1004011318643

Roeselers, G., Loosdrecht, M. C. M., and van Muyzer, G. (2008). Phototrophic biofilms and their potential applications. *J. Appl. Phycol.* 20, 227–235. doi: 10.1007/s10811-007-9223-2

Rousk, J., and Bengtson, P. (2014). Microbial regulation of global biogeochemical cycles. *Terr. Microbiol.* 5:103. doi: 10.3389/978-2-88919-297-7

Sanli, K., Bengtsson-Palme, J., Nilsson, R. H., Kristiansson, E., Alm Rosenblad, M., Blanck, H., et al. (2015). Metagenomic sequencing of marine periphyton: taxonomic and functional insights into biofilm communities. *Front. Microbiol.* 6:1192. doi: 10.3389/fmicb.2015.01192

Satinsky, B. M., Fortunato, C. S., Doherty, M., Smith, C. B., Sharma, S., Ward, N. D., et al. (2015). Metagenomic and metatranscriptomic inventories of the lower Amazon River. *Microbiome* 3:39. doi: 10.1186/s40168-015-0099-0

Siggins, A., Gunnigle, E., and Abram, F. (2012). Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol. Ecol.* 80, 265–280. doi: 10.1111/j.1574-6941.2011.01284.x

Smith, J. S., Angel, T. E., Chavkin, C., Orton, D. J., Moore, R. J., and Smith, R. D. (2014). Characterization of individual mouse cerebrospinal fluid proteomes. *Proteomics* 14, 1102–1106. doi: 10.1002/pmic.201300241

Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A.-J. L., Bunk, D. M., et al. (2010). Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* 9, 761–776. doi: 10.1021/pr9006365

Tabita, F. R., Hanson, T. E., Li, H., Satagopan, S., Singh, J., and Chan, S. (2007). Function, structure, and evolution of the rubisCO-Like proteins and their RubisCO homologs. *Microbiol. Mol. Biol. Rev.* 71, 576–599. doi: 10.1128/MMBR.00015-07

Tang, K.-H., Feng, X., Tang, Y. J., and Blankenship, R. E. (2009). Carbohydrate Metabolism and carbon fixation in *Roseobacter denitrificans* OCh114. *PLoS ONE* 4:e7233. doi: 10.1371/journal.pone.0007233

Taylor, R. C., Robertson, B.-J. M. W., Markillie, L. M., Serres, M. H., Linggi, B. E., Aldrich, J. T., et al. (2013). Changes in translational efficiency is a dominant regulatory mechanism in the environmental response of bacteria. *Integr. Biol.* 5, 1393–1406. doi: 10.1039/c3ib40120k

Thureborn, P., Franzetti, A., Lundin, D., and Sjöling, S. (2016). Reconstructing ecosystem functions of the active microbial community of the Baltic Sea oxygen depleted sediments. *PeerJ* 4:e1593. doi: 10.7717/peerj.1593

Vadeboncoeur, Y., and Steinman, A. D. (2002). Periphyton function in lake ecosystems. *Sci. World J.* 2, 1449–1468. doi: 10.1100/tsw.2002.294

Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, 11033–11033. doi: 10.1093/nar/gkw880

Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisolm, S. W. (2012). Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS ONE* 7:e43432. doi: 10.1371/journal.pone.0043432

Wagner, K., Besemer, K., Burns, N. R., Battin, T. J., and Bengtsson, M. M. (2015). Light availability affects stream biofilm bacterial community composition and function, but not diversity. *Environ. Microbiol.* 17, 5036–5047. doi: 10.1111/1462-2920.12913

Willsky, G. R., and Malamy, M. H. (1980). Characterization of two genetically separable inorganic phosphate transport systems in *Escherichia coli*. *J. Bacteriol.* 144, 356–365.

Wright, E. S., Yilmaz, L. S., Corcoran, A. M., Ökten, H. E., and Noguera, D. R. (2014). Automated design of probes for rRNA-targeted fluorescence in situ hybridization reveals the advantages of using dual probes for accurate identification. *Appl. Environ. Microbiol.* 80, 5124–5133. doi: 10.1128/AEM.01685-14

Yoon, J.-H., Yeo, S.-H., and Oh, T.-K. (2004). *Hongiella marincola* sp. nov., isolated from sea water of the East Sea in Korea. *Int. J. Syst. Evol. Microbiol.* 54, 1845–1848. doi: 10.1099/ijs.0.63170-0

Yurkov, V., and Csotonyi, J. T. (2009). "New light on aerobic anoxygenic phototrophs," in *Advances in Photosynthesis and Respiration. The Purple*

Phototrophic Bacteria, eds C. N. Hunter, F. Daldal, M. C. Thurnauer, and J. T. Beatty (Dordrecht: Springer), 31–55. doi: 10.1007/978-1-4020-8815-5_3

Yurkov, V. V., and Beatty, J. T. (1998). Aerobic anoxygenic phototrophic bacteria. *Microbiol. Mol. Biol. Rev.* 62, 695–724.

Zachara, J. M., Moran, J. J., Resch, C. T., Lindemann, S. R., Felmy, A. R., Bowden, M. E., et al. (2016). Geo- and biogeochemical processes in a heliothermal hypersaline lake. *Geochim. Cosmochim. Acta* 181, 144–163. doi: 10.1016/j.gca.2016.02.001

Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., et al. (2000). Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14674–14679. doi: 10.1073/pnas.97.26.14674

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Lindemann, Mobberley, Cole, Markillie, Taylor, Huang, Chrisler, Wiley, Lipton, Nelson, Fredrickson and Romine. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 05 December 2017
doi: 10.3389/fmicb.2017.02384

Niche Partitioning of the N Cycling Microbial Community of an Offshore Oxygen Deficient Zone

Clara A. Fuchsman, Allan H. Devol, Jaclyn K. Saunders, Cedar McKay and Gabrielle Rocap**

School of Oceanography, University of Washington, Seattle, WA, United States

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Alyson E. Santoro, University of California, Santa Barbara, United States
Katharina Kujala, University of Oulu, Finland

***Correspondence:**
Clara A. Fuchsman
cfuchsm1@u.washington.edu
Gabrielle Rocap
rocap@u.washington.edu

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *15 August 2017*
Accepted: *20 November 2017*
Published: *05 December 2017*

Citation:
Fuchsman CA, Devol AH, Saunders JK, McKay C and Rocap G (2017) Niche Partitioning of the N Cycling Microbial Community of an Offshore Oxygen Deficient Zone. Front. Microbiol. 8:2384. doi: 10.3389/fmicb.2017.02384

Microbial communities in marine oxygen deficient zones (ODZs) are responsible for up to half of marine N loss through conversion of nutrients to N_2O and N_2. This N loss is accomplished by a consortium of diverse microbes, many of which remain uncultured. Here, we characterize genes for all steps in the anoxic N cycle in metagenomes from the water column and >30 μm particles from the Eastern Tropical North Pacific (ETNP) ODZ. We use an approach that allows for both phylogenetic identification and semi-quantitative assessment of gene abundances from individual organisms, and place these results in context of chemical measurements and rate data from the same location. Denitrification genes were enriched in >30 μm particles, even in the oxycline, while anammox bacteria were not abundant on particles. Many steps in denitrification were encoded by multiple phylotypes with different distributions. Notably three N_2O reductases (*nosZ*), each with no cultured relative, inhabited distinct niches; one was free-living, one dominant on particles and one had a C terminal extension found in autotrophic S-oxidizing bacteria. At some depths >30% of the community possessed nitrite reductase *nirK*. A *nirK* OTU linked to SAR11 explained much of this abundance. The only bacterial gene found for NO reduction to N_2O in the ODZ was a form of *qnorB* related to the previously postulated "nitric oxide dismutase," hypothesized to produce N_2 directly while oxidizing methane. However, similar *qnorB-like* genes are also found in the published genomes of many bacteria that do not oxidize methane, and here the *qnorB-like* genes did not correlate with the presence of methane oxidation genes. Correlations with N_2O concentrations indicate that these *qnorB-like* genes likely facilitate NO reduction to N_2O in the ODZ. In the oxycline, *qnorB-like* genes were not detected in the water column, and estimated N_2O production rates from ammonia oxidation were insufficient to support the observed oxycline N_2O maximum. However, both *qnorB-like* and *nosZ* genes were present within particles in the oxycline, suggesting a particulate source of N_2O and N_2. Together, our analyses provide a holistic view of the diverse players in the low oxygen nitrogen cycle.

Keywords: denitrification, anammox, Eastern Tropical North Pacific, oxygen minimum zone, particles

INTRODUCTION

Naturally occurring oxygen deficient zones (ODZs), defined here as water containing <10 nM oxygen, constitute <1% of the ocean volume, but contribute 30–50% of N loss from the marine system through N_2 production (DeVries et al., 2013). In the absence of oxygen, microbes in these waters use a variety of different terminal electron acceptors, including oxidized nitrogen (nitrate and nitrite), resulting in the production of both N_2 and the greenhouse gas N_2O. Due to the temperature-dependent solubility of oxygen, concentrations are predicted to be further reduced in the future, increasing the size of ODZs (Deutsch et al., 2011) with implications for increased N_2 production and consequent N limitation of global primary production. Thus, understanding the microbial community responsible for N_2 production is important to predicting the nitrogen budget of a changing ocean.

Denitrification and anammox are the two known pathways for N_2 production. Denitrification is the multi-step reduction of NO_3^- to N_2 paired with oxidization of organic matter, reduced sulfur, or methane. Nitrate is reduced through four reduction steps, each encoded by a different enzyme. However, not all denitrifying organisms possess all the enzymes needed to reduce NO_3^- all the way to N_2. Many encode genes for only a subset of the steps and rely on other microbes for the production or consumption of their reactants and products (Zumft, 1997). This separation of individual reactions may be possible because most steps in denitrification occur in the periplasm (Zumft, 1997), which should ease transport of reactants and products across the outer membrane. Genes in the denitrification pathway are found widespread throughout the microbial tree of life (Zumft, 1997). Some horizontal gene transfer of denitrification genes has occurred and some denitrification genes have even been found on plasmids (Zumft, 1997). However, duplication and divergence of genes followed by incomplete lineage sorting maybe more important than horizontal gene transfer in accounting for the diversity of denitrification genes (Jones et al., 2008). Denitrifiers are also metabolically diverse, with heterotrophic denitrifiers coupling N reduction with organic matter oxidation and autotrophic denitrifiers coupling N reduction with sulfur or methane oxidation to fuel carbon fixation (Hannig et al., 2007; Ettwig et al., 2010; Babbin et al., 2014). In contrast, autotrophic bacteria in the order Brocadiales in the Planctomycetes phylum are the other organisms known to carry out the anammox process, which involves reduction of NO_2^- with NH_4^+ to produce N_2. The process occurs in an internal compartment where the membrane is composed of ladderane lipids; these tightly fitting lipids are needed to isolate the toxic intermediate hydrazine (N_2H_4) from the rest of the cell and prevent hydrazine leakage, which would make anammox energetically unfeasible (Sinninghe Damsté et al., 2002). This complex cellular structure may be one reason why anammox is restricted phylogenetically, with *Scalindua* the only genera of anammox bacteria thus far found in the marine environment (van de Vossenberg et al., 2013).

The relative importance of anammox compared to denitrification in marine N_2 production is still unclear (Lam et al., 2009; Ward et al., 2009; Dalsgaard et al., 2012). If the breakdown of organic matter by heterotrophic denitrifiers is the source of ammonia to anammox, then stoichiometrically heterotrophic denitrification should contribute 70% and anammox 30% to N_2 production (Devol, 2003) though the consumption of nitrite through nitrite oxidation can shift this ratio to >40% anammox (Penn et al., 2016). However, in older studies, denitrifiers, identified molecularly by targetting the nitrite reductase gene *nirS* were in low abundance (Lam et al., 2009, 2011; Jensen et al., 2011; Jayakumar et al., 2013; Kalvelage et al., 2013) and had low diversity (Bowen et al., 2015). More recently, using primer-independent approaches the copper-dependent nitrite reductase encoded by *nirK* was found in addition to *nirS*, and was more abundant (Ganesh et al., 2014, 2015; Glass et al., 2015; Lüke et al., 2016). In the Arabian Sea and ETSP, this *nirK* was predominantly from ammonia oxidizing archaea while in the ETNP *nirK* from nitrite oxidizer *Nitrospina* was found in addition (Glass et al., 2015; Lüke et al., 2016). However, neither ammonia oxidizing archaea or nitrite oxidizing bacteria are complete denitrifiers, so the role of *nirK* in N_2 production is still unclear.

Anammox and denitrifying bacteria may have different niches in the ODZ. Size fractionated studies in the ETSP and the Black Sea indicate that marine anammox bacteria are primarily free-living (Fuchsman et al., 2012b; Ganesh et al., 2014, 2015) as are nitrate reducers (Ganesh et al., 2015), but the last two steps in the denitrification pathway are enriched in >1.6 μm suspended particles (Ganesh et al., 2014, 2015). The most abundant nitrate reducer in the ETNP was a free-living SAR11 in the ODZ which has two different nitrate reductases transferred from quite distinct bacteria (gammaproteobacteria and candidate phyla OP1) but lacks genes for the last two steps in denitrification (Tsementzi et al., 2016). The organisms containing genes for the last step in denitrification, N_2O reductase *nosZ,* are largely unknown (Ganesh et al., 2014, 2015). The addition of sterilized sediment trap material significantly increased denitrification and anammox rates in all three marine ODZs (Babbin et al., 2014; Chang et al., 2014). Denitrification rates increase more than anammox (Babbin et al., 2014; Chang et al., 2014), probably because denitrification uses organic matter directly while anammox uses ammonia from organic matter degradation. N_2 production has also been found inside particles composed of diatom aggregates or zooplankton carcasses at hypoxic oxygen concentrations (Stief et al., 2016, 2017).

Denitrification in marine ODZs is generally attributed to heterotrophic denitrification. However, some autotrophic denitrification has been measured in incubations with sulfide in the coastal ETSP (Canfield et al., 2010) and autotrophic N_2 production by methane oxidizers has been proposed in the ETNP (Padilla et al., 2016). *Cand*. Methylomirabilis oxyfera, isolated from fresh water, can oxidize methane with nitrite, forming N_2 from nitrite without an N_2O intermediate (Ettwig et al., 2010). The gene in question was dubbed nitric oxide dismutase (*nod*) based on *in silico* analysis (Ettwig et al., 2012) and has been found in both transcript and gene data from the ETNP (Padilla et al., 2016). However, rates of methane oxidation in the ETNP ODZ are the slowest rates measured in the ocean ($0.034–15 \times 10^{-3}$

nmol CH_4 L^{-1} d^{-1}), suggesting that methane oxidation is not a dominant process in the ETNP (Pack et al., 2015).

Nitrite and ammonium oxidation have been shown previously in the upper ODZ in the ETNP (Peng et al., 2015; Garcia-Robledo et al., 2017). Although the ETNP ODZ contains <10 nM oxygen (Tiano et al., 2014), the low oxygen K_m for nitrite oxidation 0.5 ± 4 nM (Bristow et al., 2016), indicates nitrite oxidation is possible when oxygen concentrations below detection. Oxygen needed for nitrite oxidation may be provided by photosynthesis by *Prochlorococcus* in the upper ODZ or by mixing of waters from the oxycline above (Peters et al., 2016; Garcia-Robledo et al., 2017). The K_m for ammonia oxidizers is 333 ± 130 nM, which is significantly higher than for nitrite oxidation (Bristow et al., 2016). Correspondingly, ammonia oxidation rates drop off more rapidly than nitrite oxidation rates in the upper ODZ (Peng et al., 2015).

Here, we take a holistic approach to the low oxygen nitrogen cycle using metagenomics combined with existing chemical and rate measurements. We performed a phylogenetic analysis of key functional genes in the N cycle from assemblies of a metagenomic 10-depth profile in the water column and the >30 μm particle attached community at three key depths in the offshore ETNP ODZ. We identified previously unknown phylotypes for many key N cycling genes. We then used a read placement approach to quantify the distributions of all phylotypes for each gene in a semi-quantitative manner and found multiple differences both with respect to depth and presence in the particle attached vs. whole water community.

METHODS

Hydrographic Data

Samples were collected in April 2012 aboard the R/V Thompson TN278 using 10 L Niskin bottles on a 24 bottle CTD-rosette. A Seabird 911 Conductivity Temperature Density meter, a Seabird SBE 43 Dissolved Oxygen Sensor, a WETLabs ECO Chlorophyll Fluorometer, and a Biospherical/Licor PAR/Irradiance Sensor were attached to the rosette. Nutrient samples were filtered (GF/F glass fiber; nominal pore size 0.7 μm) before analysis. Nutrient analyses were performed by members of the University of Washington Marine Chemistry Laboratory on board the ship using a Technicon AAII system as described by the World Ocean Circulation Experiment (WOCE) Hydrographic Program protocol (Gordon et al., 1995). Ammonium was measured on board the ship using the fluorometric orthophthaldialdehyde (OPA) method due to the low detection limit (10 nM) of this method (Holmes et al., 1999). Hydrographic and nutrient data from the cruise are deposited at http://data.nodc.noaa.gov/accession/0109846. Eight day averaged satellite chlorophyll (April 7, 2012) from satellite MODIS Aqua R2014 was downloaded from http://www.science.oregonstate.edu/ocean.productivity.

N_2 Gas Concentrations

N_2 gas samples were collected from St 136 and analyzed as in Chang et al. (2010, 2012). Very briefly, duplicate gas samples were collected in evacuated 185 mL glass flasks sealed with a Louwers-Hapert valve and containing dried mercuric chloride as a preservative. To prevent air contamination when sampling, samples were transferred from the Niskin bottle to the sample flask under a local CO_2 atmosphere. Head-space gases were cryogenically processed to completely remove CO_2 and residual water vapor and run through an inline CuO furnace to remove oxygen. Then gases were measured at the Stable Isotope Lab, School of Oceanography, University of Washington on a Finnigan Delta XL isotope ratio mass spectrometer. The anoxic samples were measured against a standard containing zero oxygen. Background N_2/Ar ratios from representative water outside the ODZ were removed as in Chang et al. (2012), leaving concentrations of biologically produced N_2.

Metagenomic Data

DNA samples were obtained from station 136 (106.543° W 17.043° N; cast 136) at 10 depths including the oxycline and anoxic zones. Two liters of Niskin water were vacuum filtered onto a 0.2 μm SUPOR filter. At station BB2, a nearby station (107.148° W 16.527° N; cast 141), ∼4 L were prefiltered through >30 μm filters at 100, 120, and 150 m depths and subsequently filtered onto 0.2 μm SUPOR filters. >30 μm filters were sequenced for all three depths, but only the <30 μm filter from 120 m was sequenced. Particles >30 μm should be composed of sinking as well as large suspended particles (Clegg and Whitfield, 1990). Station 136 and BB2 were only 83 km apart and hydrographic conditions were very similar (Figures S1, S2). DNA was extracted from filters using freeze thaw followed by incubation with lysozyme and proteinase K and phenol/chloroform extraction. A Rubicon THRUPLEX kit was used for library prep using 50 ng of DNA per sample. Four libraries were sequenced on an Illumina HiSeq 2500 in rapid mode (∼25 million 150 bp paired-end reads per sample) at Michigan State. The other 10 libraries were sequenced on an Illumina HiSeq 2500 in high output mode (∼40–70 million 125 bp paired-end reads per sample) at the University of Utah (Table S1). Sequences were quality checked, trimmed, and remaining adapter sequences were removed using Trimmomatic (Bolger et al., 2014). Paired reads that overlapped were combined with Flash (Magoc and Salzberg, 2011).

Metagenomic sequences from each sample were assembled independently into larger contigs. For de novo assembly we pre-processed reads with the khmer software package (Crusoe et al., 2015), first using normalize-by-median which implements a Digital normalization algorithm (Brown et al., 2012) to reduce high coverage reads to 20x coverage, followed by filter-abund.py to trim reads of kmers with an abundance below 2, and finally we used filter-below-abund.py to trim kmers with counts above 50 (Zhang et al., 2015). We assembled the khmer processed reads with the VELVET (1.2.10) assembler (Zerbino, 2010), using a kmer size of 45. The N50, or median length, for assembled contigs ranged between 1,300 and 1,800 bp in the anoxic zone, with ∼30% of reads assembled (Table S1). The Prokka annotation pipeline (Seemann, 2014) was used for gene calling, which relies on the Prodigal algorithm for identification of coding sequence coordinates on the contigs (Hyatt et al., 2010), and preliminary functional annotation identified through similarity searching

with BLAST (Altschul et al., 1997) against UniProt (Apweiler et al., 2004) and RefSeq (Pruitt et al., 2007) databases and with HMMER v. 3.1 (Eddy, 2011) against protein domain databases Pfam (Punta et al., 2012) and TIGRFAMs (Haft et al., 2013). ETNP 2012 metagenomic reads and assembled contigs can be found at NCBI GenBank bioproject PRJNA350692.

For each gene of interest, a maximum likelihood amino acid phylogenetic tree was constructed using published full-length gene sequences as well as full or nearly full-length sequences assembled from the metagenomes themselves. Rather than rely on Prokka annotations, potential gene sequences of interest were identified from the metagenome assemblies by searching a custom blast database (Altschul et al., 1997) of all our assembled open reading frames as called by Prodigal, using representative published sequences from each section of the phylogenetic tree as query sequences. All sequences with an e-value cut-off of < -60 were included in the phylogenetic tree for further identification. Assembled genes with Ns were removed. In addition, full-length published gene sequences of closely related genes to the gene of interest were included to act as outgroups in the trees. All assembled sequences recruited from blast were combined with the previously published full-length gene sequences and aligned in amino acid space with MUSCLE v. 3.8.1551 (Edgar, 2004). Maximum likelihood phylogenetic trees were constructed with the reference sequence alignments of the genes of interest using the program RAxML v. 8.1.20 (Stamatakis, 2014). In this process, sequences with exactly identical amino acid sequences were de-duplicated (Stamatakis, 2014). The trees were constructed with a gamma model of rate heterogeneity, and appropriate amino acid substitution models were determined for each tree, and bootstrap analyses ($n = 100$) were performed.

A phylogenetic placement approach was used to characterize short metagenomic reads related to the targeted genes of interest (Berger et al., 2011) in a semi-quantitative and phylogenetically specific manner (Saunders and Rocap, 2016). For read placement, the short metagenomic reads were recruited via tblastn search of the metagenomes using an e-value cut-off of <-5 (Altschul et al., 1997). The recruited reads were trimmed to the edge of the gene of interest to remove any overhang of up or down-stream sequence, trimmed to the proper reading frame of the blast results, and converted to amino acid space. Any sequence ambiguities and stop codons were removed. Presence of sequence ambiguities and stop codons were negligible. Only sequences longer than 100 bp (33 amino acids) after quality trimming were used for placement analysis. These amino acid translated reads were aligned to the reference sequences in amino acid space using PaPaRa: Parsimony-based Phylogeny-Aware Read Alignment program v. 2.5 (Berger and Stamatakis, 2011). Following the PaPaRa alignment, paired end reads were combined into one sequence in the same alignment using a python script and placed as one read on the tree using the EPA: Evolutionary Placement Algorithm portion of RAxML (Stamatakis, 2014). Reads that placed with outgroups on each phylogenetic tree were not counted toward that gene's total. This was particularly important for closely related genes such as nitrate reductase *narG*, with the *nxrA* outgroup and for nitrite oxidoreductase *nxrB* with the *narH* outgroup.

To take into account differences in sequencing effort between samples and recruitment capacity among the different genes, read placement length normalized occurrence of the target genes were normalized to the length normalized abundance of the universal single copy core gene RNA polymerase (*rpoB;* Figure S3).

$$\%\ prokaryotic\ community = \frac{\frac{Gene\ A\ reads}{Length\ A}}{\frac{rpoB\ reads}{Length\ rpoB}}.$$

Since RNA polymerase is, to our knowledge, present as a single copy gene in all bacterial and archaeal genomes, normalization of a target gene occurrence to RNA polymerase abundance indicates what percentage of the prokaryotic community contains the gene of interest. This normalization allows our placements to be quantitative in a relative manner. However, this analysis does not take into account that the density of prokaryotic cells may, and undoubtedly do, change with depth.

Several papers have been published from this cruise, and we compare our metagenomic data to previous rate, qPCR, and lipid data. For clarification, these data and their references are listed in Table S2.

Previously Published Transcripts

In order to assess whether phylotypes observed in our metagenomes were expressed in the environment, we applied the same read placement approach to previously published metatranscriptomic data. Transcripts from 2 size fractions (0.2–1.6 and 1.6–30 μm) at five low oxygen depths from a coastal station in the ETNP in 2013 (18° 54.0′ N, 104° 54.0′ W) (Ganesh et al., 2015) were placed on our phylogenetic trees using methods as above. Transcript libraries had ∼120,000 reads per sample.

RESULTS AND DISCUSSION

We occupied 2 stations in the offshore Eastern Tropical North Pacific in April 2012. At these stations, the depth of anoxia was ∼105 m according to oxygen measurements with a STOX sensor (detection limit 2 ± 5 nM O_2; Tiano et al., 2014). The oxycline, where oxygen concentrations decrease rapidly, extended from 60 to 100 m (**Figure 1A**). STOX oxygen concentrations were 4.7 and 0.8 μM at 90 and 100 m in the lower oxycline (Tiano et al., 2014). Ammonium concentrations, as determined by the extremely sensitive OPA method, were undetectable in the anoxic zone (**Figure 1C**), but the nitrite maximum reached nearly 5 μM at 150 m (**Figure 1C**). N_2O concentrations had the usual large maximum in the oxycline, but also had a second smaller maximum at 140–150 m in the anoxic zone (**Figure 1F**; Peng et al., 2015). Biological N_2 gas increased in the anoxic zone and was between 10 and 11 μM (**Figure 1H**). To characterize the functional and taxonomic diversity of this oligotrophic ODZ community, we constructed and sequenced metagenomes from whole water from 10 depths (60–300 m) at station 136 and from >30 μm particles collected from Niskin bottles at 3 depths at station BB2, which was close to and very similar physiochemically to station 136 (Figure S2). We classified short read sequences by both function

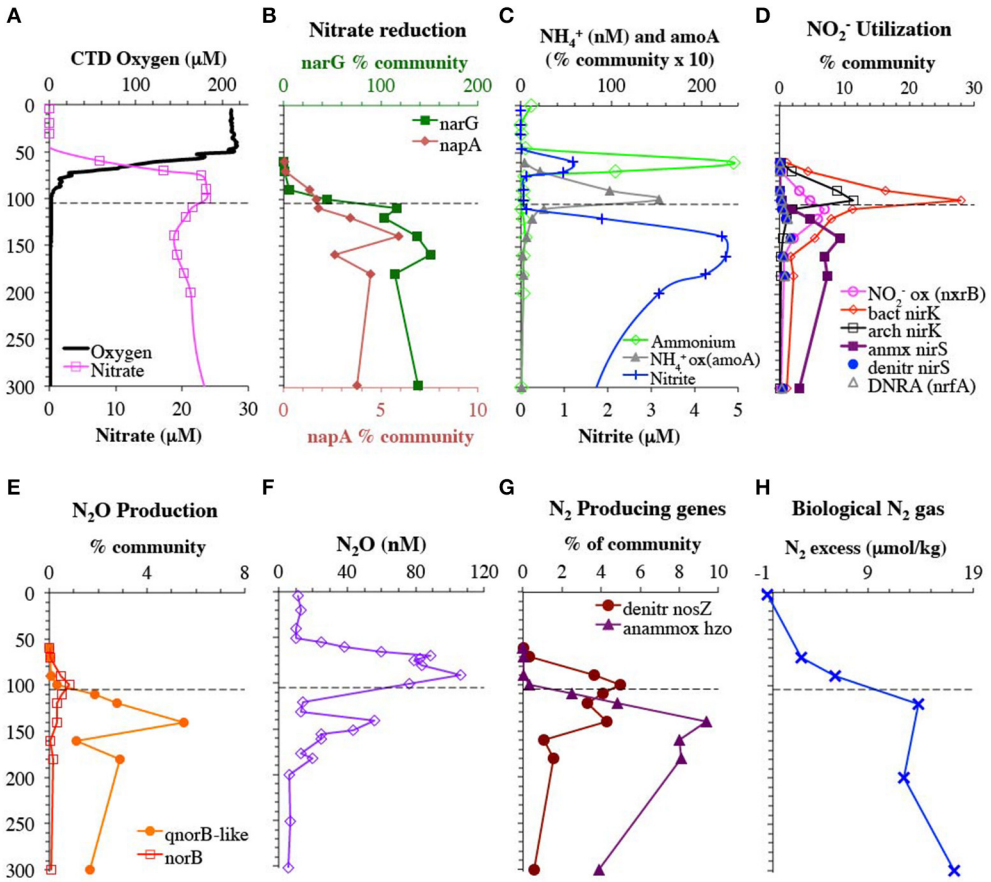


FIGURE 1 | Depth profiles of metagenomic reads encoding for enzymes used in the low oxygen N cycle and relevant chemical species. **(A)** CTD oxygen and nitrate concentrations from St 136. **(B)** Genes used in nitrate reduction. Genes for nitrate reductases *narG* and *napA* are shown on different axes for clarity. **(C)** Nitrite and OPA ammonium concentrations from St 136 as well as the gene for ammonia monooxygenase (*amoA*) used in ammonia oxidation. **(D)** Genes used for nitrite utilization: nitrite reduction to NO (bacterial and archaea *nirK*, anammox *nirS*, and bacterial *nirS*), nitrite reduction to ammonium (DNRA: *nrfA*) and nitrite oxidation (*nxrB*). **(E)** N_2O producing genes: NO reductases *norB* and *qnorB*-like. **(F)** N_2O concentration profile from St BB2 (Peng et al., 2015). **(G)** N_2 producing genes: N_2O reductase (*nosZ*) represents denitrification and hydrazine oxidoreductase (*hzo*) represents anammox. **(H)** Biological N_2 gas from St 136. Dashed line represents the top of the ODZ, which is at 105 m (Tiano et al., 2014). % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*).

and phylotype by placing reads on reference trees constructed from full length genes, including those assembled from our metagenomes.

Community Structure of Free Living and Particle Communities

We examined community structure using the RNA polymerase (*rpoB*) gene, found in all archaeal and bacterial genomes (**Figure 2**, Figure S3). SAR11 was the most abundant clade overall, making up 60% of the community at 300 m. SAR11 has previously been found to be 10–40% of the community in ODZs (Tsementzi et al., 2016). Other notable heterotrophic clades present included SAR406, SAR116, Marine Group II euryarchaeota, and Flavobacteria (**Figure 2**). Autotrophic microbes present included Cyanobacteria (photosynthesis), Marine Group I Thaumarchaeota (ammonia oxidation), *Nitrospina* (nitrite oxidation), and *Cand.* Scalindua (anammox), and together they made up <20% of the community in the oxycline and anoxic water column (**Figure 2**). Autotrophic S oxidizers are also known to be active in ODZs (Stewart et al., 2012), but known S oxidizers, including SUP05, were not identified here. However, it should be noted that a large number of the *rpoB* sequences were novel phylotypes including novel clades of Actinobacteria, Chloroflexi, Acidobacteria, and gammaproteobacteria whose metabolic lifestyles are unknown (Figure S3).

Particles may be hotspots of heterotrophic activity in the ocean and harbor a microbial community distinct from those free-living in the water column (Delong et al., 1993; Ploug et al., 1999). To further examine the microbial community on particles that make up the sinking organic matter in the ODZ, we sequenced metagenomes from >30 µm particles collected from Niskin bottles at 3 depths at the offshore station BB2. Particles >30 µm should be composed of sinking as well as large suspended particles (Clegg and Whitfield, 1990). Estimates of 16S rRNA abundance in the ETNP using qPCR indicated that 8–15

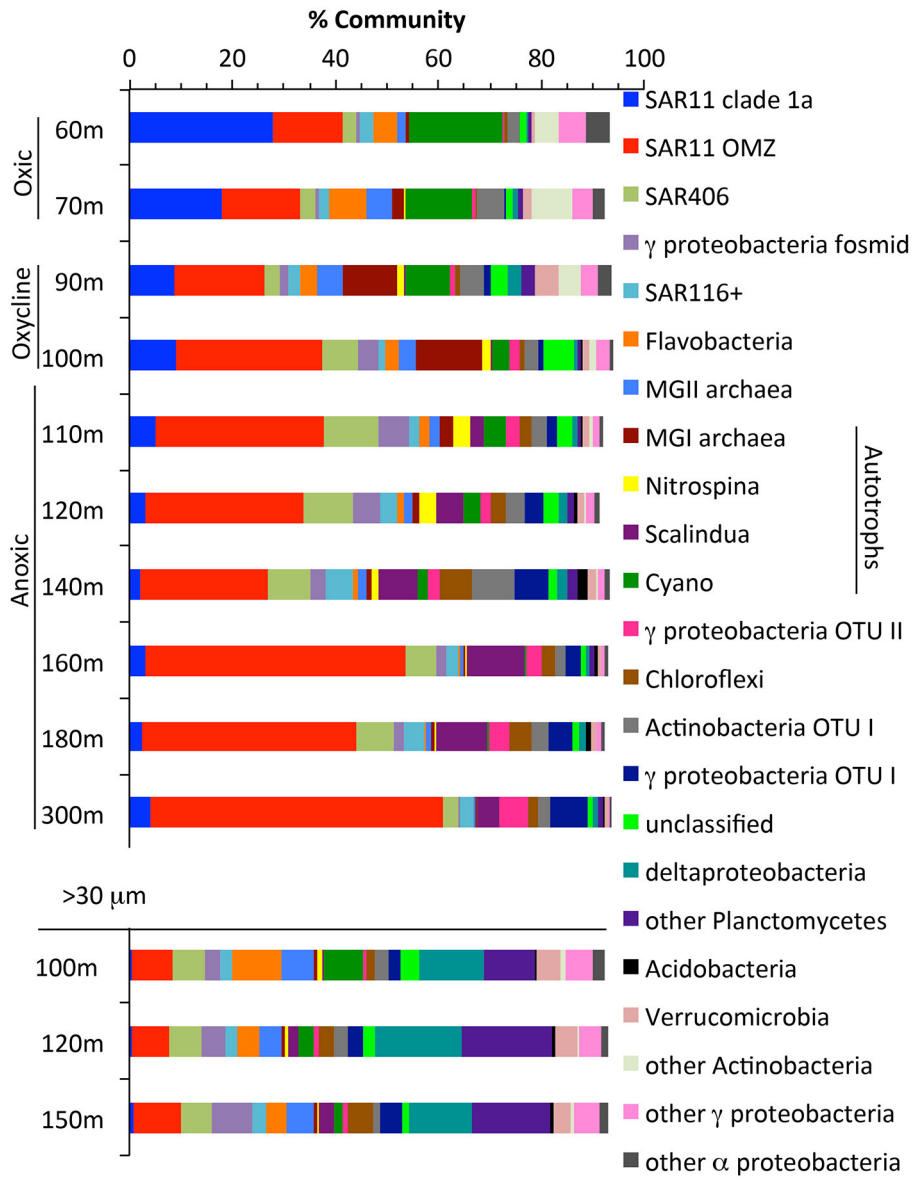


FIGURE 2 | The community in the water column and in >30 micron particles as determined by RNA polymerase gene *rpoB*. Groups found on the *rpoB* phylogenetic tree (Figure S2). Known autotrophs are labeled in the legend. Gamma fosmid represents relatives of HOT fosmid GU567967, a heterotrophic Salinisphaeraceae.

times more 16S rRNA was found per mL in free-living (<1.6 μm) communities compared to particle (>1.6 μm) fractions (Ganesh et al., 2015). Thus, the free-living community likely dominates bulk water samples. In our dataset, the microbial community on particles (>30 μm) differed from the community in the bulk water (**Figure 2**). Among others, SAR11, Marine Group I archaea (ammonia oxidizing) and *Cand.* Scalindua (anammox), all known to be free-living (Fuchsman et al., 2011, 2012b, Ganesh et al., 2014, 2015), were a much smaller proportion of the particle community. Delta proteobacteria, Planctomycetes, Flavobacteria, Verrucomicrobia, and MGII euryarchaeota were enriched in particles, consistent with observations in other environments (Delong et al., 1993; Fuchsman et al., 2011, 2012b; Ganesh et al., 2014; Glass et al., 2015; Orsi et al., 2015). In contrast to previous observations in the ETNP (Ganesh et al., 2015), the nitrite-oxidizer *Nitrospina* was not enriched in particles. This could be due to the different size filter used to define "particles" (1.6 vs. 30 μm used here) because individual *Nitrospina* cells can be 6 μm long (Spieck and Bock, 2015).

Genetic Capacity for Denitrification Enhanced on Particles

To examine the role of sinking and large suspended particles in production of N_2 by the ODZ microbial community, we mapped metagenomic reads onto reference phylogenetic trees constructed for key genes in the anoxic N cycle (**Figure 3**).

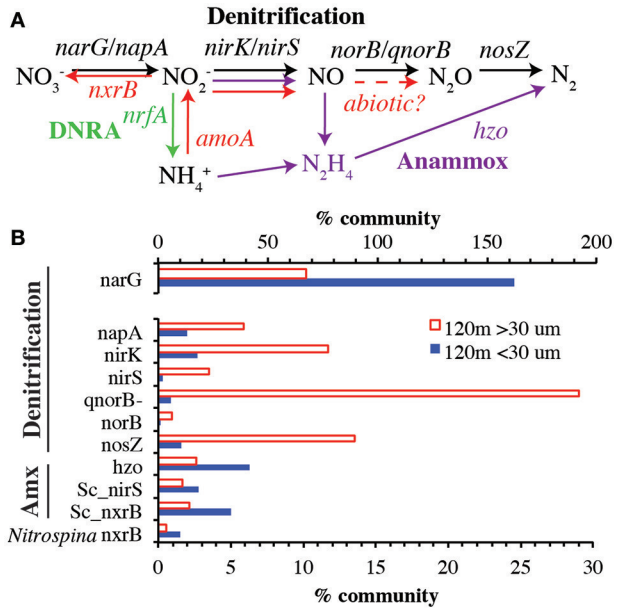


FIGURE 3 | (A) The low oxygen N cycle with genes examined in this paper indicated. Black arrows indicate denitrification. Purple arrows indicate anammox. Red arrows indicate processes mediated by ammonia and nitrite oxidizing organisms. **(B)** Comparison of genes between >30 μm and <30 μm fractions at 120 m. Note that nitrate reduction gene *narG* is on its own axis. *amoA* and *nrfA* were <1% of community at 120 m, so are not shown. Sc_*nirS* and Sc_*nxrB* indicate genes belonging to anammox bacteria *Cand*. Scalindua. *Nitrospina nxrB* represents nitrite oxidizing bacteria. % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*).

A comparison of >30 μm and <30 μm communities at 120 m revealed that most of the genes encoding the steps of denitrification were enriched in the particulate fraction while anammox genes and the nitrate reductase *narG* represented a greater fraction of the community in the free-living fraction (**Figure 3**). The particle associated denitrification genes were not present in equal amounts, ranging from more than 25% of the community possessing the NO reduction enzyme (*qnorB-like*) to <5% with the nitrite reductase *nirS*. Overall, nitrate reductase *narG* was the most abundant gene examined, present in the equivalent of >150% of the free-living community and >50% of the particle community. This apparent overestimate is due to the presence of multiple copies of *narG* in many microbial genomes, including the abundant ODZ SAR11 phylotypes (Tsementzi et al., 2016), but nevertheless underscores the important role of nitrate reduction in this environment, both on particles and in the water column. The habitat partitioning observed here, with anammox and *narG* primarily in the free living fraction and other denitrification genes enriched in sinking particles is consistent with data from >1.6 μm suspended particles (Ganesh et al., 2014, 2015). In the coastal ETNP, removal of >1.6 μm particles decimated nitrate reduction rates (Ganesh et al., 2015), reinforcing the idea that even the free-living bacteria in the ODZ are dependent on organic matter fluxes. Consistent with the *rpoB* data, nitrite oxidoreductase (*nxrB*) for nitrite-oxidizer *Nitrospina* was more abundant in the <30 μm fraction.

Gene Depth Profiles

We next examined the depth distribution and taxonomic diversity of these same key genes (**Figure 3A**) in the oxycline and upper 200 m of the ODZ water column. We compare these depth profiles to measurements of chemical species that serve as reactants and products for the reactions catalyzed by the enzymes these genes encode (**Figure 1**).

Nitrification Genes

The ammonia monooxygenase (*amoA*) gene for ammonia-oxidizing archaea had a maximum of 15% of community at 100 m at the bottom of the oxycline (**Figure 1C**). This maximum in *amoA* was below the primary maxima in ammonium and nitrite concentrations found at 60 m. This depth profile combined with ammonia oxidation rates from Peng et al. (2015) indicate that nitrite is being produced in the 80–100 m region, despite the lack of measurable nitrite in this region. The gene encoding the nitrite oxidizing enzyme nitrite oxidoreductase (*nxrB*) is unmeasurable at 70 m, but 3% of the community at 90 m, and had a maximum of 7% of the community in the ODZ at 110 m that can be solely attributed to *Nitrospina* (**Figure 1D**, Figure S6). The presence of *Nitrospina nxrB* in the upper ODZ is consistent with *nxrB* transcript data from the ETNP (Garcia-Robledo et al., 2017). The presence of *Nitrospina* may help explain the lack of measurable nitrite in the 80–100 m region.

Nitrate Reduction

The first step in denitrification, nitrate reduction to nitrite, can be carried out by nitrate reductases encoded by either *narG* or *napA*, both of which were detected here. However, *narG* was an order of magnitude more abundant than all the other denitrification genes (**Figure 1B**). The *narG* maximum at 160 m corresponded with a reduction in nitrate concentrations and the secondary nitrite maximum (**Figures 1A–C**), implying nitrate reduction activity. Again, after normalizing with the single copy core gene, more than 100% of the community contained the *narG*, implying multiple copies per genome in some bacteria. SAR11, which is very abundant in our *rpoB* data (**Figure 2**), is known to have two distinct *narG* in the same cell (Tsementzi et al., 2016) and we found both SAR11 *narG* types here along with six other phylotypes (**Figure 4**). If we assume that all SAR11 genomes have 2 types of *narG*, we calculate that 75–105% of the total microbial community has *narG*. This number may suggest that other groups in addition to SAR11 also possess duplicate copies of *narG*. In contrast, the two phylotypes of *napA* totaled only 5% of the community in the ODZ (**Figure 1B**, Figure S4). The capacity for nitrate reduction is clearly prevalent in the community. Examination of selected long contigs containing *narG* indicated multiple nitrate reductase subunits and at least one nitrate transporter (*narU*, *narT*, or *narK*) in all cases (Figure S5). Contigs associated with OTU I contained a transposase, indicating the potential for horizontal gene transfer (Figure S5).

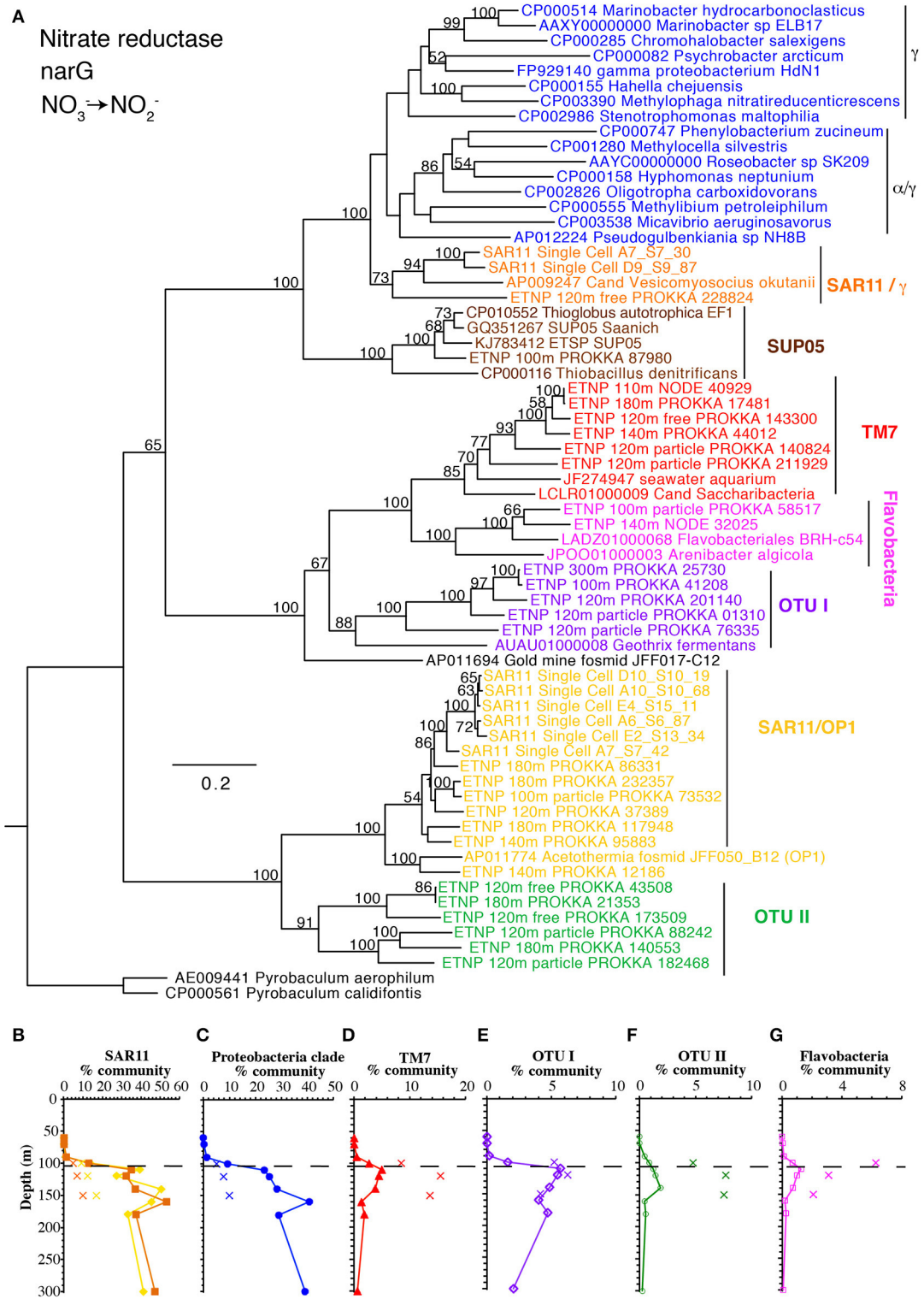


FIGURE 4 | (A) Phylogenetic tree of membrane bound nitrate reductase gene *narG.* Names with ETNP indicate sequences assembled from our metagenomic samples. Labels and colors on the tree match phylotype depth profiles **(B–G)**. **(B)** Depth profile of both forms of *narG* belonging to SAR11. **(C)** Combined profile of α and γ proteobacteria, **(D)** uncultured group TM7, **(E)** unknown OTU I, **(F)** unknown OTU II, and **(G)** Flavobacteria. In all depth profiles, Xs indicate particulate (>30 μm) samples. Dashed line indicated the top of the ODZ. % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*).

Nitrite Reduction

The nitrite produced by nitrate reduction can have many fates, including further reduction to NO, oxidation back to nitrate, and reduction to ammonia. We found evidence for existence of all of these pathways, in differing amounts. As mentioned previously, the gene encoding the nitrite oxidizing enzyme nitrite oxidoreductase (*nxrB*) had a maximum in the upper ODZ (**Figure 1D**). In contrast *nrfA,* encoding the DNRA nitrite reductase to ammonia, was present but always <2% of the community (**Figure 1D**). Only one phylotype of *nrfA* was present (Figure S7).

The second step in denitrification, nitrite reduction to NO, was dominated by *nirK* (copper containing nitrite reductase) although *nirS* (iron containing nitrite reductase) was also detected in much lower abundances (**Figure 1D**). *nirK* had a maximum at the top of the ODZ at 100 m, where it was possessed by ∼30% of the community (**Figure 1D**). At least 10 phylotypes were detected and five were present in >1% of the community (**Figure 5**). Three of these phylotypes, (OTU I, Chloroflexi, and MGI thaumarchaeota) had a maximum in the lower oxycline and were undetectable below 140 m (**Figures 5B,D,E**), while the *nirK* from nitrite oxidizer *Nitrospina* had maxima within the ODZ at 110 m **(Figure 5F)**. Only one *nirK* phylotype (OTU II) was clearly particle attached and this phylotype also had a maximum in the ODZ (**Figure 5C**). OTU II *nirK* was found on an assembled contig with a *nosZ* gene (**Figure 6**), which is discussed with that *nosZ* phylotype below. We note that unlike previous metagenomic examination of *nirK* (Glass et al., 2015; Lüke et al., 2016), MGI Thaumarchaeota were not the dominant *nirK* containing organism. Instead bacterial OTU I was 21% of the community at 100 m (**Figure 5**). Since transcripts from the 2013 ETNP cruise examined in Glass et al. (2015) place on our OTU I *nirK*, it seems this difference between reports may be methodological rather than due to interannual variability. Unfortunately, assemblies of OTU I *nirK* from our study were short contigs. However, a published SAR11 metagenomic contig from the coastal ETNP contained a partial *nirK* (scaffold 00818) (Tsementzi et al., 2016). This partial SAR11 *nirK* sequence was too short (242 bp) to be a branch on our phylogenetic tree, but aligned with 98.6% identity to the OTU I cluster here using MUSCLE (Edgar, 2004). The published representative of cluster OTU I on our phylogenic tree is from a fosmid obtained Station ALOHA (HF0770-09N23) which also appears to be a SAR11 relative (best BLAST hit for the fosmid using the nt database: SAR11 relative CP003809 HIMB5, E-value: 0.0). Presence in some SAR11 bacteria would explain how OTU I of *nirK* could be so abundant (**Figure 5B**). While *nirK* was more abundant on particles than in the water column at 120 m (**Figure 3**), at 100 m, *nirK* was more abundant in bulk water (26%) at station 136 than on particles (11%) at BB2, (Figure S8). This difference is due to the dominance of SAR11 *nirK* OTU I in the water column at the top of the ODZ and is consistent with the free-living nature of SAR11 determined from *rpoB* (**Figure 2**).

Although *nirS* was present in a much lower percentage of the community than *nirK*, at least five distinct phylotypes were still present (Figure S9). With the exception of anammox *nirS* (**Figure 1D**), the other 4 phylotypes were present in <2% of the community, and all had maxima within the ODZ at 140 m. Two of these phylotypes (OTU II and III) were particle-attached (Figure S9). The prevalence of the *nirK* copper nitrite reductase over *nirS* is consistent with other untargeted approaches that have also detected *nirK* (Ganesh et al., 2014, 2015; Glass et al., 2015; Lüke et al., 2016) in ODZs but it is at odds with prior primer-based approaches that indicated *nirK* was not environmentally relevant in ODZs (Jayakumar et al., 2013).

Nitric Oxide Reduction

The third step in denitrification, NO reduction to N_2O, is mediated by nitric oxide reductase, encoded by genes *norB* or *qnorB* (Lam et al., 2011). However, the canonical forms of these genes were present in very low abundance in this water column (<0.9% *norB*; **Figure 1**). Instead most of our assembled sequences and short read sequences cluster (**Figure 7**) with a form of the *qnorB* suggested by *in silico* analysis to be *nod* (nitric oxide dismutase) in the methane-utilizing denitrifier *Cand*. Methylomirabilis from the NC10 phylum, which was theorized to reduce NO straight to N_2 without a N_2O intermediate (Ettwig et al., 2012). However, our data is not consistent with all of the genes detected here encoding a nitric oxide dismutase. Many bacteria that have *qnorB* genes in this cluster are not known to dismutate nitric oxide and some also contain nitrous oxide reductase in their genomes (**Figure 7**). The *qnorB-like* gene has a maximum at 140 m at the same depth as the second N_2O peak (**Figures 1E,F**) where N_2O production rates are modeled to have a maximum (Babbin et al., 2015). However, neither *norB* nor *qnorB* was present at these depths (**Figure 1**). Theoretically, NO released from a cell could abiotically or non-enzymatically produce N_2O with iron or thiols under anoxic conditions (Hughes, 2008; Kampschreur et al., 2011; Kozlowski et al., 2016), but this has not been shown in the environment. Thus, if the *qnorB*-like gene is not involved in N_2O production, there are no known genes present to produce N_2O in the ODZ. Though we can't rule out the presence of an unknown novel gene for N_2O production, it would have to be in a completely different gene family from *norB*/*qnorB*/*qnorB*-like to be missed by our methods. When the quinol binding site and active site of *qnor*, *qnor*-like, putative *nod*, and ETNP assembled contigs are compared, the quinol binding site appears to have more variability between gene types than the active site (Figure S10). While our assembled ETNP sequences do share the differences in the active site seen in the putative nod enzyme, these changes are also seen in the other *qnorB*-like genes (Figure S10). We suggest that at least some, potentially all, of the *qnorB*-like phylotypes detected here retain their function as nitric oxide reductases like their homologs *norB* and *qnorB*.

Sequences for *nod*-like/*qnorB*-like genes from metatranscript assemblies in the coastal ETNP were combined with NC10 16S rRNA sequences related to *Cand*. Methylomirabilis as evidence to suggest a role for methane oxidation in N_2 production in ODZs (Padilla et al., 2016). The transcripts from the 2013 cruise place on our tree at OTUs II, III, and IV (**Figure 7**) and assembled sequences from Padilla et al. (2016) belong to

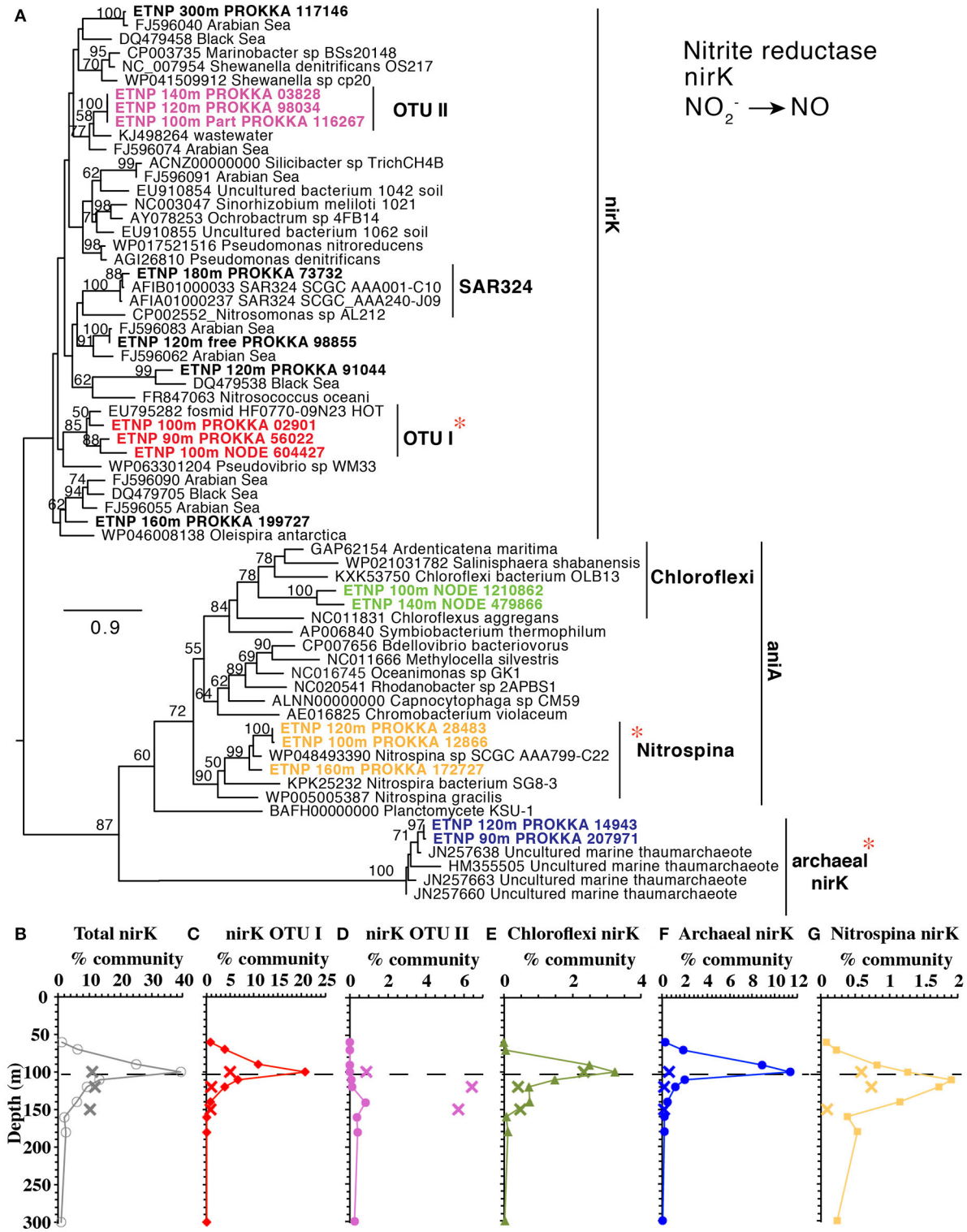


FIGURE 5 | (A) Phylogenetic tree of the *nirK/aniA* gene encoding a NO producing nitrite reductase containing copper. Names with ETNP indicate sequences assembled from our metagenomes. Labels colors on the tree match individual OTU depth profiles below **(C–G)**. Depth profiles of **(B)** total *nirK* and *nirK* phylotypes: **(C)** OTU I, **(D)** OTU II, **(E)** Chloroflexi **(F)** ammonia-oxidizing archaea, **(G)** nitrite oxidizer *Nitrospina*. In all depth profiles, Xs indicate particulate (>30 μm) samples. Dashed line represents the top of the ODZ. % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*). Red asterisks indicate placement of ETNP transcripts from Ganesh et al. (2015).

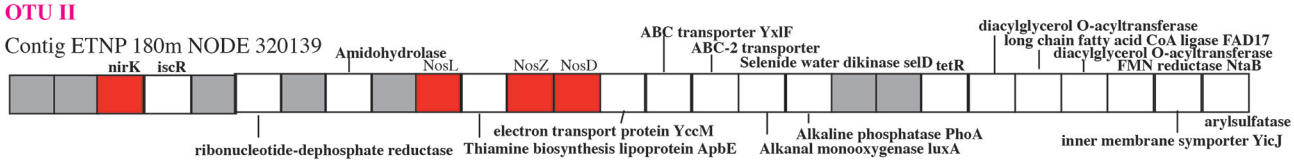


FIGURE 6 | Schematic of a contig containing a *nirK* gene and a *nosZ* gene. Hypothetical proteins are shown in gray and genes of interest are in red. NosL is a copper chaperone and NosD is part of a copper transporter.

our OTU III (Figure S10). In our dataset, however, there is no correlation between any *qnorB*-like gene OTUs and subunits of methane mono-oxygenase, *pmoA,* and *pmoB*. NC10 *pmoA* was possessed by at most of 0.1% of the community, and *pmoB* was barely above detection and no NC10 *pmoB* was detectable in our metagenomes. In contrast, the *qnorB*-like gene was present in >25% of the community in particles (**Figure 3**). Thus, the large difference in % community between gene in these two pathways makes it seems unlikely that the abundant organisms containing the *qnorB*-like gene are involved in methane oxidation. Additionally, when our long contigs containing *qnorB-like* gene are examined, none of the contigs had BLAST hits in National Center for Biotechnology Information (NCBI) nucleotide collection database (nt) (NCBI Resource Coordinators, 2017) with E values < −20. Contig 120 m particle NODE 148559 has competence genes right next to the *qnorB-like* gene, implying that *qnorB-like* gene could have been transferred (Figure S11), and in general, the contigs appear to be dissimilar to each other, even between multiple contigs in the same *qnorB*-like phylotype (OTU II; Figure S11). Thus, the *qnorB*-like gene appears to be in multiple organisms in the ETNP, which are not related to *Cand.* Methylomirabilis or other bacteria in the NCBI nucleotide database.

N_2 Gas Production

The final step in denitrification, N_2 production from N_2O, is mediated by *nosZ*, which was elevated from 90 to 140 m (**Figure 1G**). The *nosZ* gene depth profile had two maxima at 100 and 140 m (**Figure 1**), which corresponded to the two N_2O concentration maxima in the lower oxycline and at 140 m. Potential N_2O reduction rates also had a maximum at 140 m (2.7 nM/d, Babbin et al., 2015). The N_2O reductase gene *nosZ* had 5 phylotypes in the ETNP. Notably these ETNP *nosZ* phylotypes were completely different from sequences from the coastal ETSP (Castro-González et al., 2015) (**Figure 8A**). ETNP N_2O reductase (*nosZ*) phylotypes could be separated into two groups based on their depth distribution with maxima corresponding to the two N_2O concentration maxima (**Figure 8**). It is possible that these different depth profiles represent different tolerances of the organisms to oxygen. The Flavobacterial *nosZ* was abundant at the top of the ODZ with a maximum at 100 m, corresponding to the upper *nosZ* gene maxima (**Figure 8B**). This group is represented by one extensive 11 gene contig (ETNP 120 m NODE 73975) containing *nosD*, *nosZ*, and cytochrome c (Figure S12). The best BLAST hit for this contig was the Flavobacteria *Lutibacter* sp. LPI (#CP013355), confirming the Flavobacteria affinity of this *nosZ*. The other 4 phylotypes represent novel clades. One clade of *nosZ* sequences clustered with *Chloroflexi*, but this clade was not represented by multi-gene contigs. The Flavobacteria and potential *Chloroflexi* phylotypes both had their maxima at the upper N_2O peak and made up roughly similar % community in the water and on particles (**Figures 8B,C**, Figure S13). The other three unidentified phylotypes each had maxima at the second N_2O maximum at 140 m (**Figures 8D–F**). One of the nosZ phylotypes had the C terminal extension typical of S oxidizing bacteria, suggesting it may belong to an autotrophic denitrifier, and this phylotype was enriched in particles (**Figures 8A,F**, Figure S13). This clade was represented by a contig containing 23 gene sequences (ETNP 120 m NODE 137405) including a photosystem I gene *psaC* along with *nosL*, *nosD*, and *nosZ* (Figure S12). The fourth *nosZ* phylotype, OTU I, was predominately free-living (**Figure 8D**, Figure S13). This phylotype was represented by three extensive contigs (Figure S12). Multiple genes on the contigs identified this organism as a heterotroph, including beta-galactosidase. Particles were dominated by the fifth *nosZ* phylotype, OTU II (**Figure 8E,** Figure S13). Two extensive contigs represent the upper branch of OTU II (Figure S12). Along with genes for *nosL*, *nosD*, and *nosZ*, contig ETNP 180 m NODE 320139 also contained a *nirK* gene belonging to *nirK* OTU II, a phylotype that was also particle attached. Thus, *nosZ* phylotype OTU II and *nirK* OTU II are both the same unidentified organism (**Figure 6**).

Genes representing the second N_2 producing pathway, anammox, had a maximum from 120 to 180 m (**Figures 2D,G**, **9A**). Notably the depth profile of the gene for hydrazine oxidoreductase (*hzo*), the final step in the anammox pathway, was consistent with both the *nirS* and *rpoB* reads that branched with *Cand.* Scalindua (**Figure 9A**). As expected, only one phylotype of anammox bacteria was present on both *hzo* and *nirS* phylogenetic trees (Figures S9, S14). The combined maxima in anammox genes and *nosZ* corresponded to the upper N_2 gas maximum (**Figure 1H**). This depth profile for anammox genes is consistent with intact ladderane lipid data from the same station (Sollai et al., 2015) and with calculated anammox rates (**Figure 9A**). Anammox rates were estimated by subtracting N_2O reduction rates (Babbin et al., 2015) from total N_2 production rates from BB2 (Babbin et al., 2014) and these differential rates showed some variability, but still had a maximum at the same depth as the metagenomic reads (**Figure 9A**).

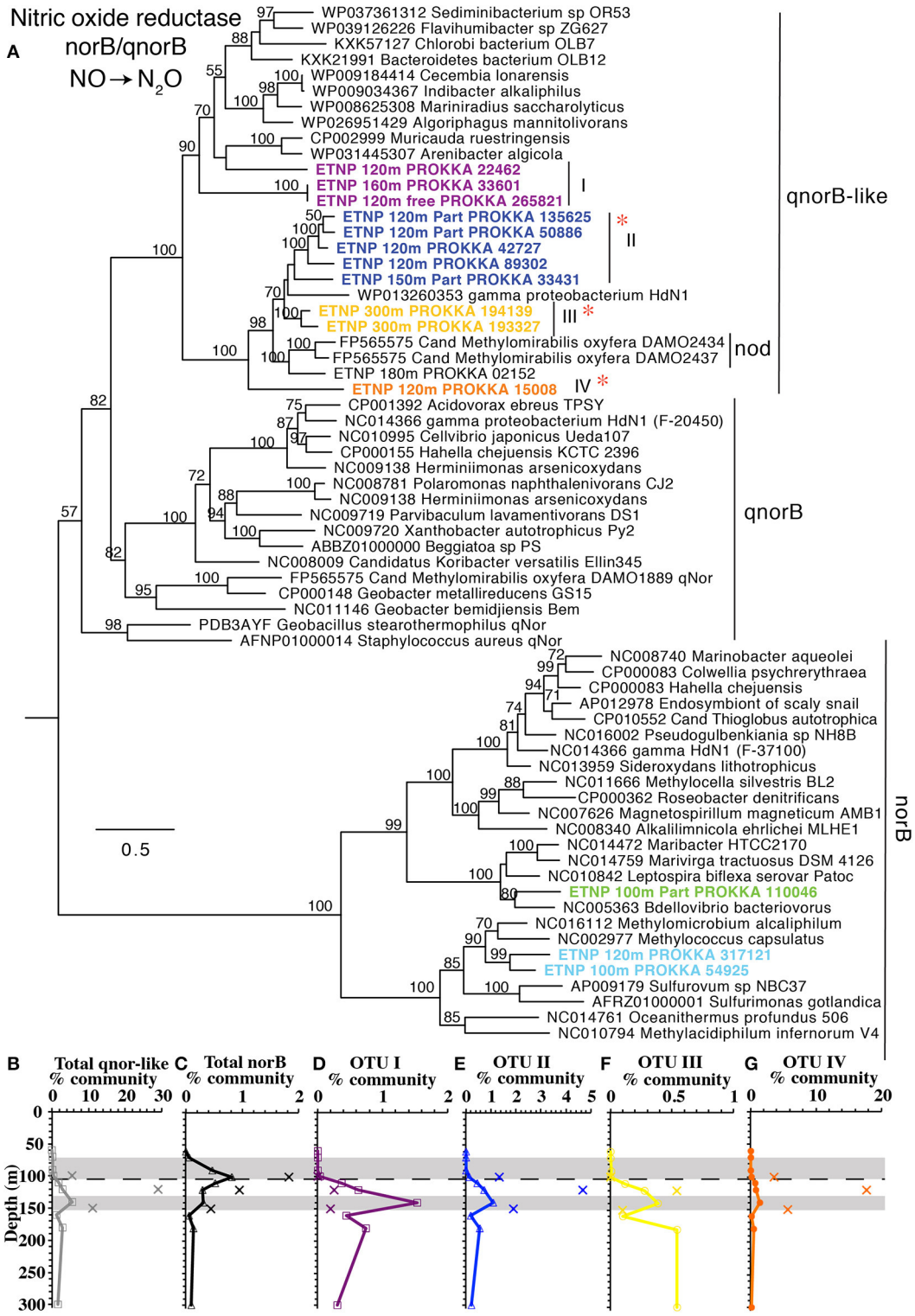


FIGURE 7 | (A) Phylogenetic tree of nitric oxide reducing enzymes *norB, qnorB,* and *q-norB-like*. Putative *nod* genes are indicated. Names with ETNP indicate sequences assembled from our metagenomes. Labels and colors on the tree match individual depth profiles **(D–G)**. Depth profiles of **(B)** total *qnorB*-like, and **(C)** total *norB* sequences as well as *q-norB-like* phylotypes: **(D)** OTU I, **(E)** OTU II, **(F)** OTU III, and **(G)** OTU IV. In all depth profiles, Xs indicate particulate (>30 μm) samples. Shaded area indicates the depth of the N_2O maxima (**Figure 1**). Dashed line represents the top of the ODZ. % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*). Red asterisks indicate placement of ETNP transcripts from Ganesh et al. (2015).

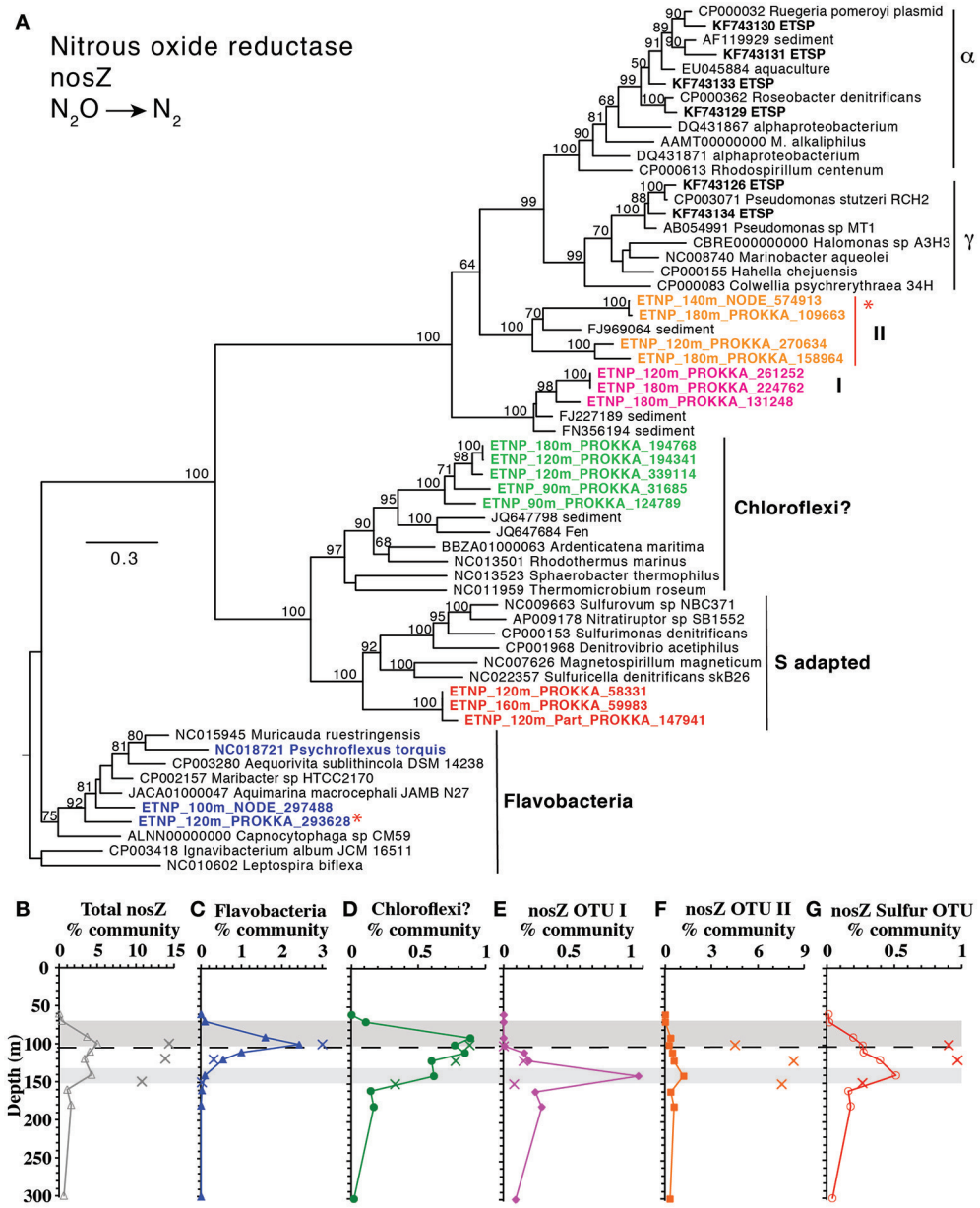


FIGURE 8 | (A) Phylogenetic tree of N_2O reductase enzyme *nosZ*. Names with ETNP indicate full-length sequences assembled from our metagenomic samples. Colors on the tree match individual OTU depth profiles **(C–G)**. Depth profiles of **(B)** total *nosZ* and of *nosZ* phylotypes: **(C)** Flavobacteria, **(D)** Chloroflexi, **(E)** OTU I, **(F)** OTU II, and **(G)** Sulfur adapted. In all depth profiles, Xs indicate particulate (>30 μm) samples. Shading on depth profiles indicates depth of the two N_2O maxima (**Figure 1**). Dashed line represents the top of the ODZ. % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*). Red asterisks indicate placement of ETNP transcripts from Ganesh et al. (2015).

Correlation of Functional Gene Abundance with Activity

Although gene presence assessed by metagenomics only confirms the potential for a given function, the anammox data suggest a close correspondence in space between the presence of an organism and its activity in this environment. We further assessed this relationship in two other well-characterized microbial groups, the ammonia-oxidizing MGI Thaumarchaeota, and the nitrite oxidizer *Nitrospina*. The depth profile of the ammonia monooxygenase gene *amoA* was tightly correlated with MGI Thaumarchaeota specific *nirK* and *rpoB* reads, and all three genes had a maximum at 100 m (**Figure 9B**). Previously determined qPCR measurements of *amoA* at BB2 also had a maximum at 100 m (Peng et al., 2015), and the overall depth profiles were consistent with a slope of 1106.2 cells per mL/% community with *amoA* ($R^2 = 0.7$). Rates of ammonia oxidation, which were calculated including ammonia oxidized to both nitrite and nitrate (Peng et al., 2015) had a maximum at

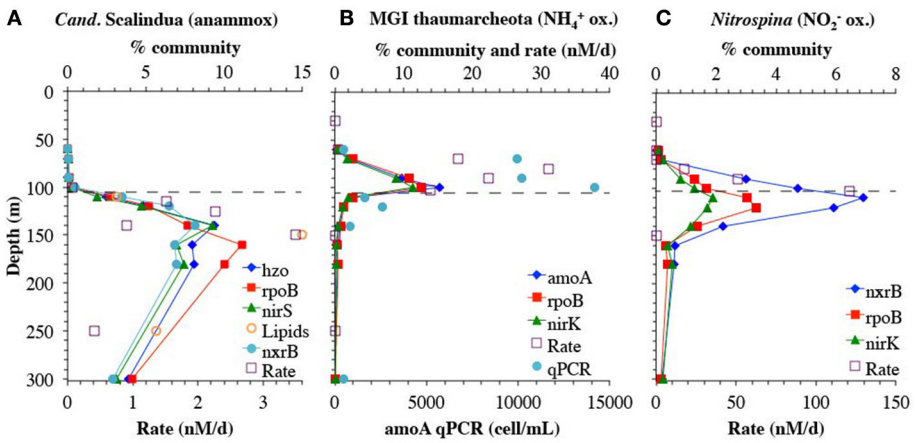


FIGURE 9 | A comparison between genes and rates. **(A)** Genes for anammox bacteria *Cand*. Scalindua for nitrite reductase (*nirS*), hydrazine oxidoreductase (*hzo*) and single copy core gene *rpoB*. Anammox rates are calculated by subtracting N_2O reduction rates (Babbin et al., 2015) from total N_2 production rates from BB2 (Babbin et al., 2014). PC-monoether ladderane lipid values are from St 136 in Sollai et al. (2015), but are normalized to fit on the % community axis; values are 15 pg/L at 110 m, 72 pg/L at 150 m, 27 pg/L at 250 m, and 25 pg/L at 350 m. **(B)** genes for ammonia oxidizing MGI Thaumarcheota for ammonia monooxygenase (*amoA*), nitrite reductase (*nirK*), and single copy core gene *rpoB*. Ammonia oxidizing rates and *amoA* qPCR are from BB2 in Peng et al. (2015). **(C)** genes from nitrite oxidizing bacterium *Nitrospina* for nitrite reductase (*nirK*), nitrite oxidoreductase (*nxrB*), and single copy core gene *rpoB*. Nitrite oxidizing rates are from BB2 in Peng et al. (2015). Dashed line represents the top of the ODZ. % Community is calculated in comparison to the single copy core gene RNA polymerase (*rpoB*).

80 m, above the maxima in both qPCR and metagenomic reads (**Figure 9B**).

The gene encoding the nitrite oxidizing enzyme nitrite oxidoreductase (*nxrB*) had a maximum in the ODZ at 110 m, again consistent with *Nitrospina* specific *nirK* and *rpoB* reads. Unlike in the Arabian Sea (Lüke et al., 2016), here nitrite oxidoreductase can be solely attributed to *Nitrospina* (**Figure 9C**, Figure S6). Like isolate *Nitrospina gracilis* (Lücker et al., 2013), it appears that the ODZ *Nitrospina* has two copies of the *nxrB* gene per genome, as determined by comparison with the single copy core gene *rpoB* (**Figure 9C**). This genetic potential for nitrite oxidation within the ODZ is consistent with measured nitrite oxidation rates of >100 nM/d at BB2 (**Figure 9C**) (Peng et al., 2015) and nitrite oxidation rates had a similar depth profile as the metagenomic read depth profile for *Nitrospina*. Nitrite oxidizers in the ETSP had a high affinity for oxygen with a low oxygen K_m of 0.5 ± 4.0 nM (Bristow et al., 2016). It has been suggested that oxygen production by *Prochlorococcus* in the ODZ can fuel nitrite oxidation (Garcia-Robledo et al., 2017).

The three examples above, *Cand.* Scalindua, MGI Thaumarchaeota, and *Nitrospina,* demonstrate that the read placement technique produces consistent results when applied to multiple genes in the same organism, supporting the use of normalized read placement as a semi quantitative method to examine microbial depth profiles. This approach is also valuable for understanding distributions of as yet uncharacterized organisms. For example, previously, the phylogeny of metagenomic and metatranscriptomic reads in ODZs have been identified with BLAST using the NCBI nucleotide collection database (nt/nr) (Ganesh et al., 2014, 2015; Glass et al., 2015). When re-examining this previous data with our read placement approach that incorporates the corresponding metagenome assemblies on the reference tree, we find the same functional gene identification (*narG*, *nirK* etc) as found with BLAST, but the phylogenetic affiliation of each gene can be better determined. BLAST identifications depend greatly on the composition of the database (Fuchsman and Rocap, 2006), and is a determination via local alignment. The placement approach can provide a higher resolution determination among closely related organisms in part because the whole sequence is used for comparison to a known reference (Berger et al., 2011). For example, using BLAST against NCBI nt/nr database, the phylogenetic identity of nitric oxide reductase transcripts from the coastal ETNP in 2013 was highly variable with depth (Ganesh et al., 2015). However, here these transcripts consistently place on three uncultured ETNP *qnor*-like OTUs assembled from our metagenome, which were not present in the original NCBI nt/nr BLAST database (**Figure 7**). Similarly, the identity of N_2O reductase *nosZ* transcripts from this same ETNP metatranscriptome were reported as predominantly unknown prokaryote (Ganesh et al., 2015). Read placement of those transcripts on a phylogenetic tree indicates that the unknown prokaryote transcripts are all OTU II (**Figure 8**), which can now be linked to other N cycling genes (**Figure 6**) through our assembly and to a particle-attached lifestyle (**Figure 8**). Thus, read placement techniques in combination with assemblies from the relevant environment allow us to take the next step in understanding the community in ODZs.

N_2O and N_2 Production in the Oxycline

We apply this holistic approach to analysis of N cycling genes with chemical measurements and rate data to the oxycline above the ETNP ODZ, which has been implicated as a potential source of N_2O, a potent greenhouse gas, to the atmosphere (Cohen and Gordon, 1978; Yamagishi et al., 2007; Babbin et al., 2015).

Here, concentrations of N_2O were highest (106 nM) at 90 m within the oxycline (**Figure 1F**, Peng et al., 2015). In addition to production by nitric oxide reductases during denitrification, N_2O is also produced by ammonia oxidizing archaea (Santoro et al., 2011). The enzymes involved in this process are unclear and the last step of N_2O production may be non-enzymatic in archaea (Kozlowski et al., 2016). From isotopomer analysis of N_2O in the upper ETNP oxycline, ammonia oxidizers contribute more than denitrifiers to N_2O production, but in the lower oxycline both processes may be present (Yamagishi et al., 2007). Indeed, in the ETSP incubation experiments indicated that both ammonia oxidation and denitrification contributed to N_2O production in the lower oxycline (Ji et al., 2015). Here, at station BB2, ammonia oxidation rates, which were solely attributed to archaea, were measurable throughout the oxycline, corresponding with the N_2O maximum, and were still 13 nM/d at 103 m (Peng et al., 2015). This is consistent with our metagenomic data indicating MGI Thaumarchaeota made up to 12% of the community at 100 m, corresponding to the upper N_2O maximum (**Figures 2E,F, 5**). However, if ammonia-oxidizing archaea were the only source of N_2O here, the by-product N_2O would be ∼20% of oxidized ammonia in the oxycline at BB2 (Peng et al., 2015). N_2O yields of marine archaea under normal oxygen conditions are <1% (Santoro et al., 2011; Loscher et al., 2012) though yields up to 1.6% were found in the oxycline of the ETSP (Ji et al., 2015). Thus, some form of denitrification in the lower oxycline (90–100 m) is necessary to explain the N_2O measurements (Babbin et al., 2015; Peng et al., 2015). However, here no nitric oxide reductases (either *norB* or *qnorB-like*) were detected in the water column at 90 m and *norB* was only present in 1% of the community at 100 m in our whole water samples (**Figure 1**). We suggest production of N_2O inside particles can explain these N_2O measurements. Although *qnorB-like* nitric oxide reductase was not detected in the water column in the oxycline at station 136, it was present in particles (5.5% of the particle community) at the base of the oxycline (100 m) at station BB2 and *norB* was also present (2.2% of the particle community) (Figure S8). Thus, particle communities in the oxycline have the capacity to produce N_2O. This is understandable as there are oxygen gradients inside marine organic particles (Ploug et al., 1997), and measurements of diatom aggregates indicate that at ∼100 μmol L^{-1} ambient O_2 and below aggregates contained anoxic regions from which N_2O and N_2 production could be measured (Stief et al., 2016). Zooplankton carcasses may also be sites of denitrification below ∼10 μmol L^{-1} ambient O_2 (Stief et al., 2017). Thus, based on the measured STOX oxygen concentrations (4.7 and 0.8 μM at 90 and 100 m) the majority of particles should contain anoxic zones in the ETNP lower oxycline. N_2O production from sediment trap material under oxic conditions has been shown previously in the North Pacific (Wilson et al., 2014).

N_2O is also potentially consumed in the lower oxycline, forming N_2. Although denitrification can be 50% inhibited by 200–300 nM O_2 (Dalsgaard et al., 2014), N_2 production in the oxycline is consistent with the N_2 gas concentration profile (**Figure 1H**). Potential N_2O reduction rates extend into the oxycline (1 nM/d at 85 m), though vials for these rates were incubated anaerobically (Babbin et al., 2015). Importantly, measured gas in the water column, either N_2O or N_2, represents both water column and particulate production. As with N_2O production, particles may also be important for N_2 production in the oxycline. N_2O reductase (*nosZ*) was found in 3–5% community in bulk water in the oxycline at station 136 (**Figure 1G**), and only two phylotypes were present (*Flavobacteria* and *Chloroflexi*) (**Figures 8B,C**). This suggests that all denitrifiers may not have the same oxygen inhibition threshold or that *Flavobacteria* and *Chloroflexi* are facultative denitrifiers. Notably, *nosZ* was also enriched in the 100 m particles at station BB2 where it was contained by 13% of the community (Figure S8) represented by *nosZ* OTU II and Flavobacteria *nosZ* (**Figure 8**). Thus, much of the N_2 gas production in the oxycline may occur on particles.

Niche Partitioning

Microbes containing N cycling genes examined here separate into three groups with distinct niches: particle-attached, free-living with tolerance to low levels of oxygen, free-living with preference to complete anoxia. Genes in bacteria attached to particles could either be taking advantage of the abundant organic matter there or of the more reduced conditions found inside particles. It appears that three phylotypes of *narG*, one phylotype of *nirK*, two phylotype of *nirS*, two phylotype of *qnor*-like and two phylotype of *nosZ* genes are particle attached, including the *nosZ* affiliated with S-oxidizers (**Figures 4**, **5**, **7**, **8**, Figures S9, S13). In the water column, genes can be separated into two niches based on oxygen content. Genes found in the upper ODZ are likely exposed to nanomolar oxygen either by mixing or by O_2 production by ODZ *Prochlorococcus* (Garcia-Robledo et al., 2017). Some OTUs appear to be tolerant of oxygen while others are only abundant below 120 m after all possible O_2 has been removed. Genes from oxygen utilizing microbes, ammonia oxidizing archaea and nitrite oxidizing bacteria were found in the upper ODZ (**Figure 9**). Genes for anammox bacteria were abundant below 120 m (**Figure 9**), which could imply a lack of tolerance to oxygen or competition with ammonia and nitrite oxidizers for its reactant ammonium (Penn et al., 2016). All of the *nirK* phylotypes that are enriched in bulk water are found in the upper ODZ while all the *nirS* OTUs are abundant below 120 m (**Figure 5**, Figure S9), implying different oxygen tolerances for bacteria with these genes. Additionally, all the *qnor*-like genes were abundant below 120 m (**Figure 7**). Two water column *nosZ* phylotypes were found in the upper ODZ while one free-living *nosZ* was abundant below 120 m (**Figure 8**, Figure S13), indicating diversity of oxygen tolerance between types of denitrifiers. The free-living SAR11 *narG* phylotypes were abundant throughout the ODZ, but not in the oxycline, so did not clearly fall into our defined water-column niches (**Figure 4**, Figure S13). Due to their differing depth profiles and niches, it is possible that not all these free-living microbes have all the genes necessary to perform complete denitrification. In all, these data highlight the diversity of N reducing microbes in the ODZ.

Synthesis

Here we describe gene abundances and distinct phylotypes for all key steps in the low oxygen N cycle in an open-ocean ODZ community. Although gene presence only indicates potential activity, relative gene abundances here correlated closely with features in chemical profiles and measured rates for denitrification, anammox, ammonia oxidation, and nitrite oxidation (Babbin et al., 2014, 2015; Peng et al., 2015). Both anammox and N_2 producing denitrifiers were present in the water column with denitrifiers found at shallower depths, reaching the oxycline. Overall, denitrification genes, with the exception of nitrate reductase, were enriched in $>30\,\mu m$ particles. However, when examined at a phylotype level, two N_2O reductase phylotypes were enriched on particles while three other *nosZ* phylotypes were actually not enriched in particles. Furthermore, *Flavobacteria* and *Chloroflexi nosZ* phylotypes had maxima within lower oxycline, while the remaining *nosZ* phylotypes had maxima within the ODZ. These data highlight the diversity of denitrifiers both phylogenetically and potentially functionally.

While we did not find evidence for denitrification without an N_2O intermediate by autotrophic methane-oxidizers, autotrophic denitrification is still a distinct probability in the ETNP. The presence of a C-terminal extension on *nosZ* in a group of assembled contigs indicated the presence of a S-oxidizing autotrophic denitrifying phylotype on particles. A S-oxidizing autotrophic denitrifier was previously found on particles in the suboxic zone of the Black Sea (Fuchsman et al., 2012a). These data support the possibility of low level but widespread S cycling in particles under low oxygen conditions.

The largest N_2O maxima in the ETNP are in the oxycline above the ODZ. Since ammonia oxidation rates are too low to support all the N_2O production in the oxycline (Peng et al., 2015), an additional source of N_2O production is likely (Babbin et al., 2015). Our metagenomic data supports production of N_2O inside particles in the oxycline, rather than denitrification in the water column.

In a warming ocean with lower oxygen concentrations, the area of both ODZs and the oxycline above them may expand. Understanding the diversity and function of water column and particle associated communities in these regions may be critical for correctly predicting the magnitude of N loss and N_2O release to the atmosphere.

AUTHOR CONTRIBUTIONS

CF: collected the samples, created metagenomic libraries, designed, and performed analyses and wrote the paper; JS: wrote code essential to the analyses and edited the paper; CM: assembled the metagenomes, wrote code used in analyses, and edited the paper; AD and GR: helped design analyses and edited the paper.

ACKNOWLEDGMENTS

We thank William Brazelton for helpful discussions. Captain and crew of the R/V Thompson. Bonnie X. Chang for OPA ammonia data. Aaron Morello for shipboard nutrient analyses. This study was funded by NSF grants to AD. OCE-1029316 and to GR. OCE-1138368.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02384/full#supplementary-material

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119. doi: 10.1093/nar/gkh131

Babbin, A. R., Bianchi, D., Jayakumar, A., and Ward, B. B. (2015). Rapid nitrous oxide cycling in the suboxic ocean. *Science* 348, 1127–1129. doi: 10.1126/science.aaa8380

Babbin, A. R., Keil, R. G., Devol, A. H., and Ward, B. B. (2014). Oxygen control nitrogen loss in the ocean. *Science* 344:406. doi: 10.1126/science.1248364

Berger, S. A., Krompass, D., and Stamatakis, A. (2011). Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* 60, 291–302. doi: 10.1093/sysbio/syr010

Berger, S. A., and Stamatakis, A. (2011). Aligning short reads to reference alignments and trees. *Bioinformatics* 27, 2068–2075. doi: 10.1093/bioinformatics/btr320

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bowen, J. L., Weisman, D., Yasuda, M., Jayakumar, A., Morrison, H. G., and Ward, B. B. (2015). Marine oxygen-deficient zones harbor depauperate denitrifying communities compared to novel genetic diversity in coastal sediments. *Microb. Ecol.* 70, 311–321. doi: 10.1007/s00248-015-0582-y

Bristow, L. A., Dalsgaard, T., Tiano, L., Mills, D. B., Bertagnolli, A. D., Wright, J. J., et al. (2016). Ammonium and nitrite oxidation at nanomolar oxygen concentrations in oxygen minimum zone waters. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10601–10606. doi: 10.1073/pnas.1600359113

Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., and Brom, T. H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. *arXiv:1203.4802v2.*

Canfield, D. E., Stewart, F. J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., DeLong, E. F., et al. (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the chilean coast. *Science.* 330:1375. doi: 10.1126/science.1196889

Castro-González, M., Ulloa, O., and Farías, L. (2015). Structure of denitrifying communities reducing N 2 O at suboxic waters off northern Chile and Perú. *Rev. Biol. Mar. Oceanogr.* 50, 95–110. doi: 10.4067/S0718-19572015000100008

Chang, B. X., Devol, A. H., and Emerson, S. R. (2010). Denitrification and the nitrogen gas excess in the eastern tropical South Pacific oxygen deficient zone. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 57, 1092–1101. doi: 10.1016/j.dsr.2010.05.009

Chang, B. X., Devol, A. H., and Emerson, S. R. (2012). Fixed nitrogen loss from the eastern tropical North Pacific and Arabian Sea oxygen deficient zones determined from measurements of N 2 :Ar. *Global Biogeochem. Cycles* 26:GB004207. doi: 10.1029/2011GB004207

Chang, B. X., Rich, J. R., Jayakumar, A., Naik, H., Pratihary, A. K., Keil, R. G., et al. (2014). The effect of organic carbon on fixed nitrogen loss in the eastern tropical

South Pacific and Arabian Sea oxygen deficient zones. *Limnol. Oceanogr.* 59, 1267–1274. doi: 10.4319/lo.2014.59.4.1267

Clegg, S. L., and Whitfield, M. (1990). A generalized model for the scavenging of trace metals in the open ocean—I. Particle cycling. *Deep Sea Res. Part A. Oceanogr. Res. Pap.* 37, 809–832. doi: 10.1016/0198-0149(90)90008-J

Cohen, Y., and Gordon, L. (1978). Nitrous oxide in the oxygen minimum of eastern tropical North Pacific: evidence for its consumption during denitrification and possble mechanisms for its production. *Deep Sea Res.* 25, 509–524. doi: 10.1016/0146-6291(78)90640-9

Crusoe, M. R., Alameldin, H. F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., et al. (2015). The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res.* 4:900. doi: 10.12688/f1000research.6924.1

Dalsgaard, T., Stewart, F. J., Thamdrup, B., De Brabandere, L., Revsbech, P., and Ulloa, O. (2014). Oxygen at nanomolar levels reversibly suppresses process rates and gene expression in anammox and denitrification in the oxygen minimum zone off Northern Chile. *MBio* 5, e01966–e01914. doi: 10.1128/mBio.01966-14

Dalsgaard, T., Thamdrup, B., Farías, L., and Peter Revsbech, N. (2012). Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnol. Oceanogr.* 57, 1331–1346. doi: 10.4319/lo.2012.57.5.1331

Delong, E. F., Franks, D. G., and Alldredge, A. L. (1993). Phylogenetic diversity of aggregate-attached marine bacterial assemblages. *Limnol. Oceanogr.* 38, 924–934. doi: 10.4319/lo.1993.38.5.0924

Deutsch, C., Brix, H., Ito, T., Frenzel, H., and Thompson, L. (2011). Climate-forced variability of ocean hypoxia. *Science* 333, 336–339. doi: 10.1126/science.1202422

Devol, A. H. (2003). Solution to a marine mystery. *Nature* 422:575. doi: 10.1038/422575a

DeVries, T., Deutsch, C., Rafter, P. A., and Primeau, F. (2013). Marine denitrification rates determined from a global 3-D inverse model. *Biogeosciences* 10, 2481–2496. doi: 10.5194/bg-10-2481-2013

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Ettwig, K., Butler, M., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M., et al. (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 464:543. doi: 10.1038/nature08883

Ettwig, K. F., Speth, D. R., Reimann, J., Wu, M. L., Jetten, M. S. M., and Keltjens, J. T. (2012). Bacterial oxygen production in the dark. *Front. Microbiol.* 3:273. doi: 10.3389/fmicb.2012.00273

Fuchsman, C. A., Kirkpatrick, J. B., Brazelton, W. J., Murray, J. W., and Staley, J. T. (2011). Metabolic strategies of free-living and aggregate-associated bacterial communities inferred from biologic and chemical profiles in the Black Sea suboxic zone. *FEMS Microbiol. Ecol.* 78, 586–603. doi: 10.1111/j.1574-6941.2011.01189.x

Fuchsman, C. A., Murray, J. W., and Staley, J. T. (2012a). Stimulation of autotrophic denitrification by intrusions of the Bosporus Plume into the anoxic Black Sea. *Front. Microbiol.* 3:257. doi: 10.3389/fmicb.2012.00257

Fuchsman, C. A., and Rocap, G. (2006). Whole-Genome reciprocal BLAST analysis reveals that planctomycetes do not share an unusually large number of genes with Eukarya and Archaea. *Appl. Environ. Microbiol.* 72, 6841–6844. doi: 10.1128/AEM.00429-06

Fuchsman, C. A., Staley, J. T., Oakley, B. B., Kirkpatrick, J. B., and Murray, J. W. (2012b). Free-living and aggregate-associated planctomycetes in the black sea. *FEMS Microbiol. Ecol.* 80, 402–416. doi: 10.1111/j.1574-6941.2012.01306.x

Ganesh, S., Bristow, L. A., Larsen, M., Sarode, N., Thamdrup, B., and Stewart, F. J. (2015). Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone. *ISME J.* 9, 2682–2696. doi: 10.1038/ismej.2015.44

Ganesh, S., Parris, D. J., DeLong, E. F., and Stewart, F. J. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* 8, 187–211. doi: 10.1038/ismej.2013.144

Garcia-Robledo, E., Padilla, C. C., Aldunate, M., Stewart, F. J., Ulloa, O., Paulmier, A., et al. (2017). Cryptic oxygen cycling in anoxic marine zones. *Proc. Natl. Acad. Sci.U.S.A.* 114, 8319–8324. doi: 10.1073/pnas.1619844114

Glass, J. B., Kretz, C. B., Ganesh, S., Ranjan, P., Seston, S. L., Buck, K. N., et al. (2015). Meta-omic signatures of microbial metal and nitrogen cycling in marine oxygen minimum zones. *Front. Microbiol.* 6:998. doi: 10.3389/fmicb.2015.00998

Gordon, L. I., Jennings, J. C., Ross, A. R., and Krest, J. M. (1995). *A Suggested Protocol for Continuous Flow Automoated Analysis of Seawater Nutrients in the WOCE Hydrographic Program and the Hoint Global Ocean Fluxes Study*. Technical Report, Oregon State University.

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, 387–395. doi: 10.1093/nar/gks1234

Hannig, M., Lavik, G., Kuypers, M. M. M., Woebken, D., Martens-Habbena, W., and Jurgens, K. (2007). Shift from denitrification to anammox after inflow events in the central Baltic Sea. *Limnol. Oceanogr.* 52, 1336–1345. doi: 10.4319/lo.2007.52.4.1336

Holmes, R. M., Aminot, A., Kerouel, R., Hooker, B. A., and Peterson, B. J. (1999). A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can. J. Fish. Aquat. Sci.* 56, 1801–1808. doi: 10.1139/f99-128

Hughes, M. N. (2008). Chemistry of nitric oxide and related species. *Methods Enzymol.* 436, 3–19. doi: 10.1016/S0076-6879(08)36001-7

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jayakumar, A., Peng, X., and Ward, B. (2013). Community composition of bacteria involved in fixed nitrogen loss in the water column of two major oxygen minimum zones in the ocean. *Aquat. Microb. Ecol.* 70, 245–259. doi: 10.3354/ame01654

Jensen, M. M., Lam, P., Revsbech, N. P., Nagel, B., Gaye, B., Jetten, M. S., et al. (2011). Intensive nitrogen loss over the Omani Shelf due to anammox coupled with dissimilatory nitrite reduction to ammonium. *ISME J.* 5, 1660–1670. doi: 10.1038/ismej.2011.44

Ji, Q., Babbin, A. R., Jayakumar, A., Oleynik, S., and Ward, B. B. (2015). Nitrous oxide production by nitrification and denitrification in the eastern tropical south pacific oxygen minimum zone. *Geophys. Res. Lett.* 42, 755–764. doi: 10.1002/2015GL066853

Jones, C. M., Stres, B., Rosenquist, M., and Hallin, S. (2008). Phylogenetic analysis of nitrite, nitric oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history for denitrification. *Mol. Biol. Evol.* 25, 1955–1966. doi: 10.1093/molbev/msn146

Kalvelage, T., Lavik, G., Lam, P., Contreras, S., Arteaga, L., Löscher, C. R., et al. (2013). Nitrogen cycling driven by organic matter export in the South Pacific oxygen minimum zone. *Nat. Geosci.* 6, 228–234. doi: 10.1038/ngeo1739

Kampschreur, M. J., Kleerebezem, R., de Vet, W. W. J. M., and Van Loosdrecht, M. C. M. (2011). Reduced iron induced nitric oxide and nitrous oxide emission. *Water Res.* 45, 5945–5952. doi: 10.1016/j.watres.2011.08.056

Kozlowski, J. A., Stieglmeier, M., Schleper, C., Klotz, M. G., and Stein, L. Y. (2016). Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME J.* 10, 1836–1845. doi: 10.1038/ismej.2016.2

Lam, P., Jensen, M. M., Kock, A., Lettmann, K. A., Plancherel, Y., Lavik, G., et al. (2011). Origin and fate of the secondary nitrite maximum in the Arabian Sea. *Biogeosciences* 8, 1565–1577. doi: 10.5194/bg-8-1565-2011

Lam, P., Lavik, G., Jensen, M. M., van de Vossenberg, J., Schmid, M., Woebken, D., et al. (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4752–4757. doi: 10.1073/pnas.0812444106

Loscher, C. R., Kock, A., Konneke, M., LaRoche, J., Bange, H. W., and Schmitz, R. A. (2012). Production of oceanic nitrous oxide by ammonia-oxidizing archaea. *Biogeosciences* 9, 2419–2429. doi: 10.5194/bg-9-2419-2012

Lücker, S., Nowka, B., Rattei, T., Spieck, E., and Daims, H. (2013). The genome of *Nitrospina gracilis* illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front. Microbiol.* 4:27. doi: 10.3389/fmicb.2013.00027

Lüke, C., Speth, D. R., Kox, M. A. R., Villanueva, L., and Jetten, M. S. M. (2016). Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *PeerJ.* 4:e1924. doi: 10.7717/peerj.1924

Magoc, T., and Salzberg, S. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 21. doi: 10.1093/bioinformatics/btr507

NCBI Resource Coordinators (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acid Res.* 4, 152. doi: 10.1093/nar/gkw1071

Orsi, W. D., Smith, J. M., Wilcox, H. M., Swalwell, J. E., Carini, P., Worden, A. Z., et al. (2015). Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J.* 9, 1747–1763. doi: 10.1038/ismej.2014.260

Pack, M. A., Heintz, M. B., Reeburgh, W. S., Trumbore, S. E., Valentine, D. L., Xu, X., et al. (2015). Methane oxidation in the eastern tropical North Pacific Ocean water column. *J. Geophys. Res. Biogeosci.* 120, 1078–1092. doi: 10.1002/2014JG002900

Padilla, C. C., Bristow, L. A., Sarode, N., Garcia-robledo, E., Ramírez, E. G., Benson, C. R., et al. (2016). NC10 bacteria in marine oxygen minimum zones. *ISME J.* 10, 2067–2071. doi: 10.1038/ismej.2015.262

Peng, X., Fuchsman, C. A., Jayakumar, A., Oleynik, S., and Martens-habbena, W. (2015). Ammonia and nitrite oxidation in the Eastern Tropical North Pacific. *Global Biogeochem. Cycles* 29, 2034–2049. doi: 10.1002/2015GB005278

Penn, J., Weber, T., and Deutsch, C. (2016). Microbial functional diversity alters the structure and sensitivity of oxygen deficient zones. *Geophys. Res. Lett.* 43, 9773–9780. doi: 10.1002/2016GL070438

Peters, B. D., Babbin, A. R., Lettmann, K. A., Mordy, C. W., Ulloa, O., Ward, B. B., et al. (2016). Vertical modeling of the nitrogen cycle in the eastern tropical South Pacific oxygen deficient zone using high-resolution concentration and isotope measurements. *Global Biogeochem. Cycles* 30:GB005415. doi: 10.1002/2016GB005415

Ploug, H., Grossart, H.-P., Azam, F., and Jorgensen, B. B. (1999). Photosynthesis, respiration, and carbon turnover in sinking marine snow from surface waters of Southern California Bight : implications for the carbon cycle in the ocean. *Mar. Ecol. Prog. Ser.* 179, 1–11. doi: 10.3354/meps179001

Ploug, H., Kiihl, M., Buchholz-cleven, B., and Jergensen, B. B. (1997). Anoxic aggregates - an ephemeral phenomenon in the pelagic environment ? *Aquat. Microb. Ecol.* 13, 285–294. doi: 10.3354/ame013285

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35, 501–504. doi: 10.1093/nar/gkl842

Punta, M., Coggill, P., Eberhardt, R., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families databases. *Nucleic Acids Res.* 40, D290–D301. doi: 10.1093/nar/gkr1065

Santoro, A. E., Buchwald, C., McIlvin, M. R., and Casciotti, K. L. (2011). Isotopic signature of N2O produced by marine ammonia-oxidizing archaea. *Science* 333, 1282–1285. doi: 10.1126/science.1208239

Saunders, J. K., and Rocap, G. (2016). Genomic potential for arsenic efflux and methylation varies among global Prochlorococcus populations. *ISME J.* 10, 197–209. doi: 10.1038/ismej.2015.85

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sinninghe Damsté, J. S., Strous, M., Rijpstra, W. I. C., Hopmans, E. C., Geenevasen, J. A. J., van Duin, A. C. T., et al. (2002). Linearly concatenated cyclobutane lipids form a dense bacterial membrane. *Nature* 419, 708–712. doi: 10.1038/nature01128

Sollai, M., Hopmans, E. C., Schouten, S., Keil, R. G., and Sinninghe Damsté, J. S. (2015). Intact polar lipids of Thaumarchaeota and anammox bacteria as indicators of N-cycling in the Eastern Tropical North Pacific oxygen deficient zone. *Biogeosci. Discuss.* 12, 4833–4864. doi: 10.5194/bgd-12-4833-2015

Spieck, E., and Bock, E. (2015). "Nitrifying bacteria," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed W. B. Whitman (John Wiley & Sons, Inc.), 1–8. doi: 10.1002/9781118960608.bm00016

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stewart, F. J., Ulloa, O., and DeLong, E. F. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ. Microbiol.* 14, 23–40. doi: 10.1111/j.1462-2920.2010.02400.x

Stief, P., Kamp, A., Thamdrup, B., and Glud, R. N. (2016). Anaerobic nitrogen turnover by sinking diatom aggregates at varying ambient oxygen levels. *Front. Microbiol.* 7:98. doi: 10.3389/fmicb.2016.00098

Stief, P., Lundgaard, A. S. B., Morales-Ramírez, Á., Thamdrup, B., and Glud, R. N. (2017). Fixed-nitrogen loss associated with sinking zooplankton carcasses in a coastal oxygen minimum zone (Golfo Dulce, Costa Rica). *Front. Mar. Sci.* 4:152. doi: 10.3389/fmars.2017.00152

Tiano, L., Garcia-Robledo, E., Dalsgaard, T., Devol, A. H., Ward, B. B., Ulloa, O., et al. (2014). Oxygen distribution and aerobic respiration in the north and south eastern tropical Pacific oxygen minimum zones. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 94, 173–183. doi: 10.1016/j.dsr.2014.10.001

Tsementzi, D., Wu, J., Deutsch, S., Nath, S., Rodriguez-r, L. M., Burns, A. S., et al. (2016). SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* 536, 179–183. doi: 10.1038/nature19068

van de Vossenberg, J., Woebken, D., Maalcke, W. J., Wessels, H. J., Dutilh, B. E., Kartal, B., et al. (2013). The metagenome of the marine anammox bacterium "*Candidatus* Scalindua profunda" illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ. Microbiol.* 15, 1275–1289. doi: 10.1111/j.1462-2920.2012.02774.x

Ward, B. B., Devol, A. H., Rich, J. J., Chang, B. X., Bulow, S. E., Naik, H., et al. (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature* 461:78. doi: 10.1038/nature08276

Wilson, S. T., del Valle, D. A., Segura-Noguera, M., and Karl, D. M. (2014). A role for nitrite in the production of nitrous oxide in the lower euphotic zone of the oligotrophic North Pacific Ocean. *Deep. Res. Part I Oceanogr. Res. Pap.* 85, 47–55. doi: 10.1016/j.dsr.2013.11.008

Yamagishi, H., Westley, M. B., Popp, B. N., Toyoda, S., Yoshida, N., Watanabe, S., et al. (2007). Role of nitrification and denitrification on the nitrous oxide cycle in the eastern tropical North Pacific and Gulf of California. *J. Geophys. Res. Biogeosci.* 112:G02015. doi: 10.1029/2006JG000227

Zerbino, D. R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinformatics* Unit−11.5, 1–13. doi: 10.1002/0471250953.bi1105s31

Zhang, Q., Awad, S., and Brown, C. (2015). Crossing the streams: a framework for streaming analysis of short DNA sequencing reads. *PeerJ* 3:e890v1. doi: 10.7287/peerj.preprints.890v1

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* 61, 533–616.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Fuchsman, Devol, Saunders, McKay and Rocap. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 August 2017
doi: 10.3389/fmicb.2017.01621

FixK$_2$ Is the Main Transcriptional Activator of *Bradyrhizobium diazoefficiens nosRZDYFLX* Genes in Response to Low Oxygen

María J. Torres‡, Emilio Bueno†‡, Andrea Jiménez-Leiva, Juan J. Cabrera, Eulogio J. Bedmar, Socorro Mesa and María J. Delgado**

Department of Soil Microbiology and Symbiotic Systems, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, Granada, Spain

The powerful greenhouse gas, nitrous oxide (N_2O) has a strong potential to drive climate change. Soils are the major source of N_2O and microbial nitrification and denitrification the main processes involved. The soybean endosymbiont *Bradyrhizobium diazoefficiens* is considered a model to study rhizobial denitrification, which depends on the *napEDABC*, *nirK*, *norCBQD*, and *nosRZDYFLX* genes. In this bacterium, the role of the regulatory cascade FixLJ-FixK$_2$-NnrR in the expression of *napEDABC*, *nirK*, and *norCBQD* genes involved in N_2O synthesis has been previously unraveled. However, much remains to be discovered regarding the regulation of the respiratory N_2O reductase (N_2OR), the key enzyme that mitigates N_2O emissions. In this work, we have demonstrated that *nosRZDYFLX* genes constitute an operon which is transcribed from a major promoter located upstream of the *nosR* gene. Low oxygen was shown to be the main inducer of expression of *nosRZDYFLX* genes and N_2OR activity, FixK$_2$ being the regulatory protein involved in such control. Further, by using an *in vitro* transcription assay with purified FixK$_2$ protein and *B. diazoefficiens* RNA polymerase we were able to show that the *nosRZDYFLX* genes are direct targets of FixK$_2$.

Keywords: climate change, denitrification, greenhouse gas, nitrous oxide, nitrous oxide reductase, regulation

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Stephen Spiro, University of Texas at Dallas, United States
Rosa María Martínez-Espinosa, University of Alicante, Spain
James Moir, University of York, United Kingdom

****Correspondence:***
María J. Delgado
mdelgado@eez.csic.es
Socorro Mesa
socorro.mesa@eez.csic.es

†Present address:
Emilio Bueno, Laboratory for Molecular Infection Medicine Sweden (MIMS), Department of Molecular Biology, Umeå University, Umeå, Sweden

‡These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *07 June 2017*
Accepted: *09 August 2017*
Published: *30 August 2017*

Citation:
Torres MJ, Bueno E, Jiménez-Leiva A, Cabrera JJ, Bedmar EJ, Mesa S and Delgado MJ (2017) FixK$_2$ Is the Main Transcriptional Activator of Bradyrhizobium diazoefficiens nosRZDYFLX Genes in Response to Low Oxygen. Front. Microbiol. 8:1621. doi: 10.3389/fmicb.2017.01621

INTRODUCTION

Nitrous oxide (N_2O) is a powerful greenhouse gas (GHG) and a major cause of ozone layer depletion with an atmospheric lifetime of 114 years and, based on its radiative capacity, an estimated 300-fold greater potential for global warming compared with that of carbon dioxide (CO_2). Hence, N_2O accounts for approximately 10% of total emissions with respect to the impact of each individual GHGs on global warming (Intergovernmental Panel on Climate Change [IPCC], 2014). Due to its environmental impact, a better understanding of the pathways implicated in the generation and consumption of N_2O has received great interest (Thomson et al., 2012).

Despite the existence of multiple pathways for N_2O generation in soils such as nitrifier denitrification, nitrite oxidation, heterotrophic denitrification, ammonia oxidation, anaerobic ammonium oxidation (anammox) and dissimilatory nitrate reduction to ammonium (DNRA), it is generally assumed that nitrification and denitrification are the principal processes that contribute to the emissions of N_2O from terrestrial ecosystems (for a review see

Stein, 2011; Schreiber et al., 2012; Butterbach-Bahl et al., 2014). Denitrification is widespread within the domain of *Bacteria* being dominant within *Proteobacteria* (Shapleigh, 2006). However, it has been evinced that some archaea (Treusch et al., 2005) and fungi (Takaya, 2002; Prendergast-Miller et al., 2011) may also denitrify. Most of the studies about denitrification have been focused on Gram-negative bacteria that occupy terrestrial niches, using the alpha-proteobacterium *Paracoccus* (*Pa.*) *denitrificans* as well as the gamma-proteobacteria *Pseudomonas* (*Ps.*) *stutzeri* and *Ps. aeruginosa* as model organisms (Zumft, 1997). The reactions of denitrification are catalyzed by periplasmic (Nap) or membrane-bound (Nar) nitrate reductase, nitrite reductases (NirK/NirS), nitric oxide (NO) reductases (cNor, qNor, or Cu_ANor) and nitrous oxide reductase (N_2OR) encoded by *nap*/*nar*, *nirK*/*nirS*, *nor,* and *nos* genes, respectively. The physiological, biochemical and molecular aspects of denitrification have been covered by a collection of reviews published elsewhere (Zumft, 1997; van Spanning et al., 2005, 2007; Kraft et al., 2011; Richardson, 2011; Bueno et al., 2012).

In contrast to the numerous sources of N_2O, nitrous oxide reductase (NosZ) is the only known biological enzyme involved in its removal by reduction to N_2 (reviewed by Thomson et al., 2012). A new cluster of atypical *nosZ* genes, designated clade II, have been recently identified (Sanford et al., 2012; Jones et al., 2013) which are also present in genomes lacking the *nirS* and/or *nirK* gene. This suggests that non-denitrifiers also contribute to N_2O removal (Jones et al., 2013).

Nitrous oxide reductase is a homodimer with molecular weight of 120−160 kDa, a copper content of ∼12 Cu atoms, and a sulfide content of ∼2 S^{2-} ions per dimer (Rasmussen et al., 2000). The enzyme contains two copper sites: Cu_A, and Cu_Z, a tetranuclear μ4-sulfide-bridged cluster liganded by seven histidine residues, which has been proposed to be the active center for N_2O reduction. The expression, maturation, and maintenance of the NosZ catalytic subunit require several other auxiliary proteins (Zumft, 2005) being all encoded together by a typical gene cluster that contains six genes (*nosRZDFYL*). This core cluster is, in some cases, associated with an additional gene, *nosX* (reviewed by Zumft and Kroneck, 2007). Mutation analyses demonstrated that NosDFY or NosL are involved in the maturation of the NosZ Cu_Z, but not in the biogenesis of the Cu_A site (reviewed by Zumft and Kroneck, 2007; van Spanning, 2011). NosR and NosX do not participate in Cu_Z biogenesis but do play a role in N_2O reduction *in vivo* altering the state of the Cu_Z site during turnover and supporting the catalytic activity of NosZ (Wunsch and Zumft, 2005). NosR, apart from its putative role as electron donor to NosZ, might also act as a regulator, since it is needed for *Ps. stutzeri nosZ* and *nosD* transcription (Honisch and Zumft, 2003).

Low O_2 conditions and NO have been suggested as the main signal molecules for induction of *nos* genes expression (reviewed by Zumft and Kroneck, 2007). Both signals are perceived and transduced via transcriptional regulators belonging to the cyclic AMP receptor protein (CRP)/fumarate and nitrate reductase (FNR) superfamily. This family carries diverse mnemonics, such as ANR, DNR, NNR, NnrR, FNR or FixK but all refer to the same type of regulatory protein with similar domain structure. Proteins that form part of the DNR clade such as DNR/DnrD/NNR from *Ps. aeruginosa, Ps. stutzeri*, and *Pa. denitrificans*, respectively (van Spanning et al., 1999; Vollack and Zumft, 2001; Zumft and Kroneck, 2007; Arai et al., 2013), control *nos* genes expression in response to NO, while low oxygen is perceived by [4Fe-4S]$^{2+}$ cluster-containing FNR- and FnrP-type proteins such as *Pa. denitrificans* FnrP (Bergaust et al., 2012) or *Ps. aeruginosa* ANR (Trunk et al., 2010).

Bradyrhizobium diazoefficiens (Delamuta et al., 2013; formerly *B. japonicum*), the endosymbiont of soybeans, possesses the ability to denitrify under both free-living and symbiotic lifestyles. In *B. diazoefficiens* the denitrification process depends on the *napEDABC, nirK, norCBQD,* and *nosRZDYFLX* genes, coding for Nap, copper-containing NirK, *c*-type Nor and the N_2OR, respectively (Velasco et al., 2001, 2004; Mesa et al., 2002; Delgado et al., 2003; Bedmar et al., 2005).

Expression of *B. diazoefficiens* denitrification genes required low oxygen tension and in the case of *norCBQD* genes the presence of NO is also needed (Bueno et al., 2017). In this bacterium, perception and transduction of the 'low-oxygen' signal are mediated by a complex network comprising two interconnected regulatory cascades, the FixLJ–$FixK_2$–NnrR and the RegSR–NifA (Sciotti et al., 2003). In the latter cascade, an oxygen concentration at or below 0.5% is required for activation of the oxygen-sensitive NifA protein and subsequent induction of essential nitrogen fixation genes (Sciotti et al., 2003). Under anoxic conditions in the presence of NO_3^-, NifA is also necessary for the maximal expression of *napE-lacZ*, *nirK-lacZ,* and *norC-lacZ* fusions (Bueno et al., 2010). Moreover, global transcription analyses of a *regR* mutant in comparison to the wild-type (WT), both grown in anoxic denitrifying conditions showed that RegR is also involved in the regulation of *B. diazoefficiens norCBQD* and *nosRZDYFLX* genes (Torres et al., 2014).

In contrast as reported for the RegSR-NifA cascade, activation of expression of the FixLJ-$FixK_2$-NnrR-dependent targets requires a moderate decrease in the oxygen concentration in the gas phase (≤5%), where the haem-based sensory kinase FixL senses the 'low-oxygen' signal, phosphorylates itself and transfers the phosphoryl group to the FixJ response regulator. Then, FixJ activates transcription of the *fixK*$_2$ gene, encoding the $FixK_2$ protein, a CRP/FNR-like transcriptional regulator. $FixK_2$ induces, in turn, expression of the *napEDABC, nirK,* and *norCBQD* denitrification genes involved in N_2O production (Velasco et al., 2001; Mesa et al., 2002; Robles et al., 2006) as well as other regulatory genes [e.g., *rpoN*$_1$, *fixK*$_1$, and *nnrR*; (Nellen-Anthamatten et al., 1998; Mesa et al., 2003, 2008)]. The latter, the CRP/FNR-type NnrR protein adds an additional control level to the FixLJ-$FixK_2$ cascade integrating the NOx signal necessary for induction of *norCBQD* genes expression (Mesa et al., 2003; Bueno et al., 2017). Within the CRP/FNR family, $FixK_2$ belongs to the FixK subgroup, whose members, in contrast to the O_2-sensitive proteins *Ps. aeruginosa* ANR and *Pa. denitrificans* FnrP, lack the cysteine motif required to bind an [4Fe-4S]$^{2+}$ cluster (reviewed in Korner et al., 2003; Mesa et al., 2006). Particularly, $FixK_2$ activity is subjected to posttranslational control by oxidation of its singular cysteine residue at position 183 (Mesa et al., 2009).

B. diazoefficiens NnrR forms part of the NnrR clade, proteins that cover a similar function to the one defined for DNR-type proteins on the control of denitrification genes expression in response to NO (Bueno et al., 2017). Recently, we observed that *B. diazoefficiens napEDABC, nirK,* and *norCBQD* promoters exhibited differences with regard to their dependence on low oxygen (microoxia), NOx, and the regulatory proteins $FixK_2$ and NnrR. While microoxic conditions were sufficient to induce expression of *napEDABC* and *nirK* genes and this control directly depends on $FixK_2$, *norCBQD* genes expression depends on NO, NnrR being the candidate that directly interacts with *norCBQD* promoter (Bueno et al., 2017).

As described for other CRP/FNR members, $FixK_2$ acts as a dimeric form which binds to a twofold symmetric DNA sequence present at distinct distances within the promoter region of regulated genes (Browning and Busby, 2004). Specifically, the $FixK_2$ box corresponds to TTG(A/C)-N_6-(T/G)CAA (Bonnet et al., 2013), which matches reasonably well with the previously described consensus binding site for FixK-type proteins (TTGA-N_6-TCAA) (Fischer, 1994; Dufour et al., 2010).

While substantial progress has been made on the external signals (microxia and NO) and the manner by which the $FixK_2$ and NnrR proteins control the expression of *B. diazoefficiens napEDABC*, *nirK,* and *norCBQD* genes involved in N_2O synthesis, the regulation of *nosRZDYFLX* genes involved in N_2O reduction to N_2, the key step to N_2O mitigation, has been very poorly explored in this bacterium. In the present work, we show the transcriptional arrangement of the *nosRZDYFLX* genes in *B. diazoefficiens*. We also expanded the knowledge on *nosRZDYFLX* regulation by studying the involvement of low oxygen, and NOx in *nos* expression as well as the role of $FixK_2$ and NnrR regulatory proteins in this control. By using *in vitro* transcription (IVT) activation assays we demonstrated, for first time, that the *nosRZDYFLX* genes are direct targets of $FixK_2$.

MATERIALS AND METHODS

Bacterial Strains, Media, and Growth Conditions

Bacterial strains used in this work are compiled in **Table 1**. *Escherichia coli* cells were cultivated in Luria Bertani medium (Miller, 1972) at 37°C. When needed, antibiotics were used at the following concentrations (in μg/ml): ampicillin, 200; kanamycin, 30; spectinomycin, 25; streptomycin, 25; tetracycline, 10.

Bradyrhizobium diazoefficiens cells were cultured oxically and microoxically basically as described earlier (Bueno et al., 2017). While Peptone-Salts-Yeast extract (PSY) medium (Regensburger and Hennecke, 1983; Mesa et al., 2008) was employed in routine oxic cultures, Yeast Extract-Mannitol (YEM) medium (Daniel and Appleby, 1972) was used as standard medium in our experiments. After growth under oxic conditions in PSY medium, cells were collected by centrifugation (8.000 *g* for 10 min at 4°C), and washed twice with YEM medium. Next, washed cells were used to inoculate, at a 600 nm optical density (OD_{600}) of 0.2, 17 ml or 500 ml rubber stoppered tubes or Erlenmeyer flasks containing 3 ml or 150 ml of YEM medium amended or not with 10 mM KNO_3, respectively. Next, cells were incubated for 24 h under low oxygen conditions, either at initial 0.5% O_2 or at 2% O_2 (in this case the headspace was exchanged every 8–16 h). The latter conditions were chosen to study the specific control of the $FixK_2$ and NnrR regulatory proteins. To analyze the effect of the different NOx, microoxically incubated cells were subsequently exposed for 5 h to 10 mM KNO_3, 500 μM $NaNO_2$, 50 μM NO (from a saturated NO solution [1.91 mM at 20°C]), and 0.15% (30 mM) N_2O. 10 μM or 100 μM of the NO-scavenger cPTIO [2-(4-Carboxyphenyl)-4,4,5,5-tetramethylimidazoline-1-oxyl-3-oxide; carboxy-PTIO potassium salt; Sigma] was added from the beginning to the WT and Δ*nnrR* strain cultures grown microoxically (2% O_2) in the presence of 10 mM KNO_3 for 24 h, in order to analyze the effect of removing the excess of NO on the expression of *nosR-lacZ* or N_2OR activity, respectively. Antibiotics were added to the *B. diazoefficiens* cultures at the following concentrations (μg/ml); chloramphenicol, 20; streptomycin, 200; kanamycin, 200; tetracycline, 100 (solid cultures), 25 (liquid cultures); spectinomycin, 200.

Plasmids and Bacterial Strains Construction

Plasmids used in this study are listed in **Table 1**. Primer sequences in this work are compiled in Supplementary Table S1. For construction of transcriptional reporter fusion plasmids, 5′ DNA fragments for the *nosR* (558; 132; 128 and 75 bp), *nosZ* (1024 bp) and *nosD* (875 pb) promoter regions were amplified using primers' pair a1/PnosR.r, PnosRfull.f/PnosR.r, PnosRhalf.f/PnosR.r, PnosRno.f/PnosR.r, PnosZ.f/PnosZ.r and c1/c2, respectively (Supplementary Table S1). The PCR products were then individually ligated into the pGEM®-T vector (Promega), digested with EcoRI or EcoRI-PstI and cloned into the *lacZ* fusion suicide vector pSUP3535 (Mesa et al., 2003), to yield plasmids pBG0301, pBG0304, pBG0305, pBG0306, pBG0302, and pBG0303, respectively (see **Table 1** for details). The correct orientation of the inserts was verified by sequencing. Plasmids pBG0301, pBG0302, pBG0303, pBG0304, pBG0305, and pBG0306 were integrated by homologous recombination into the chromosome of WT *B. diazoefficiens* 110*spc*4, yielding strains 110*spc*4-BG0301, 110*spc*4-BG0302, 110*spc*4-BG0303, 110*spc*4-BG0304, 110*spc*4-BG0305, 110*spc*4-BG0306. Plasmid pBG0301 was also integrated into the chromosome of *napA* (GRAP1), *nirK* (GRK308), *fixK*$_2$ (9043), and *nnrR* (8678) mutants, yielding strains GRPA1-BG0301, GRK308-BG0301, 9043-BG0301, and 8678-BG0301, respectively (**Table 1**). Correct recombination into the chromosome of the corresponding recipient strain was checked by PCR analyses.

The plasmid used as transcription template was based on the plasmid pRJ9519 which contains a *B. diazoefficiens rrn* transcriptional terminator (Beck et al., 1997). The *nosRZDFYLX* promoter was PCR-amplified with nosR_For_Transc and nosR_Rev_Transc primers, subsequently restricted with XbaI and EcoRI, and finally cloned as a 486-bp fragment into pRJ9519, yielding plasmid pDB4020. The correct nucleotide sequence was confirmed by sequencing.

TABLE 1 | Bacterial strains and plasmids used in this study.

Strains and plasmids	Relevant description	Source of reference
Strains		
E. coli		
DH5α	*supE44* Δ*lacU*169 (ϕ80 *lacZ* ΔM15) *hsdR17 recA1 gyrA96 thi-1 relA1*	Bethesda Research Laboratories Inc., Gaithersburg, MD, United States.
S17.1	Tp^r Sm^r Spc^r *thi, pro*, recA, *hsdR*, *hsdM*, RP4Tc::Mu, Km::Tn7	Simon et al., 1983
BL21 (DE3)	F^- *opmT* $hsdS_B(rB^-$ $mB^-)$ *gal dcm* (DE3)	Novagen Inc.
B. diazoefficiens		
USDA110	Cm^r wild-type	United States Department of Agriculture, Beltsville, MD, United States
110*spc*4	Cm^r Sp^r wild-type, a spectinomycin-resistant derivative of USDA110	Regensburger and Hennecke, 1983
GRPA1	Cm^r Spc^r Sm^r *napA*::Ω	Delgado et al., 2003
GRK308	Cm^r Spc^r Sm^r *nirK*::Ω	Velasco et al., 2001
9043	Cm^r Spc^r Sm^r Δ$fixK_2$::Ω	Nellen-Anthamatten et al., 1998
8678	Cm^r Spc^r Km^r Δ*nnrR*::*aphII*	Mesa et al., 2003
GRZ3035	Cm^r Spc^r Sm^r *nosZ*::Ω	Velasco et al., 2004
110*spc*4-BG0301	Cm^r Sp^r Tc^r *nosR-lacZ* chromosomally integrated into 110*spc*4	This work
110*spc*4-BG0302	Cm^r Sp^r Tc^r *nosZ-lacZ* chromosomally integrated into 110*spc*4	This work
110*spc*4-BG0303	Cm^r Sp^r Tc^r *nosD-lacZ* chromosomally integrated into 110*spc*4	This work
110*spc*4-BG0304	Cm^r Sp^r Tc^r *nosR-lacZ* chromosomally integrated into 110*spc*4 full $FixK_2$-like box	This work
110*spc*4-BG0305	Cm^r Sp^r Tc^r *nosR-lacZ* chromosomally integrated into 110*spc*4 half $FixK_2$-like box	This work
110*spc*4-BG0306	Cm^r Sp^r Tc^r *nosR-lacZ* chromosomally integrated into 110*spc*4 no $FixK_2$-like box	This work
GRPA1-BG0301	Cm^r Sp^r Sm^r Tc^r *nosR-lacZ* chromosomally integrated into GRPA1	This work
GRK308-BG0301	Cm^r Sp^r Sm^r Tc^r *nosR-lacZ* chromosomally integrated into GRK308	This work
9043-BG0301	Cm^r Sp^r Sm^r Tc^r *nosR-lacZ* chromosomally integrated into 9043	This work
8678-BG0301	Cm^r Sp^r Tc^r *nosR-lacZ* chromosomally integrated into 8678	This work
Plasmids		
pSUP3535	Tc^r transcriptional *lacZ* fusion suicide vector	Mesa et al., 2003
pRJ9519	Ap^r [pBluescript SK(+) 308-bp BstXI-KpnI fragment containing the *B. diazoefficiens rrn* terminator cloned into the HincII and KpnI sites]	Beck et al., 1997
pBG0301	Tc^r (pSUP3535) *nosR* 5′ region on a 558-bp EcoRI-PstI fragment	This work
pBG0302	Tc^r (pSUP3535) *nosZ* 5′ region on a 1024-bp EcoRI fragment	This work
pBG0303	Tc^r (pSUP3535) *nosZ* 5′ region on a 875-bp EcoRI fragment	This work
pBG0304	Tc^r (pSUP3535) *nosR* 5′ region containing the intact $FixK_2$-like box on a 224-bp EcoRI fragment	This work
pBG0305	Tc^r (pSUP3535) *nosR* 5′ region containing a partial $FixK_2$-like box on a 218-bp EcoRI fragment	This work
pBG0306	Tc^r (pSUP3535) *nosR* 5′ region lacking the $FixK_2$-like box on a 168-bp EcoRI fragment	This work
pDB4020	Ap^r (pRJ9519) *nosRZDFYLX* promoter on a 486-bp XbaI-EcoRI fragment	This work
pRJ0004	Km^r [pET-24c(+)] with a 701-bp NdeI/NotI fragment encodingC183S-$FixK_2$-His_6	Bonnet et al., 2013

Analysis of *nosRZDFYLX* Genes Co-transcription by RT-PCR

End-point reverse transcription-polymerase chain reaction (RT-PCR) was performed to investigate the transcriptional architecture of *nosRZDFYLX* genes. First, *B. diazoefficiens* cells were grown under 0.5% initial O_2 concentration to an OD_{600} of ∼0.4 in YEM medium supplemented with 10 mM KNO_3. Cell harvest and isolation of total RNA were done as described previously (Hauser et al., 2007; Lindemann et al., 2007; Mesa et al., 2008). First strand cDNA synthesis was performed with the SuperScript II reverse transcriptase (Invitrogen) according to the supplier's guidelines, using 1 μg of total RNA and primers c2 and g2 that hybridize in the complementary sequence of *nosD* and *nosX* genes. The obtained cDNA was next used for amplification of putative intergenic regions between *nosR* and *nosX* (blr0314-blr0320) using primers' pairs labeled as b1/b2-to-g1/g2 and flanking regions using primers' pair labeled as a1/a2 and h1/h2 (Supplementary Table S1), essentially as described

by Sambrook and Russell (2001). In negative controls, reverse transcriptase was omitted in the reaction. Positive control PCR reactions were performed with *B. diazoefficiens* genomic DNA as template.

5′ RACE of *B. diazoefficiens nosRZDFLYX* Genes

The transcription start sites of *nos* genes were determined with the RACE (Rapid Amplification of cDNA Ends) method as described by Sambrook and Russell (2001). Cell cultivation and harvest as well as total RNA isolation were carried out as described above for the RT-PCR experiments. First strand cDNA synthesis was performed with the SuperScript II reverse transcriptase (Invitrogen) according to the supplier's guidelines, using 0.8 μg of total RNA and primer SP1_nosR. After the reaction, dNTPs and primers were removed with the GeneJET PCR Purification Kit (Thermo Fisher Scientific) and products were eluted in 15 μl of 10 mM Tris-HCl, pH 8.5. Poli-A tails were added to 5′ end of cDNAs with the terminal deoxynucleotidyl transferase (Thermo Fisher Scientific) and final products were diluted with purified water to final volume of 1 ml. Amplification reactions were carried out with primers $(dT)_{17}$-adaptor-primer, adaptor-primer and SP2_ nosR primers using the following PCR program: 95°C for 5 min; (95°C for 30 s; 48°C for 30 s; 72°C for 45 s) × 5 cycles; (95°C for 30 s; 55°C for 30 s; 72°C for 45 s) × 30 cycles; 72°C for 10 min and hold at 4°C. DNA libraries were constructed by cloning the PCR products into pGEM-T easy vector (Promega). Plasmid DNA of individual clones was purified with QIAprep Spin Miniprep Kit (Qiagen) and Sanger sequenced using SP6 as primer. Transcription start sites were identified as the first nucleotide sequenced after the poly-A sequence.

Analysis of *nosRZDFYLX* Gene Expression by qRT-PCR

Expression of *nosR* was also analyzed by qRT-PCR using an iQTM5 Optical System (Bio-Rad, Foster City, CA, United States). *B. diazoefficiens* WT and *napA, nirK, fixK*$_2$, and *nnrR* mutant strains were grown in YEM medium amended with 10 mM NO_3^- under initial 0.5% O_2 (WT, *napA* and *nirK* mutant strains) or 2% O_2 (WT, *fixK*$_2$ and *nnrR* mutant strains) for 24 h. Cell harvest, isolation of total RNA and cDNA synthesis were done as described previously (Hauser et al., 2007; Lindemann et al., 2007; Mesa et al., 2008). Primers for the PCR reactions (nosR_qRT_PCR_F/ nosR_qRT_PCR_R; Supplementary Table S1) were designed with the Clone Manager Suite 9 software to have melting temperatures between 57 and 62°C and generate PCR products of 50–100 bp. Each PCR reaction contained 9.5 μl of iQTM SYBR Green Supermix (Bio-Rad), 2 μM (final concentration) of individual primers and appropriate dilutions of different cDNA samples in a total volume of 19 μl. Reactions were run in triplicate. Melting curves were generated to verify the specificity of the amplification. Relative changes in gene expression were calculated as described by Pfaffl (2001). Expression of the 16S *rrn* gene was used as reference for normalization (primers 16S_qRT_For and 16S_qRT_Rev; Supplementary Table S1).

β-Galactosidase Activity Determination

β-galactosidase activity was determined by using permeabilised cells from at least three independently grown cultures assayed in triplicate essentially as previously described (Cabrera et al., 2016). Specific activities were calculated in Miller units (Miller, 1972).

N_2OR Activity

B. diazoefficiens cells were incubated microoxically (2% O_2) for 24 h in YEM medium supplemented or not with 10 mM NO_3^-. In the latter conditions, parallel replicates were also exposed to 100 μM of the NO-scavenger cPTIO. Next, cells were washed three times with YEM medium and 30 μl gaseous aliquots of 2% N_2O in 98% N_2 (0.15% N_2O final concentration in the headspace) were injected into the rubber stoppered Erlenmeyer flasks. After 5 h of incubation at 30°C at 185 rpm, gas-liquid phase equilibration was reached and 500-μl gaseous aliquots were taken from the headspace to analyze N_2O consumption by gas chromatography as described previously (Tortosa et al., 2015).

The protein concentration was estimated using the Bradford method (Bio-Rad Laboratories) with a standard curve constructed with varying bovine serum albumin (BSA) concentrations. N_2OR activity was determined by using cells from at least three independently biological grown cultures.

Immunoblot Analyses

B. diazoefficiens cells incubated micooxically (2% O_2) in YEM medium in the presence or absence of 10 mM NO_3^- for 24 h, were harvested and the soluble fraction of the cells was obtained by following the protocol previously described by Delgado et al. (2003). The resulting membrane pellet was discarded and the supernatant, containing the soluble fraction, was concentrated to about 100 μl by using AmiconR Ultra-2 centrifugal filter devices (Millipore) and stored at −20°C until their use. Protein concentration was estimated as described above.

For immunodetection of NosZ, protein samples (10 μg of the soluble fraction) were separated by 12% SDS-polyacrylamide gel electrophoresis (PAGE) as described by Laemmli (1970). Then, proteins were transferred to nylon or PVDF membranes (Millipore). The membrane was then incubated in blocking buffer [5% non-fat dry milk in TTBS buffer containing 50 mM Tris-HCl pH 7.5, 0.15 mM NaCl and 0.1% Tween 20], with overnight shaking at 4°C. Afterward, the membrane was then washed with TTBS buffer (four times for 10 min each), before being incubated in 10 ml of blocking buffer containing 1/1000 (v/v) antibody dilution (anti-NosZ of *Pa. denitrificans*; Felgate et al., 2012). The membrane was subsequently incubated by shaking gently for 1 h at room temperature (RT). Further, the membrane was then washed with TTBS and incubated for 1 h at RT with a 1/3500 (v/v) dilution of the secondary antibody (sheep anti-IgG: peroxidase antibody produced in donkeys; A3415 Sigma–Aldrich) in blocking buffer. Next the membrane

was washed four times with TTBS before adding 500 μl of ECL Select western-blotting detection reagent (GE Healthcare, Amersham) followed by Chemiluminescent signal detection in a Chemidoc XRS (Universal Hood II, Bio-Rad). The Quantity One software (Bio-Rad) was used for image analyses.

Purification of *B. diazoefficiens* RNA Polymerase

Purification of the *B. diazoefficiens* holoenzyme was carried by using a modified protocol similar to the one described by Beck et al. (1997). 25 g (wet weight) of *B. diazoefficiens* 110*spc*4 cells grown oxically in PSY supplemented with 0.1% arabinose until late exponential phase were used for each purification batch. All purification steps were performed at 4°C. Cells were resuspended in 70 ml of TGED buffer (10 mM Tris-HCl [pH 8.0], 10% glycerol, 1 mM EDTA, 0.1 mM dithiothreitol [DTT]) containing 0.02 M NaCl and 1 mM ABSF and disrupted in a French pressure cell (three passes at 1000 psi). The crude extract was treated with polyethyleneimine to a final concentration of 0.3%. The pellet obtained after centrifugation (15 min; 27,000 × *g*) was washed with TGED buffer (0.2 M NaCl), and protein containing RNAP was washed three times in TGED buffer (0.8 M NaCl). In all recovery steps, the supernatant was collected and precipitated again by adding solid $(NH_4)_2SO_4$ to 65% final saturation (43 g per 100 ml). The precipitate was collect by centrifugation (30 min; 27,000 × *g*), dissolved in 30 ml of TGED buffer (0.02 M NaCl) and, dialyzed against 1 liter of TGED buffer (0.02 M NaCl). The dialyzed sample was loaded onto an HiTrap Q FF column (GE Healthcare), from which it was eluted by a linear 0.02–1.2 M NaCl gradient. Fractions containing RNAP (as judged by standard transcription assays) were pooled and loaded onto a heparin agarose column (HiTrap Heparin HP; GE Healthcare). Equilibration and elution buffers were similar to those used in the HiTrap Q FF chromatography. Peak fractions contained the RNAP (indicated by general IVT assays performed according Beck et al., 1997) were pooled, concentrated by ultrafiltration (YM30 membrane, Amicon), and dialyzed and stored in TGED buffer (0.02 M NaCl) containing 50% glycerol at −20° or −80°. The purity of the active fractions was tested by SDS-PAGE. Protein concentrations were determined with Bio-Rad assay solution, with BSA as the standard.

IVT Activation Assay

Multiple-round *in vitro* transcription (IVT) assays were carried out as described previously (Beck et al., 1997; Mesa et al., 2008). Plasmid pDB4020 was used as template to study the capacity of the $FixK_2$ protein to initiate transcription from the *nosRZDFYLX* promoter. Expression and purification of an oxidation-insensitive C-terminal Histidine-tagged C183S $FixK_2$ protein variant (C183S-$FixK_2$-His_6; Bonnet et al., 2013) were carried out as described in (Mesa et al., 2005). Purified $FixK_2$ protein was used at concentrations of 1.25 or 2.5 μM dimer.

Runoff transcripts of 286 and 180 nucleotides produced *in vitro* following the procedure used by Mesa et al. (2005) were used as RNA size markers. Transcripts were visualized with a PhosphorImager and signal intensities were determined with the Bio-Rad Quantity One software (Bio-Rad).

RESULTS

Transcriptional Organization of the *B. diazoefficiens nosRZDFYLX* Genes

Analysis of the *nosRZDFYLX* sequence did not reveal any predicted transcriptional termination signals[1] which is an indication that they might be transcribed as an operon. Overlapping coding regions between *nosR* and *nosZ*, as well as between *nosD*, *F*, *Y*, and *L* stop and start codons, suggest translational couplings between *nosRZ* and *nosDFYL*. However, unlike these translational couplings, there is a short intergenic region of 14 nucleotides between *nosZ* and *nosD* and 11 nucleotides between *nosL* and *nosX*.

In order to investigate the transcriptional architecture of *nosRZDFYLX* genes, end-point RT-PCR was performed to detect intergenic regions between each pair of correlative genes. To ensure that the amplified RT-PCR product was from the template mRNA, each RT-PCR reaction had a negative control (without reverse transcriptase) and a positive control (genomic DNA). First, total RNA was isolated from *B. diazoefficiens* WT cells cultured with initial 0.5% O_2 concentration in the presence of NO_3^- and subsequently reverse transcribed to cDNA. As shown in **Figure 1A**, specific cDNA products were obtained for intergenic regions designed as b-to-g, but not from those labeled as "a" and "h" corresponding to flanking regions of the *nosRZDFYLX* genes. These findings reveal that *B. diazoefficiens nosRZDFYLX* genes constitute a transcriptional unit, although we cannot discard the presence of additional internal promoters.

To test any potential transcription from the DNA regions upstream of the *nosR, nosZ,* and *nosD* genes, we determined β-Galactosidase activity of chromosomally integrated transcriptional fusions between the DNA regions preceding the annotated *nosR*, *nosZ*, *nosD* genes and the reporter gene *lacZ* (**Figure 1B**). After growing *B. diazoefficiens* cells under an initial O_2 concentration of 0.5% O_2 in the presence of NO_3^-, the highest transcriptional expression was driven from the *nosR-lacZ* fusion compared to the *nosZ-lacZ* and *nosD-lacZ* fusions (**Figure 1B**). These results strongly suggest that transcription of *nosRZDFYLX* mainly depends on a promoter present in the DNA region upstream of *nosR*. However, although β-galactosidase activity from the *nosZ-lacZ* fusion was sixfold lower to that observed from the *nosR-lacZ* fusion, we cannot exclude the possibility that another internal promoter upstream of *nosZ* might exist.

In order to map transcription initiation within the *nosR* promoter region, we identified their Transcriptional Start Sites (TSS) by using 5′-RACE. As shown in **Figure 2A**, we identify two TSS (TSS_1 and TSS_2) that initiate at a G and T, 84 and 57 bp upstream of the putative translational start codon, respectively. Analysis of the 5′ region of *nosR* revealed the

[1] http://pallab.serc.iisc.ernet.in/gester/dbsearch.php

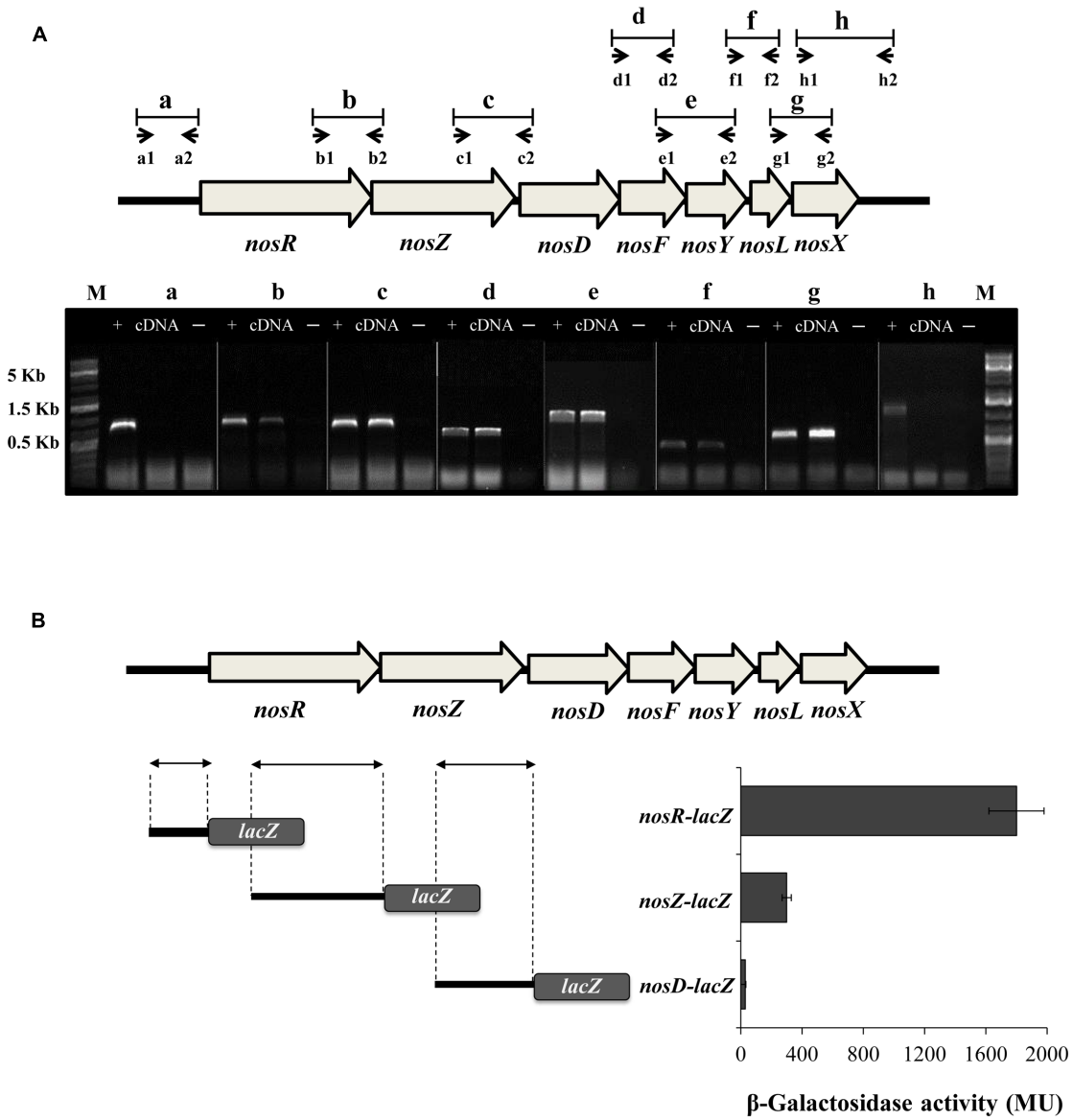


FIGURE 1 | Transcriptional organization of the *B. diazoefficiens nosRZDFYLX* genes. **(A)** Putative intergenic regions probed by RT-PCR are labeled as "b-to-g" and the small arrows below depict positions and orientation of the PCR primers used for each intergenic region. Combinations of primers a1/a2 and h1/h2 were used to amplify "a" and "h" flanking regions. Primer sequences are shown in Supplementary Table S1. RT-PCR products of each primers' pair combination were loaded on an agarose gel. Total RNA served as the template for cDNA synthesis by using gene specific primers that hybridize in the complementary sequence of *nosD* and *nosX* genes. PCR amplifications using genomic DNA as template (+) or without reverse transcriptase (–) served as positive and negative controls, respectively. The sizes of the marker (M) bands are labeled on the left side. **(B)** β-galactosidase activity from the DNA regions upstream of *nosR, nosZ,* and *nosD* genes fused to the *lacZ* reporter gene (on the right). On the left, the DNA regions fused to *lacZ* are depicted by arrows. In **(A,B)**, *B. diazoefficiens* wild-type (WT) cells were grown for 24 h under low oxygen conditions (initial 0.5% O_2) with 10 mM KNO_3. Data expressed as Miller units (MU) represent mean values and error bars from triplicate samples from at least two independent cultures.

presence of a purine-rich Shine-Dalgarno-like sequence (GAGG) four bases in front of the *nosR* putative translational start codon. Exhaustive inspection of the *nosR* promoter region failed to identify any putative conserved -35/-10- or -24/-12-type elements associated to σ^{70}-dependent or σ^{54}-dependent promoters. However, we noticed the presence of an imperfect palindromic sequence (TTGATCCAGCGCAA) positioned at 40.5 and 67.5 bp from TSS_1 and TSS_2, respectively (**Figure 2A**). This sequence resembles reasonably well the consensus sequence of the binding site for FixK-type proteins, 5′-TTGA-N_6-TCAA-3′ (Fischer, 1994; Dufour et al., 2010) and specifically the consensus $FixK_2$ binding site [TTG(A/C)-N_6-(T/G)CAA] recently reported by Bonnet et al. (2013) based on the solved $FixK_2$-DNA complex structure.

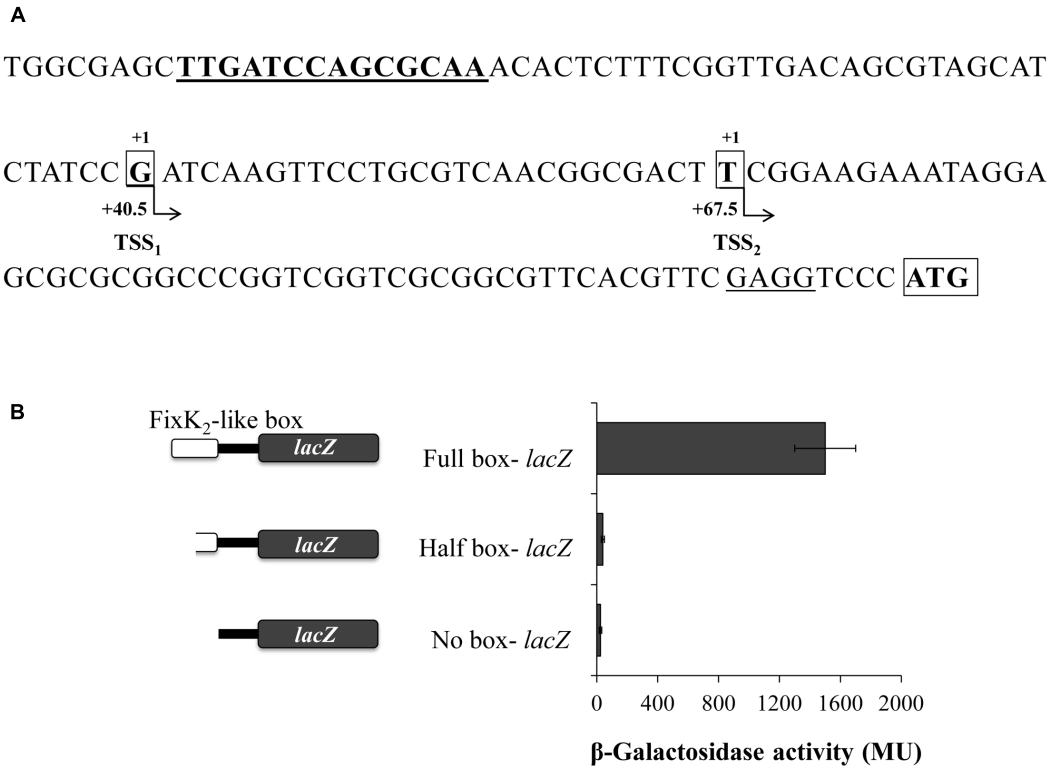


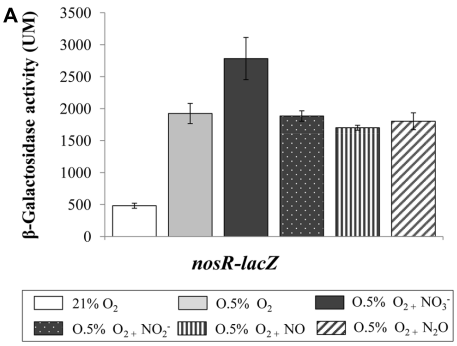
FIGURE 2 | Structure of the *B. diazoefficiens nosR* promoter. **(A)** Sequence and schematic representation of the *nosR* promoter region. Nucleotides corresponding to the transcription start sites (TSS) located at +40.5 and +67.5 bp from the axis of symmetry of the putative $FixK_2$-like box (bold and underlined) identified by RACE are shown in bold, marked with "+1" above and highlighted with an open black box. The annotated *nosR* translation start codon (ATG; http://genome.annotation.jp/RhizoBase) is shown in bold with an open black box. A putative ribosome binding site of *nosR* is underlined. **(B)** β-Galactosidase activity from *nosR-lacZ* fusions containing different lengths of the $FixK_2$-like box is indicated on the left side of the figure. Cells were cultured for 24 h under low oxygen conditions (0.5% O_2) with 10 mM KNO_3. Data expressed as Miller units (MU) represent mean values and error bars from triplicate samples from at least two independent cultures.

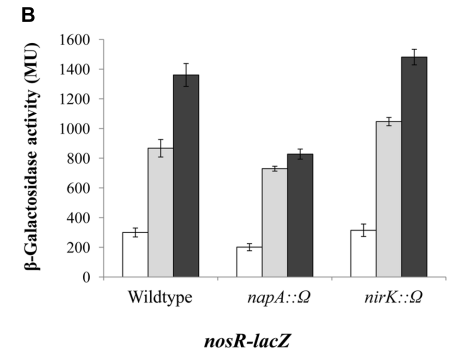
In order to examine the importance of the $FixK_2$-like box identified within the *nosR* promoter region in its transcription, we studied the transcriptional expression derived from a battery of *nosR-lacZ* fusions harboring the full or half $FixK_2$-like box, or a deletion of this box (plasmids pBG0304, pBG0305, and pBG0306, respectively) (**Figure 2B** and **Table 1**). These plasmids were integrated into *B. diazoefficiens* WT and β-galactosidase activity was measured in cells cultured under initial 0.5% O_2 with NO_3^-. In contrast to the significant induction of the *nosR-lacZ* transcriptional fusion containing the full $FixK_2$-like site, expression of *nosR-lacZ* constructs carrying half or deleted $FixK_2$-like site was basal, which showed the importance of the presence of this $FixK_2$-like binding site in the induction of *nosR* (**Figure 2B**).

Low Oxygen Is the Main Signal Which Induces Expression of the *nosRZDFYLX* Operon

To address the effect of low oxygen and NOx in the expression of the *nosRZDFYLX* operon, we analyzed β-galactosidase activity of the *nosR-lacZ* transcriptional fusion in WT cells cultured oxically or under initial 0.5% O_2, both for 24 h, and later exposed to different NOx (NO_3^-, NO_2^-, NO, or N_2O) for additional 5 h-period. As shown in **Figure 3A**, β-galactosidase activity values were basal in cells incubated under oxic conditions. Similar basal levels were observed under oxic conditions in the presence of NO_3^- (data not shown). However, when cells were cultured under 0.5% O_2, expression of the *nosR-lacZ* fusion significantly increased (about fourfold) as compared to oxic conditions (**Figure 3A**). The presence of NO_3^-, but not of NO_2^-, NO, or N_2O, slightly increased *nosR-lacZ* expression (about 1.5-fold) compared to that observed in cells incubated microoxically in the absence of any NOx (**Figure 3A**).

Next, we were interested to confirm that the lack of NO_3^- reduction products does not affect *nosR-lacZ* expression. Therefore, β-galactosidase activity from the *nosR-lacZ* fusion was individually analyzed in *napA* or *nirK* mutant strains which are unable to reduce NO_3^- or NO_2^-, respectively (Velasco et al., 2001; Delgado et al., 2003). Again, a slight induction of the *nosR-lacZ* fusion in the WT cells cultured under 0.5% O_2 in the presence of NO_3^- was observed (**Figure 3B**), however, no change was detected in the *napA* mutant cultured under the same conditions, suggesting a requirement of NO_3^- reduction on *nosR-lacZ* expression. By contrary, induction by NO_3^- of the *nosR-lacZ* fusion was retained in the *nirK* mutant indicating that





C

		Wildtype	*napA::Ω*	*nirK::Ω*
Fold Change	0.5% O_2 + NO_3^-	1	-3.18 ± 0.39	-1.69 ± 0.21

FIGURE 3 | Low-oxygen is the main inducer of *nosRZDFYLX* expression. **(A)** β-Galactosidase activity derived from a *nosR-lacZ* fusion in *B. diazoefficiens* cells grown oxically or under 0.5% O_2 for 24 h. Then, cells were incubated for another 5 h with or without 10 mM KNO_3, 500 μM $NaNO_2$, 50 μM NO, and 30 mM N_2O. **(B)** β-Galactosidase activity from the *nosR-lacZ* fusion in the *B. diazoefficiens* WT, and mutant strains *napA* and *nirK*. Cells were grown oxically (white bars) or under 0.5% O_2 in the absence (gray bars) or in the presence of 10 mM KNO_3 (black bars) during 24 h. **(C)** Expression of *nosR* measured by qRT-PCR. After RNA isolation from cells grown under 0.5% O_2 in the presence of 10 mM KNO_3, qRT-PCR reactions were performed with cDNA synthesized from three independent RNA samples assayed in three parallel reactions. Fold-change values refer to differences of expression in the *napA* and *nirK* mutants relative to the WT. In **(A,B)** data expressed as Miller units (MU) are means with standard error bars from at least two independent cultures, assayed in triplicate.

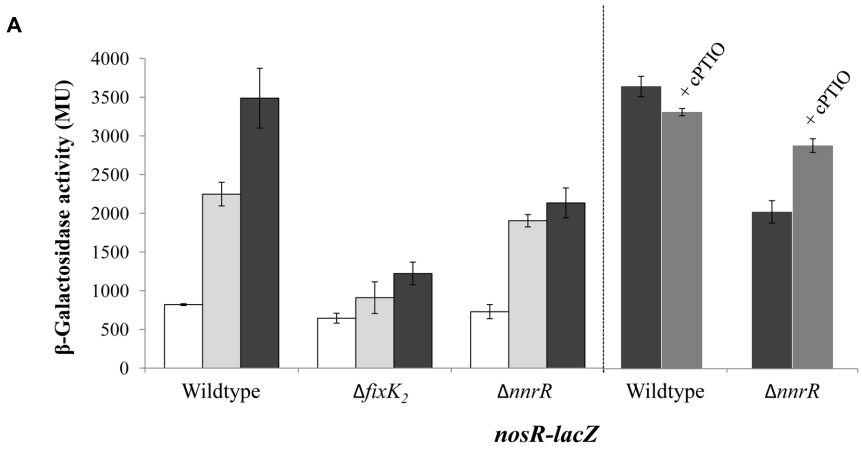
NO_2^- reduction products (NO or N_2O) are not required for activating the expression of *nosRZDFYLX* genes. These results were validated by qRT-PCR analyses (**Figure 3C**). Similarly as we observed by using the *nosR-lacZ* fusion, *nosR* expression was reduced in the *napA* mutant (3.18-fold) compared to WT cells, while it was not significantly affected in the *nirK* mutant (1.69-fold), all cultured in the presence of NO_3^-. However, we cannot conclude that the lack of NO_3^--mediated induction of *nos* genes observed in the *napA* mutant (**Figures 3B,C**) is due to the absence of NO_2^-, since the addition of NO_2^- to the medium did not increase *nosR-lacZ* expression (**Figure 3A**). Taken together, results from **Figures 3A–C** suggest that microoxia is the main signal that induces expression of *B. diazoefficiens nosRZDFYLX* genes.

Selective Regulation of *nosRZDFYLX* Genes by $FixK_2$ But Not by NnrR

In *B. diazoefficiens*, sensing and transduction of the decrease in O_2 concentration are mediated by two interlinked O_2-responsive regulatory cascades, the FixLJ-$FixK_2$-NnrR and the RegSR-NifA (Sciotti et al., 2003). A mild decrease in the O_2 concentration in the gas phase (≤5%) is sufficient to activate expression of FixLJ-$FixK_2$-dependent targets, however, a 10-fold lower O_2 concentration (≤0.5%) is necessary for NifA-mediated activation. In order to investigate how $FixK_2$ and NnrR control the microoxic expression of *nosRZDFYLX* genes, we analyzed β-Galactosidase activity from the *nosR-lacZ* fusion in the WT and Δ*fixK*$_2$ and Δ*nnrR* strains, incubated for 24 h oxically, and microoxically (2% O_2) in the absence or the presence of 10 mM of KNO_3. In these experiments, 2% O_2 concentration was chosen as a middle concentration between 5% (needed for FixLJ-$FixK_2$ cascade activation) and 0.5% (required for the activation of the low O_2-responsive NifA protein), in order to circumvent any possible influence by NifA regulation in our assays.

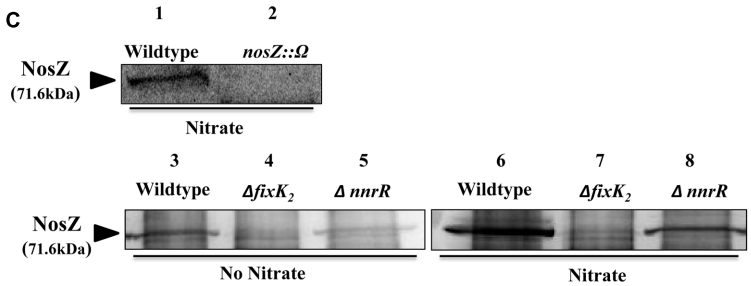
As observed in **Figure 4A**, microoxic induction of *nosR-lacZ* was completely abolished in the absence of a functional *fixK*$_2$ gene, however, it was retained in the Δ*nnrR* strain, suggesting that microoxic expression of *nosRZDFYLX* genes depends on $FixK_2$ but not on NnrR. When cells were cultured microoxically in the presence of NO_3^-, expression of the *nosR-lacZ* fusion was significantly reduced in the *fixK*$_2$ mutant (about threefold) compared to that observed in the WT cells (**Figure 4A**). However, β-galactosidase activity of the *nosR-lacZ* fusion was slightly reduced in the *nnrR* mutant (about 1.75-fold) compared to the WT (**Figure 4A**). This slight reduction of the expression of the *nosR* gene in the *nnrR* mutant is probably due to the toxic effect of NO that is accumulated in *nnrR* cells as previously reported by Bueno et al. (2017). To check this hypothesis, a NO scavenger (cPTIO) was added during growth of WT and Δ*nnrR* cells under microoxic conditions with NO_3^-. As shown in **Figure 4A**, while no effect of cPTIO was observed in WT cells, *nosR-lacZ* expression in Δ*nnrR* cells increased about 40% to that observed in the absence of cPTIO (right panel), which almost corresponds to the expression pattern of the WT. Thus, this indicates that *nos* expression could be partially recovered in the Δ*nnrR* mutant when NO was sequestered by cPTIO.

The different control of *nosR* expression by $FixK_2$ or NnrR was also confirmed by qRT-PCR analyses. When cells were cultured microoxically in the absence of NO_3^-, expression of *nosR* was reduced in the *fixK*$_2$ mutant (3.92-fold) compared to that observed in the WT cells (**Figure 4B**), however, it was almost not affected in the *nnrR* mutant (**Figure 4B**). When NO_3^- was added to medium, a significant reduction of *nosR* expression (10.38-fold) was observed in the *fixK*$_2$ mutant but only a slight decrease (2.4-fold) was detected in the *nnrR* mutant, both compared to the WT cultured in the same conditions (**Figure 4B**). Taken together,



B

		Wildtype	Δ*fixK*$_2$	Δ*nnrR*
Fold Change	2% O_2	1	- 3.92 ± 0.03	- 1.09 ± 0.04
	2% O_2 + NO_3^-	1	-10.38 ± 2.60	-2.40 ± 0.38



D

N_2OR activity

No Nitrate			Nitrate			Nitrate + cPTIO		
Wildtype	Δ*fixK*$_2$	Δ *nnrR*	Wildtype	Δ*fixK*$_2$	Δ*nnrR*	Wildtype	Δ*fixK*$_2$	Δ*nnrR*
6.23 ± 0.41	0.29 ± 0.11	3.54 ± 0.08	9.77 ± 0.57	0.19 ± 0.10	0.95 ± 0.09	12.24 ± 2.79	0.78 ± 0.20	11.78 ± 2.36

FIGURE 4 | Control of *nosRZDFYLX* expression by the regulatory proteins FixK$_2$ and NnrR. **(A)** β-Galactosidase activity expressed as Miller units (MU) from the *nosR-lacZ* transcriptional fusion chromosomally integrated in the *B. diazoefficiens* WT strain, and Δ*nnrR*, and Δ*fixK*$_2$ strains grown oxically (white bars), under 2% O_2 in the absence (light gray bars) or in the presence of 10 mM KNO_3 (black bars) for 24 h. In the right panel, 10 μM of the NO-scavenger cPTIO was added to a series of cultures containing NO_3^- (dark gray bars). **(B)** Expression of *nosR* by qRT-PCR in the WT, and Δ*nnrR*, and Δ*fixK*$_2$ strains. qRT-PCR reactions were performed with cDNA synthesized from three independent RNA samples assayed in triplicate. Fold-change values refer to differences of expression in the Δ*nnrR*, and Δ*fixK*$_2$ mutants relative to the WT. **(C)** Western-blotted SDS-PAGE gels of the soluble fraction from the WT and Δ*nnrR*, and Δ*fixK*$_2$ strains probed with anti-NosZ antibody from *Pa. denitrificans*. As control, a *B. diazoefficens nosZ* mutant was used. The size of *B. diazoefficiens* NosZ is labeled on the left side. **(D)** Nitrous oxide reductase (N_2OR) activity in the WT and Δ*nnrR*, and Δ*fixK*$_2$ strains expressed as nmol N_2O consumed × (mg prot^{-1}) h^{-1}. In **(B–D)**, cells were grown under 2% O_2 in the absence or in the presence of 10 mM KNO_3 during 24 h. 100 μM of cPTIO was added to some of the cultures containing NO_3^- in **(D)**. In **(A,B,D)**, data shown as means with standard errors from at least two independent cultures, assayed in triplicate.

these results suggest FixK$_2$ as the transcriptional activator of *nos* genes in response to microoxic conditions.

The differential dependency of *nosRZDFYLX* expression on FixK$_2$ and NnrR was also confirmed at protein level by immunoblot analyses using antibodies raised against purified *Pa. denitrificans* NosZ (Felgate et al., 2012). Firstly, we were able to identify NosZ protein in the soluble fraction of *B. diazoefficiens* cells cultured under microoxic conditions (2% O_2) with NO_3^-, since a prominent band of about 70 kDa found in the WT was readily undetectable in the *nosZ* mutant (**Figure 4C**, lanes 1 and 2). The size of this band corresponds to the predicted molecular mass of *B. diazoefficiens* NosZ subunit

(71.6 kDa; ProtParam tool[2]). NosZ was already detected in the WT cells cultured microoxically (**Figure 4C**, lane 3) but the presence of NO_3^- slightly increased NosZ steady-state levels (**Figure 4C**, lane 6). This is in line with the observed NO_3^--mediated induction of the *nosR-lacZ* fusion (**Figures 3A,B**, **4A**). Similarly as the expression pattern observed for the *nosR-lacZ* fusion, NosZ was present in the soluble fraction of Δ*nnrR* cells cultured microoxically either in the absence or in the presence of nitrate (**Figure 4C**, lanes 5 and 8), although at a slightly lower concentration than in the WT cells. As expected, the band of about 70 kDa corresponding to NosZ was absent in the soluble fractions of the Δ*fixK*$_2$, independently of the presence or absence of NO_3^- in the incubation medium (**Figure 4C**, lanes 4 and 7).

Finally, we determined N_2O reductase (N_2OR) activity in *B. diazoefficiens* WT and *fixK*$_2$ and *nnrR* mutant strains as the capacity to reduce a defined initial N_2O concentration. As shown in **Figure 4D**, values of N_2OR activity in WT cells correlated with NosZ steady-state levels in *B. diazoefficiens* cells (**Figure 4C**), where a slight induction (about 1.6-fold) of activity was observed in the WT cells in the presence of NO_3^- (**Figure 4D**) compared to that observed in exclusively microoxic conditions. In line with the expression pattern of the *nosR-lacZ* fusion (**Figure 4A**), *nosR* expression (**Figure 4B**) and NosZ detection (**Figure 4C**, lanes 4 and 7), N_2OR activity was severely impaired in the Δ*fixK*$_2$ strain cultivated microoxically independently of the presence of NO_3^- (**Figure 4D**). Under microoxic conditions, cells of the Δ*nnrR* strain showed a milder decrease of N_2OR activity (about 1.75-fold) compared to that observed in WT cells (**Figure 4D**), which was significantly diminished further (about 10-fold) in the presence of NO_3^- (**Figure 4D**). As we have mentioned above, this strong decrease is probably due to the higher NO accumulation capacity of Δ*nnrR* cells grown microoxically with nitrate compared to WT cells grown under the same conditions (Bueno et al., 2017). In fact, when cPTIO was added during growth, Δ*nnrR* cells restored its ability to reduce N_2O reaching WT N_2OR activity values (**Figure 4D**). These data discard the involvement of NnrR as direct regulator of *nos* expression and suggest that the incapacity of Δ*nnrR* to reduce N_2O under microoxic conditions with NO_3^- is probably due to the accumulation of NO. Taken together, these results pointed out that $FixK_2$ is the key transcriptional regulator involved in *nosRZDFYLX* expression.

The *nosRZDFYLX* Operon Is a Novel Direct Target of $FixK_2$

In order to investigate whether $FixK_2$ could have a direct role on *nosRZDFYLX* activation, we monitored RNA synthesis by multiple-round IVT. The *nosR* promoter region was cloned into the template plasmid pRJ9519 (Beck et al., 1997), which carries an *rrn* terminator, yielding plasmid pDB4020. In these experiments, purified C183S-$FixK_2$-His_6 (Bonnet et al., 2013), hereafter referred as $FixK_2$, and RNA polymerase (RNAP) holoenzyme from *B. diazoefficiens* that was purified in this work (see Material and Methods) were used. In the absence of $FixK_2$, *B. diazoefficiens* RNAP was unable to transcribe the *nosR* promoter efficiently (**Figure 5**, lane 3), whereas it produced a vector-encoded transcript that served as an internal reference. In the presence of $FixK_2$ (1.25 and 2.5 μM dimer), *B. diazoefficiens* RNAP transcribed the *nosRZDFYLX* promoter producing a single specific transcript larger than 286 nucleotides (**Figure 5**, lanes 4 and 5, respectively), which probably initiate at TSS_1. This suggested that the *nosR* promoter is directly activated by $FixK_2$ and that transcription from TSS_1 depends on $FixK_2$, at least, in *in vitro* conditions.

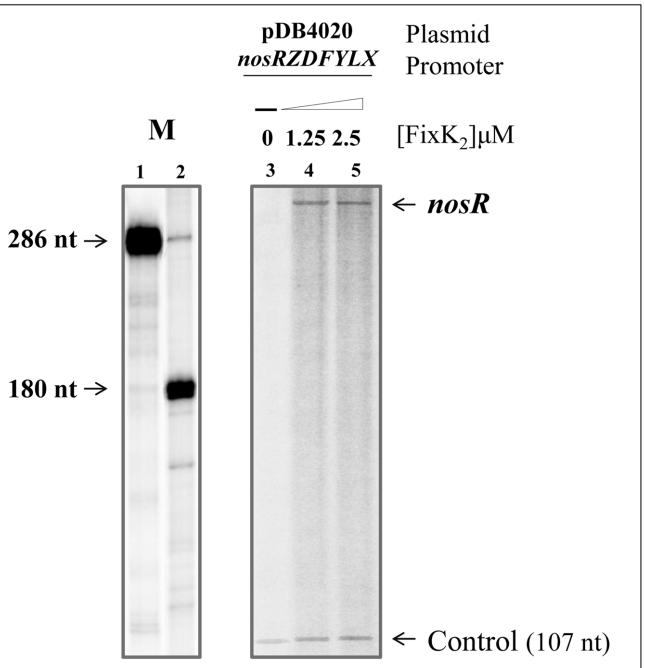


FIGURE 5 | $FixK_2$-mediated IVT activation from the *nosRZDFYLX* promoter. pDB4020 plasmid containing the *nosR* promoter cloned upstream of the *rrn* terminator was used for multiple-round IVT assays with different amounts of purified $FixK_2$ protein and RNAP from *B. diazoefficiens*. $FixK_2$ concentrations were as follows: no protein, 1.25 and 2.5 μM (lanes 3–5). Transcripts synthesized *in vitro* in the presence of [α − ^{32}P]UTP were separated on a 6% denaturing polyacrylamide gel and visualized by phosphorimager analysis of the dried gel. Markers transcripts (M) of 286 and 180 nucleotides loaded in lanes 1 and 2 were produced as described by Mesa et al. (2005). The 107-nucleotide transcript present in all lanes originates from a promoter located on the plasmid vector and serves as internal reference. Shown are the results from a transcription experiment that was repeated at least once. Both panels correspond to the same gel. nt, nucleotides.

DISCUSSION

Given the damaging effect on climate change of the powerful GHG N_2O, strategies to mitigate their emissions have to be developed in order to increase agricultural efficiency and decrease current levels of N_2O production, to satisfy the demands of continuing population growth (Richardson et al., 2009; Thomson et al., 2012). These strategies should include a better understanding of the environmental and molecular factors that contribute to the biological generation and consumption of N_2O.

[2]http://web.expasy.org/protparam/

B. diazoefficiens, the endosymbiont of soybeans, contributes to N_2O emissions given its capacity to carry out the denitrification process under both free-living and symbiotic conditions. Despite the significant knowledge available in this rhizobial species on the regulation of the three first enzymes of denitrification (Nap, NirK, and cNor) involved in N_2O production (Bueno et al., 2017), the regulatory mechanisms involved in the control of the key step in N_2O mitigation (the reduction of N_2O to N_2) in response to low oxygen and NOx has not been covered in detail. Previous studies have demonstrated that expression of a *nosZ-lacZ* fusion depends on low O_2, the presence of NO_3^- and the FixLJ, $FixK_2$ and NosR regulatory proteins (Velasco et al., 2004). The capacity of *B. diazoefficiens* to couple N_2O reduction to growth as well as a role for the NasST regulatory system on modulation of *nosZ* gene transcription has also been reported (Sánchez et al., 2013, 2014). Furthermore, recent studies have demonstrated the capacity of NasT to interact with *B. diazoefficiens nosR* 5′-leader RNA (Sánchez et al., 2017).

In this work, we have dissected, for the first time, the transcriptional organization of the *nosRZDFYLX* genes in *B. diazoefficiens*. By using RT-PCR we found that the *nosRZDFYLX* genes are transcribed as a single polycistronic mRNA and thus, they are organized as an operon. The transcriptional arrangement of the *nos* genes in other denitrifiers indicate the existence of a diversity of transcriptionally active promoters detected across the *nos* genes between different bacterial species (Zumft and Kroneck, 2007). Supporting our findings, the *Ps. aeruginosa nos* genes are arranged in a single hexacistronic *nosRZDFYL* operon (Arai et al., 2013). A single *nosZ* transcript was identified in *Ps. fluorescens* as well (Philippot et al., 2001). However, in *Ps. stutzeri* three units of monocistronic *nosR* and *nosZ*, and the *nosDFYLtatE* operon (Cuypers et al., 1992; Vollack and Zumft, 2001; Honisch and Zumft, 2003) have been proposed. Similarly, the transcriptional organization of the *nos* cluster of both *Ensifer meliloti* and *Pa. denitrificans* comprises three transcripts: *nosR*, *nosZ,* and *nosDF*(*Y*), and *nosCR*, *nosZ,* and *nosDFYLX*, respectively (Holloway et al., 1996; van Spanning, 2011). In order to confirm the results obtained by RT-PCR, we looked for transcriptionally active promoters within *B. diazoefficiens nosRZDFYLX* operon analyzing the transcriptional strength driven by the DNA regions upstream to the *nosR*, *nosZ,* and *nosD* genes. Interestingly, the highest transcriptional activity was derived from the DNA region upstream of the *nosR* gene compared to that detected from the *nosZ* gene, and no transcription was observed from the 5′ DNA region of the *nosD* gene. The presence of a transcriptionally active promoter upstream of the *nosZ* gene was previously demonstrated by using a *nosZ-lacZ* transcriptional fusion (Velasco et al., 2004) and by performing 5′-RACE (Sánchez et al., 2017). However, since a binding motif for FixK-type regulators was only present within the promoter region of *nosR*, we suggest that this promoter plays the major role in *B. diazoefficiens nosRZDFYLX* regulation.

In this work, we have identified two *nosR* TSS, i.e., TSS_1 and TSS_2, positioned at +40.5 and +67.5 bp, respectively, from the axis of symmetry of the FixK-like binding site (TTGATCCAGCGCAA). Similarly, a TSS at +40.5 from the axis of symmetry of the FixK box has been recently identified by Sánchez et al. (2017). In contrast to our results, the TSS at +67.5 bp was not identified in the latter studies. This discrepancy could be due to the different growth conditions used by Sánchez et al. (2017) where cells were cultured in HMM medium (Sameshima-Saito et al., 2006) under anoxic conditions (replacement of O_2 by N_2 in the gas phase). $FixK_2$-like boxes are present within the promoters of the *B. diazoefficiens napEDABC* (TTGATCCAGATCAA), *nirK* (TTGTTGCAGCGCAA), and *norCBDQD* (TTGCGCCCTGACAA) genes (Velasco et al., 2001; Mesa et al., 2002; Delgado et al., 2003; Supplementary Figure S1). Interestingly, only the *napEDABC*-associated $FixK_2$ box as well as the *nosR*-box identified in this work, matches quite well with the consensus $FixK_2$ box, TTG(A/C)-N_6-(T/G)CAA (Mesa et al., 2008, 2009; Bonnet et al., 2013; Supplementary Figure S1). Deletion of this $FixK_2$-like box resulted in the complete shutdown of *nosR-lacZ* expression, indicating its essential role in the transcription of the *nosRZDFYLX* operon.

Cells of *B. diazoefficiens* grown oxically showed a basal expression of the *nosR-lacZ* fusion. In this regard, previous observations showed that the *Ps. stutzeri nosZ* gene can also be expressed at high O_2 concentrations (Miyahara et al., 2010). Supporting these findings, it was recently demonstrated the capacity of both *Ps. stutzeri* and *Pa. denitrificans* to reduce N_2O under oxic conditions (Desloover et al., 2014; Qu et al., 2015).

Similarly as described for *napEDABC* genes (Bueno et al., 2017), we found that microoxia is sufficient to induce expression of the *nosR-lacZ* fusion, NosZ levels as well as N_2OR activity. In contrast to that observed for *nosR*/NosZ expression and activity, previous results reported that microoxic expression of *B. diazoefficiens norCBQD* genes required the presence of either NO_3^-, NO_2^-, or NO, the latter being the signal molecule involved in such control (Bueno et al., 2017). The slight induction of the *nosR-lacZ* fusion in WT cells cultured in the presence of NO_3^- was not observed in cells of a *napA* mutant which does not reduce NO_3^-. However, results from **Figure 3A** suggest that any of the NOx derived from NO_3^- reduction (NO_2^-, NO, or N_2O) are not inducers of *nosR-lacZ* expression. Furthermore, NO is not required for *nosR-lacZ* induction, since WT levels of *nosR* expression were observed in a *nirK* mutant which does not reduce NO_2^- to NO. Likewise as we found in this work, previous studies suggested N_2O as a weak inducer of *nosZ* genes in several bacteria (Kroneck et al., 1989; Richardson et al., 1991; Sabaty et al., 1999). Taken together, these observations suggest a very mild effect of NOx in the expression of *nos* genes. Therefore, it might be possible that a change in the cellular redox state derived from NO_3^- reduction by Nap is involved in *nosR-lacZ* induction. In fact, our own previous results demonstrated the involvement of the *B. diazoefficiens* redox-responsive regulatory protein RegR on the expression of *nos* genes (Torres et al., 2014). Alternatively, the NasST system might be involved in the NO_3^--mediated response of *nos* genes expression (Sánchez et al., 2014).

Microoxic induction of the *nosRZDFYLX* genes as well as NosZ expression in *B. diazoefficiens* depends on $FixK_2$, but not on NnrR. The dependency of *nosRZDFYLX* transcription on $FixK_2$ was demonstrated by IVT transcription experiments

carried out with oxically purified protein in collaboration with *B. diazoefficiens* RNAP. In the same manner, microoxic induction of the *B. diazoefficiens napEDABC* genes depends on $FixK_2$, but not on NnrR, probably due to its NOx-independent expression (Bueno et al., 2017). In fact, $FixK_2$ also activates transcription of *napEDABC* genes (Bueno et al., 2017).

In contrast to our results, NO has been proposed as the signal that upregulates the *nosR*, *nosZ,* and *nosD* promoters in *Ps. aeruginosa, Ps. stutzeri*, and *Pa. denitrificans* (reviewed by Zumft and Kroneck, 2007). In *Rhodobacter sphaeroides* IL106 *nosZ* expression depends on one of the reduction products of NO_3^-, suggesting NO as the signal molecule, too (Sabaty et al., 1999). Further, global gene expression analysis carried out with *E. meliloti* showed induction of *nos* genes in response to NO (Meilhoc et al., 2010). NO-dependent induction of *nos* genes in *Ps. aeruginosa*, *Ps. Stutzeri,* or *Pa. denitrificans* is processed via the regulatory proteins DNR/DnrD/NNR, respectively (van Spanning et al., 1999; Vollack and Zumft, 2001; Arai et al., 2013). While *Ps. aeruginosa* DNR is under the control of the low O_2-sensing protein ANR (Trunk et al., 2010), transcription of *dnrD* in *Ps. stutzeri* is activated in cells grown under O_2 limitation conditions, being particularly strong in denitrifying cells, but not under the control of the low-O_2 sensor FnrA (Vollack et al., 1999). A particular case constitutes *Pa. denitrificans*, where N_2O reduction is subjected to a robust regulation by FnrP and NNR in response to low oxygen (via FnrP) or NO (via NNR) (Bergaust et al., 2012).

The reduced induction of *nosR*/NosZ expression observed in Δ*nnrR* cells cultured with NO_3^- that has also been described previously for *napEDABC* genes expression (Bueno et al., 2017), might be a consequence of the higher capacity to accumulate NO by the *nnrR* mutant strain compared to the WT strain (Bueno et al., 2017). Supporting this hypothesis, when NO was removed by adding the NO-scavenger cPTIO to the Δ*nnrR* cultures with NO_3^-, *nosR-lacZ* expression as well as N_2OR activity restored to WT levels. It might be possible that the NosZ catalytic center Cu_z which remains in a redox-inert, paramagnetic state Cu_z^* (Wunsch and Zumft, 2005), is inactivated in the presence of NO accumulated by the *nnrR* mutant (Dell'Acqua et al., 2011). However, the precise mechanism involved in NosZ inactivation by NO is still unknown.

This work performed with the model rhizobial denitrifier *B. diazoefficiens* expands the understanding of the environmental and regulatory factors involved in the reduction of N_2O, the key step that mitigates N_2O emissions. We hope that our results would help to establish action plans for the development of practical strategies for mitigation of N_2O emissions from legume crops.

AUTHOR CONTRIBUTIONS

MT, EB, MD, and SM conceived and designed the study. MT, EB, AJ-L, and JC performed the experiments. MT, EB, AJ-L, JC, MD, and SM analyzed the results. MT, EB, MD, and SM wrote the manuscript. EB critically revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by Fondo Europeo de Desarrollo Regional (FEDER)-co-financed grants (AGL2013-45087-R and AGL2015-63651-P) from the Ministerio de Economía y Competitividad (Spain). Grant P12-AGR-1968 and support from the Junta de Andalucía to Group BIO-275 are also acknowledged. MT was supported by a contract funded by Grant AGL2013-45087-R. AJ-L was financed by a Ph.D. contract associated to Grant P12-AGR-1968. EB was supported by the Consejo Superior de Investigaciones Cientificas JAE-DOC Programme co-financed by European Social Fund (ESF).

ACKNOWLEDGMENTS

We are grateful to Germán Tortosa (EEZ, CSIC, Granada, Spain) for the excellent technical assistance. Juan J. Lázaro and Alfonso Lázaro (EEZ, CSIC, Granada, Spain) are acknowledged for their help in *B. diazoefficiens* RNAP purification. We also thank D. Richardson (UEA, Norwich, United Kingdom) for the gift of the *Pa. denitrificans* NosZ polyclonal antibodies. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01621/full#supplementary-material

REFERENCES

Arai, H., Roh, J. H., Eraso, J. M., and Kaplan, S. (2013). Transcriptome response to nitrosative stress in *Rhodobacter sphaeroides* 2.4.1. *Biosci. Biotechnol. Biochem.* 77, 111–118. doi: 10.1271/bbb.120601

Beck, C., Marty, R., Klausli, S., Hennecke, H., and Gottfert, M. (1997). Dissection of the transcription machinery for housekeeping genes of *Bradyrhizobium japonicum*. *J. Bacteriol.* 179, 364–369. doi: 10.1128/jb.179.2.364-369.1997

Bedmar, E. J., Robles, E. F., and Delgado, M. J. (2005). The complete denitrification pathway of the symbiotic, nitrogen-fixing bacterium *Bradyrhizobium japonicum*. *Biochem. Soc. Trans.* 33, 141–144. doi: 10.1042/BST0330141

Bergaust, L., van Spanning, R. J., Frostegard, A., and Bakken, L. R. (2012). Expression of nitrous oxide reductase in *Paracoccus denitrificans* is regulated by oxygen and nitric oxide through FnrP and NNR. *Microbiology* 158, 826–834. doi: 10.1099/mic.0.054148-0

Bonnet, M., Kurz, M., Mesa, S., Briand, C., Hennecke, H., and Grutter, M. G. (2013). The structure of *Bradyrhizobium japonicum* transcription factor FixK2 unveils sites of DNA binding and oxidation. *J. Biol. Chem.* 288, 14238–14246. doi: 10.1074/jbc.M113.465484

Browning, D. F., and Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2, 57–65. doi: 10.1038/nrmicro787

Bueno, E., Mesa, S., Bedmar, E. J., Richardson, D. J., and Delgado, M. J. (2012). Bacterial adaptation of respiration from oxic to microoxic and anoxic conditions: redox control. *Antioxid. Redox. Signal.* 16, 819–852. doi: 10.1089/ars.2011.4051

Bueno, E., Mesa, S., Sánchez, C., Bedmar, E. J., and Delgado, M. J. (2010). NifA is required for maximal expression of denitrification genes in *Bradyrhizobium japonicum*. *Environ. Microbiol.* 12, 393–400. doi: 10.1111/j.1462-2920.2009.02076.x

Bueno, E., Robles, E. F., Torres, M. J., Krell, T., Bedmar, E. J., Delgado, M. J., et al. (2017). Disparate response to microoxia and nitrogen oxides of the *Bradyrhizobium japonicum napEDABC*, *nirK* and *norCBQD* denitrification genes. *Nitric Oxide* 68, 137–149. doi: 10.1016/j.niox.2017.02.002

Butterbach-Bahl, K., Baggs, E. M., Dannenmann, M., Kiese, R., and Zechmeister-Boltenstern, S. (2014). Nitrous oxide emissions from soils: How well do we understand the processes and their controls? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20130122. doi: 10.1098/rstb.2013.0122

Cabrera, J. J., Salas, A., Torres, M. J., Bedmar, E. J., Richardson, D. J., Gates, A. J., et al. (2016). An integrated biochemical system for nitrate assimilation and nitric oxide detoxification in *Bradyrhizobium japonicum*. *Biochem. J.* 473, 297–309. doi: 10.1042/bj20150880

Cuypers, H., Viebrock-Sambale, A., and Zumft, W. G. (1992). NosR, a membrane-bound regulatory component necessary for expression of nitrous oxide reductase in denitrifying *Pseudomonas stutzeri*. *J. Bacteriol.* 174, 5332–5339. doi: 10.1128/jb.174.16.5332-5339.1992

Daniel, R. M., and Appleby, C. A. (1972). Anaerobic-nitrate, symbiotic and aerobic growth of *Rhizobium japonicum*: effects on cytochrome P450, other haemoproteins, nitrate and nitrite reductases. *Biochim. Biophys. Acta* 275, 347–354. doi: 10.1016/0005-2728(72)90215-0

Delamuta, J. R., Ribeiro, R. A., Ormeno-Orrillo, E., Melo, I. S., Martinez-Romero, E., and Hungria, M. (2013). Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. *Int. J. Syst. Evol. Microbiol.* 63, 3342–3351. doi: 10.1099/ijs.0.049130-0

Delgado, M. J., Bonnard, N., Tresierra-Ayala, A., Bedmar, E. J., and Muller, P. (2003). The *Bradyrhizobium japonicum* napEDABC genes encoding the periplasmic nitrate reductase are essential for nitrate respiration. *Microbiology* 149, 3395–3403. doi: 10.1099/mic.0.26620-0

Dell'Acqua, S., Pauleta, S. R., Moura, I., and Moura, J. J. (2011). The tetranuclear copper active site of nitrous oxide reductase: the CuZ center. *J. Biol. Inorg. Chem.* 16, 183–194. doi: 10.1007/s00775-011-0753-3

Desloover, J., Roobroeck, D., Heylen, K., Puig, S., Boeckx, P., Verstraete, W., et al. (2014). Pathway of nitrous oxide consumption in isolated *Pseudomonas stutzeri* strains under anoxic and oxic conditions. *Environ. Microbiol.* 16, 3143–3152. doi: 10.1111/1462-2920.12404

Dufour, Y. S., Kiley, P. J., and Donohue, T. J. (2010). Reconstruction of the core and extended regulons of global transcription factors. *PLoS Genet.* 6:e1001027. doi: 10.1371/journal.pgen.1001027

Felgate, H., Giannopoulos, G., Sullivan, M. J., Gates, A. J., Clarke, T. A., Baggs, E., et al. (2012). The impact of copper, nitrate and carbon status on the emission of nitrous oxide by two species of bacteria with biochemically distinct denitrification pathways. *Environ. Microbiol.* 14, 1788–1800. doi: 10.1111/j.1462-2920.2012.02789.x

Fischer, H. M. (1994). Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* 58, 352–386.

Hauser, F., Pessi, G., Friberg, M., Weber, C., Rusca, N., Lindemann, A., et al. (2007). Dissection of the *Bradyrhizobium japonicum* NifA+σ54 regulon, and identification of a ferredoxin gene (*fdxN*) for symbiotic nitrogen fixation. *Mol. Genet. Genomics* 278, 255–271. doi: 10.1007/s00438-007-0246-9

Holloway, P., McCormick, W., Watson, R. J., and Chan, Y. K. (1996). Identification and analysis of the dissimilatory nitrous oxide reduction genes, *nosRZDFY*, of *Rhizobium meliloti*. *J. Bacteriol.* 178, 1505–1514. doi: 10.1128/jb.178.6.1505-1514.1996

Honisch, U., and Zumft, W. G. (2003). Operon structure and regulation of the nos gene region of *Pseudomonas stutzeri*, encoding an ABC-Type ATPase for maturation of nitrous oxide reductase. *J. Bacteriol.* 185, 1895–1902. doi: 10.1128/JB.185.6.1895-1902.2003

Intergovernmental Panel on Climate Change [IPCC] (2014). *Climate change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: IPCC. doi: 10.1017/CBO9781107415416

Jones, C. M., Graf, D. R., Bru, D., Philippot, L., and Hallin, S. (2013). The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. *ISME J.* 7, 417–426. doi: 10.1038/ismej.2012.125

Korner, H., Sofia, H. J., and Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol. Rev.* 27, 559–592. doi: 10.1016/S0168-6445(03)00066-4

Kraft, B., Strous, M., and Tegetmeyer, H. E. (2011). Microbial nitrate respiration–genes, enzymes and environmental distribution. *J. Biotechnol.* 155, 104–117. doi: 10.1016/j.jbiotec.2010.12.025

Kroneck, P. M., Antholine, W. A., Riester, J., and Zumft, W. G. (1989). The nature of the cupric site in nitrous oxide reductase and of Cu_A in cytochrome *c* oxidase. *FEBS Lett.* 248, 212–213. doi: 10.1016/0014-5793(89)80464-8

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0

Lindemann, A., Moser, A., Pessi, G., Hauser, F., Friberg, M., Hennecke, H., et al. (2007). New target genes controlled by the *Bradyrhizobium japonicum* two-component regulatory system RegSR. *J. Bacteriol.* 189, 8928–8943. doi: 10.1128/JB.01088-07

Meilhoc, E., Cam, Y., Skapski, A., and Bruand, C. (2010). The response to nitric oxide of the nitrogen-fixing symbiont *Sinorhizobium meliloti*. *Mol. Plant Microbe Interact.* 23, 748–759. doi: 10.1094/MPMI-23-6-0748

Mesa, S., Bedmar, E. J., Chanfon, A., Hennecke, H., and Fischer, H. M. (2003). *Bradyrhizobium japonicum* NnrR, a denitrification regulator, expands the FixLJ-FixK2 regulatory cascade. *J. Bacteriol.* 185, 3978–3982. doi: 10.1128/JB.185.13.3978-3982.2003

Mesa, S., Hauser, F., Friberg, M., Malaguti, E., Fischer, H. M., and Hennecke, H. (2008). Comprehensive assessment of the regulons controlled by the FixLJ-FixK2-FixK1 cascade in *Bradyrhizobium japonicum*. *J. Bacteriol.* 190, 6568–6579. doi: 10.1128/JB.00748-08

Mesa, S., Hennecke, H., and Fischer, H. M. (2006). A multitude of CRP/FNR-like transcription proteins in *Bradyrhizobium japonicum*. *Biochem. Soc. Trans.* 34, 156–159. doi: 10.1042/BST0340156

Mesa, S., Reutimann, L., Fischer, H. M., and Hennecke, H. (2009). Posttranslational control of transcription factor FixK2, a key regulator for the *Bradyrhizobium japonicum*-soybean symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21860–21865. doi: 10.1073/pnas.0908097106

Mesa, S., Ucurum, Z., Hennecke, H., and Fischer, H. M. (2005). Transcription activation in vitro by the *Bradyrhizobium japonicum* regulatory protein FixK2. *J. Bacteriol.* 187, 3329–3338. doi: 10.1128/JB.187.10.3329-3338.2005

Mesa, S., Velasco, L., Manzanera, M. E., Delgado, M. J., and Bedmar, E. J. (2002). Characterization of the *norCBQD* genes, encoding nitric oxide reductase, in the nitrogen fixing bacterium *Bradyrhizobium japonicum*. *Microbiology* 148, 3553–3560. doi: 10.1099/00221287-148-11-3553

Miller, J. H. (1972). *Experiments in Molecular Genetics*. New York, NY: Cold Spring Harbor Laboratory.

Miyahara, M., Kim, S. W., Fushinobu, S., Takaki, K., Yamada, T., Watanabe, A., et al. (2010). Potential of aerobic denitrification by *Pseudomonas stutzeri* TR2 to reduce nitrous oxide emissions from wastewater treatment plants. *Appl. Environ. Microbiol.* 76, 4619–4625. doi: 10.1128/AEM.01983-09

Nellen-Anthamatten, D., Rossi, P., Preisig, O., Kullik, I., Babst, M., Fischer, H. M., et al. (1998). *Bradyrhizobium japonicum* FixK2, a crucial distributor in the FixLJ-dependent regulatory cascade for control of genes inducible by low oxygen levels. *J. Bacteriol.* 180, 5251–5255.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45. doi: 10.1093/nar/29.9.e45

Philippot, L., Mirleau, P., Mazurier, S., Siblot, S., Hartmann, A., Lemanceau, P., et al. (2001). Characterization and transcriptional analysis of *Pseudomonas fluorescens* denitrifying clusters containing the *nar*, *nir*, nor and *nos* genes. *Biochim. Biophys. Acta* 1517, 436–440. doi: 10.1016/S0167-4781(00)00286-4

Prendergast-Miller, M. T., Baggs, E. M., and Johnson, D. (2011). Nitrous oxide production by the ectomycorrhizal fungi *Paxillus involutus* and *Tylospora fibrillosa*. *FEMS Microbiol. Lett.* 316, 31–35. doi: 10.1111/j.1574-6968.2010.02187.x

Qu, Z., Bakken, L. R., Molstad, L., Frostegard, A., and Bergaust, L. (2015). Transcriptional and metabolic regulation of denitrification in *Paracoccus denitrificans* allows low but significant activity of nitrous oxide reductase under oxic conditions. *Environ. Microbiol.* 18, 2951–2963. doi: 10.1111/1462-2920.13128

Rasmussen, T., Berks, B. C., Sanders-Loehr, J., Dooley, D. M., Zumft, W. G., and Thomson, A. J. (2000). The catalytic center in nitrous oxide reductase, CuZ, is a copper-sulfide cluster. *Biochemistry* 39, 12753–12756. doi: 10.1021/bi001811i

Regensburger, B., and Hennecke, H. (1983). RNA polymerase from *Rhizobium japonicum*. *Arch. Microbiol.* 135, 103–109. doi: 10.1007/BF00408017

Richardson, D., Felgate, H., Watmough, N., Thomson, A., and Baggs, E. (2009). Mitigating release of the potent greenhouse gas N2O from the nitrogen cycle - could enzymic regulation hold the key? *Trends Biotechnol.* 27, 388–397. doi: 10.1016/j.tibtech.2009.03.009

Richardson, D. J. (2011). "Redox complexes of the nitrogen cycle," in *Nitrogen Cycling in Bacteria*, ed. J. W. B. Moir (Norfolk: Caister Academic Press), 23–39.

Richardson, D. J., Bell, L. C., McEwan, A. G., Jackson, J. B., and Ferguson, S. J. (1991). Cytochrome c_2 is essential for electron transfer to nitrous oxide reductase from physiological substrates in *Rhodobacter capsulatus* and can act as an electron donor to the reductase in vitro. Correlation with photoinhibition studies. *Eur. J. Biochem.* 199, 677–683. doi: 10.1111/j.1432-1033.1991.tb16170.x

Robles, E. F., Sánchez, C., Bonnard, N., Delgado, M. J., and Bedmar, E. J. (2006). The *Bradyrhizobium japonicum napEDABC* genes are controlled by the FixLJ-FixK2-NnrR regulatory cascade. *Biochem. Soc. Trans.* 34, 108–110. doi: 10.1042/BST0340108

Sabaty, M., Schwintner, C., Cahors, S., Richaud, P., and Vermeglio, A. (1999). Nitrite and nitrous oxide reductase regulation by nitrogen oxides in *Rhodobacter sphaeroides* f. sp. *denitrificans* IL106. *J. Bacteriol.* 181, 6028–6032.

Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sameshima-Saito, R., Chiba, K., Hirayama, J., Itakura, M., Mitsui, H., Eda, S., et al. (2006). Symbiotic *Bradyrhizobium japonicum* reduces N_2O surrounding the soybean root system via nitrous oxide reductase. *Appl. Environ. Microbiol.* 72, 2526–2532. doi: 10.1128/AEM.72.4.2526-2532.2006

Sánchez, C., Itakura, M., Mitsui, H., and Minamisawa, K. (2013). Linked expressions of *nap* and *nos* genes in a *Bradyrhizobium japonicum* mutant with increased N_2O reductase activity. *Appl. Environ. Microbiol.* 79, 4178–4180. doi: 10.1128/AEM.00703-13

Sánchez, C., Itakura, M., Okubo, T., Matsumoto, T., Yoshikawa, H., Gotoh, A., et al. (2014). The nitrate-sensing NasST system regulates nitrous oxide reductase and periplasmic nitrate reductase in *Bradyrhizobium japonicum*. *Environ. Microbiol.* 16, 3263–3274. doi: 10.1111/1462-2920.12546

Sánchez, C., Mitsui, H., and Minamisawa, K. (2017). Regulation of nitrous oxide reductase genes by NasT-mediated transcription antitermination in *Bradyrhizobium diazoefficiens*. *Environ. Microbiol. Rep.* 9, 389–396. doi: 10.1111/1758-2229.12543

Sanford, R. A., Wagner, D. D., Wu, Q., Chee-Sanford, J. C., Thomas, S. H., Cruz-Garcia, C., et al. (2012). Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19709–19714. doi: 10.1073/pnas.1211238109

Schreiber, F., Wunderlin, P., Udert, K. M., and Wells, G. F. (2012). Nitric oxide and nitrous oxide turnover in natural and engineered microbial communities: biological pathways, chemical reactions, and novel technologies. *Front. Microbiol.* 3:372. doi: 10.3389/fmicb.2012.00372

Sciotti, M. A., Chanfon, A., Hennecke, H., and Fischer, H. M. (2003). Disparate oxygen responsiveness of two regulatory cascades that control expression of symbiotic genes in *Bradyrhizobium japonicum*. *J. Bacteriol.* 185, 5639–5642. doi: 10.1128/JB.185.18.5639-5642.2003

Shapleigh, J. P. (2006). "The denitrifying prokaryotes," in *The Prokaryotes: A Handbook on the Biology of Bacteria*, eds M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (New York, NY: Springer Science+Business Media), 769–792.

Simon, R., Priefer, U., and Pühler, A. (1983). "Vector plasmids for *in vivo* and *in vitro* manipulation of gram-negative bacteria", in *Molecular Genetics of the Bacteria-Plant Interaction*, ed A. Pühler (Heidelberg: Springer), 98–106.

Stein, L. Y. (2011). Surveying N_2O-producing pathways in bacteria. *Methods Enzymol.* 486, 131–152. doi: 10.1016/B978-0-12-381294-0.00006-7

Takaya, N. (2002). Dissimilatory nitrate reduction metabolisms and their control in fungi. *J. Biosci. Bioeng.* 94, 506–510. doi: 10.1016/S1389-1723(02)80187-6

Thomson, A. J., Giannopoulos, G., Pretty, J., Baggs, E. M., and Richardson, D. J. (2012). Biological sources and sinks of nitrous oxide and strategies to mitigate emissions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1157–1168. doi: 10.1098/rstb.2011.0415

Torres, M. J., Argandona, M., Vargas, C., Bedmar, E. J., Fischer, H. M., Mesa, S., et al. (2014). The global response regulator RegR controls expression of denitrification genes in *Bradyrhizobium japonicum*. *PLoS ONE* 9:e99011. doi: 10.1371/journal.pone.0099011

Tortosa, G., Hidalgo, A., Salas, A., Bedmar, E. J., Mesa, S., and Delgado, M. J. (2015). Nitrate and flooding induce N2O emissions from soybean nodules. *Symbiosis* 67, 125–133. doi: 10.1007/s13199-015-0341-3

Treusch, A. H., Leininger, S., Kletzin, A., Schuster, S. C., Klenk, H. P., and Schleper, C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ. Microbiol.* 7, 1985–1995. doi: 10.1111/j.1462-2920.2005.00906.x

Trunk, K., Benkert, B., Quack, N., Munch, R., Scheer, M., Garbe, J., et al. (2010). Anaerobic adaptation in *Pseudomonas aeruginosa*: definition of the Anr and Dnr regulons. *Environ. Microbiol.* 12, 1719–1733. doi: 10.1111/j.1462-2920.2010.02252.x

van Spanning, R. J. (2011). "Structure, function, regulation and evolution of the nitrite and nitrous oxide reductase: denitrification enzymes with a *b*-propeller fold," in *Nitrogen Cycling in Bacteria*, ed. J. W. B. Moir (Norfolk: Caister Academic Press), 135–161.

van Spanning, R. J., Delgado, M. J., and Richardson, D. J. (2005). "The nitrogen cycle: denitrification and its relationship to N2 fixation," in *Nitrogen Fixation in Agriculture, Forestry, Ecology and the Environment*, eds D. Werner and W. E. Newton (Dordrecht: Springer), 277–342. doi: 10.1007/1-4020-3544-6_13

van Spanning, R. J., Houben, E., Reijnders, W. N., Spiro, S., Westerhoff, H. V., and Saunders, N. (1999). Nitric oxide is a signal for NNR-mediated transcription activation in *Paracoccus denitrificans*. *J. Bacteriol.* 181, 4129–4132.

van Spanning, R. J., Richardson, D. J., and Ferguson, S. J. (2007). "Introduction to the biochemistry and molecular biology of denitrification," in *Biology of the Nitrogen Cycle*, eds H. Bothe, S. J. Ferguson, and W. E. Newton (Amsterdam: Elsevier Science), 3–20.

Velasco, L., Mesa, S., Delgado, M. J., and Bedmar, E. J. (2001). Characterization of the *nirK* gene encoding the respiratory, Cu-containing nitrite reductase of *Bradyrhizobium japonicum*. *Biochim. Biophys. Acta* 1521, 130–134. doi: 10.1016/S0167-4781(01)00279-2

Velasco, L., Mesa, S., Xu, C. A., Delgado, M. J., and Bedmar, E. J. (2004). Molecular characterization of *nosRZDFYLX* genes coding for denitrifying nitrous oxide reductase of *Bradyrhizobium japonicum*. *Antonie Van Leeuwenhoek* 85, 229–235. doi: 10.1023/B:ANTO.0000020156.334242470.db

Vollack, K. U., Hartig, E., Korner, H., and Zumft, W. G. (1999). Multiple transcription factors of the FNR family in denitrifying *Pseudomonas stutzeri*: characterization of four *fnr*-like genes, regulatory responses and cognate metabolic processes. *Mol. Microbiol.* 31, 1681–1694. doi: 10.1046/j.1365-2958.1999.01302.x

Vollack, K. U., and Zumft, W. G. (2001). Nitric oxide signaling and transcriptional control of denitrification genes in *Pseudomonas stutzeri*. *J. Bacteriol.* 183, 2516–2526. doi: 10.1128/JB.183.8.2516-2526.2001

Wunsch, P., and Zumft, W. G. (2005). Functional domains of NosR, a novel transmembrane iron-sulfur flavoprotein necessary for nitrous oxide respiration. *J. Bacteriol.* 187, 1992–2001. doi: 10.1128/JB.187.6.1992-2001.2005

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* 61, 533–616.

Zumft, W. G. (2005). Biogenesis of the bacterial respiratory CuA, Cu-S enzyme nitrous oxide reductase. *J. Mol. Microbiol. Biotechnol.* 10, 154–166. doi: 10.1159/000091562

Zumft, W. G., and Kroneck, P. M. (2007). Respiratory transformation of nitrous oxide (N2O) to dinitrogen by *Bacteria* and *Archaea*. *Adv. Microb. Physiol.* 52, 107–227. doi: 10.1016/S0065-2911(06)52003-X

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Torres, Bueno, Jiménez-Leiva, Cabrera, Bedmar, Mesa and Delgado. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 March 2018
doi: 10.3389/fmicb.2018.00575

Comparative Analysis of the Microbiota Between Sheep Rumen and Rabbit Cecum Provides New Insight Into Their Differential Methane Production

Lan Mi[1,2], Bin Yang[1], Xialu Hu[1], Yang Luo[1], Jianxin Liu[1], Zhongtang Yu[2] and Jiakun Wang[1]**

[1] Laboratory of Ruminant Nutrition, Institute of Dairy Science, College of Animal Sciences, Zhejiang University, Hangzhou, China, [2] Department of Animal Sciences, The Ohio State University, Columbus, OH, United States

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council (CONICET), Argentina

Reviewed by:
Wang Min, Institute of Subtropical Agriculture (CAS), China
Sanjay Kumar, University of Pennsylvania, United States
Stephan Schmitz-Esser, Iowa State University, United States

***Correspondence:**
Zhongtang Yu
yu.226@osu.edu
Jiakun Wang
jiakunwang@zju.edu.cn

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: 18 December 2017
Accepted: 13 March 2018
Published: 27 March 2018

Citation:
Mi L, Yang B, Hu X, Luo Y, Liu J, Yu Z and Wang J (2018) Comparative Analysis of the Microbiota Between Sheep Rumen and Rabbit Cecum Provides New Insight Into Their Differential Methane Production. Front. Microbiol. 9:575. doi: 10.3389/fmicb.2018.00575

The rumen and the hindgut represent two different fermentation organs in herbivorous mammals, with the former producing much more methane than the latter. The objective of this study was to elucidate the microbial underpinning of such differential methane outputs between these two digestive organs. Methane production was measured from 5 adult sheep and 15 adult rabbits, both of which were placed in open-circuit respiratory chambers and fed the same diet (alfalfa hay). The sheep produced more methane than the rabbits per unit of metabolic body weight, digestible neutral detergent fiber, and acid detergent fiber. pH in the sheep rumen was more than 1 unit higher than that in the rabbit cecum. The acetate to propionate ratio in the rabbit cecum was more than threefold greater than that in the sheep rumen. Comparative analysis of 16S rRNA gene amplicon libraries revealed distinct microbiota between the rumen of sheep and the cecum of rabbits. Hydrogen-producing fibrolytic bacteria, especially *Butyrivibrio*, *Succiniclastium*, *Mogibacterium*, *Prevotella*, and *Christensenellaceae*, were more predominant in the sheep rumen, whereas non-hydrogen producing fibrolytic bacteria, such as *Bacteroides*, were more predominant in the rabbit cecum. The rabbit cecum had a greater predominance of acetogens, such as those in the genus *Blautia*, order *Clostridiales*, and family *Ruminococcaceae*. The differences in the occurrence of hydrogen-metabolizing bacteria probably explain much of the differential methane outputs from the rumen and the cecum. Future research using metatranscriptomics and metabolomics shall help confirm this premise and understand the factors that shape the differential microbiota between the two digestive organs. Furthermore, our present study strongly suggests the presence of new fibrolytic bacteria in the rabbit cecum, which may explain the stronger fibrolytic activities therein.

Keywords: acetogen, cecum, fibrolytic bacteria, hydrogen, methane, microbiota, pH, rumen

Abbreviations: A, acetate; ADF, acid detergent fiber; BW, body weight; $BW^{0.75}$, metabolic body weight; CMCase, carboxymethyl cellulase; CP, crude protein DM, dry matter; DMI, dry matter intake; Eh, redox potential; *fhs*, formyltetrahydrofolate synthetase gene; *frd*A, fumarate reductase gene α subunit; MCCase, microcrystalline cellulose cellulase; MCP, crude microbial protein; *mcr*A, methyl CoM reductase gene α subunit; NDF, neutral detergent fiber; NZW rabbit, New Zealand white rabbit; P, propionate; RCC, rumen cluster C; VFA, volatile fatty acid.

INTRODUCTION

Mammalian herbivores do not synthesize the enzymes needed to digest cellulose or hemicellulose. They depend on a symbiotic relationship with a community of microbes (primarily bacteria) with fibrolytic ability in either their foregut (i.e., the rumen of ruminants and the pseudo-ruminants) or their hindgut (i.e., the cecum and colon of non-ruminant herbivores) for fiber digestion (Furness et al., 2015). Both foregut and hindgut fermenters produce methane (CH_4) as an inevitable by-product during feed fermentation. As a greenhouse gas, CH_4 is 23 times more potent than carbon dioxide (CO_2) (IPCC, 2014). A significant portion of the ingested feed energy is also lost as CH_4, ranging from 1.5 to 12% of the gross energy intake in cattle (Johnson and Johnson, 1995; Franz et al., 2010). Ruminants are the main producing animals of meat and milk, but they also produce more CH_4 than monogastric animals per unit of $BW^{0.75}$ or product (Franz et al., 2010). Indeed, up to 20% of the global anthropogenic CH_4 is emitted by ruminants (Bhatta et al., 2007). Intensive research has aimed to mitigate CH_4 emission to ensure sustainable production of beef, lamb, and dairy products.

Methanogenesis in the rumen and hindgut is predominately driven via the hydrogenotrophic pathway using hydrogen (H_2) and CO_2 (also formate) as the substrates (Liu and Whitman, 2008) though some CH_4 is also produced through the methylotrophic methanogenesis pathway using methanol and methylamines as the substrates (Poulsen et al., 2013). The genus *Methanobrevibacter* is the most ubiquitous and predominant hydrogenotrophic methanogens found in the foregut and hindgut of herbivores (Kušar and Avguštin, 2010; St-Pierre and Wright, 2013) although several species of *Methanomassiliicoccales* can use methyl substrates (Poulsen et al., 2013). Although affected by several factors, such as pH, the rate and CH_4 output from herbivores are primarily determined by the availability of methanogenic substrates (i.e., H_2 and CO_2), which are in turn determined by the rates of production and consumption. Fermentative acetate production accompanied with H_2 production is thermodynamically favored, especially when forage-based diets are fed because more ATP is synthesized (Russell and Rychlik, 2001). CH_4 output can vary among cows or sheep fed the same diet (Kittelmann et al., 2014; Wallace et al., 2015), and CH_4 output was found to be positively associated with bacterial populations that ferment ingested feed to relatively more hydrogen in sheep (Kittelmann et al., 2014). Furthermore, hydrogenotrophic methanogens are thermodynamic favored than acetogens when competing for hydrogen in rumen (Cord-Ruwisch et al., 1988; Joblin, 1999), but in the foregut (tubular) of kangaroos, acetogens outcompete methanogens for CO_2 and H_2 and can synthesize acetate via the acetyl-CoA pathway, providing a significant energetic benefit to the host animal (Attwood and McSweeney, 2008). Similar hydrogen disposal pathway was thought to be present in the cecum of rabbits, but no acetogens were reported (Piattoni et al., 1996). In a recent study, species of *Blautia* including *B. coccoides*, *B. hydrogenotrophica*, and *B. schinikii*, which are known acetogens, were found at high predominance in rabbit cecum (Yang et al., 2016). We hypothesized that the hindgut of hindgut fermenters probably also has a distinct microbiota than the rumen of ruminants, and such difference may be the main reason for the differential CH_4 outputs between these two types of herbivores. In the present study, we tested this hypothesis using sheep as ruminants and New Zealand White (NZW) rabbits as a non-ruminant herbivore, with alfalfa hay as the only diet. Feed consumption, fermentation characteristics, CH_4 emission, and the microbiota in the sheep rumen and the rabbit cecum were comparatively analyzed. The differences determined in the above measurements will help understand the physiological and microbial underpinnings of differential CH_4 production between ruminat and non-ruminant herbivores, and the knowledge on correlations between the microbiota and CH_4 production might be useful for targeted intervention of rumen microbiota to mitigate CH_4 production from ruminants.

MATERIALS AND METHODS

Animals, Diets, and Experimental Design

All experiments involving animals and the animal use protocols were approved by the Animal Care Committee of Zhejiang University (Hangzhou, China). Five 1.5-years old healthy male sheep (63.91 $\pm$ 6.18 kg BW) each with a permanent ruminal cannula were each allocated to an open circuit respiration chamber, which was constructed using aluminum frames and resin sheets, allowing animals in neighboring chambers see each other. Temperature and humidity inside the chambers were respectively maintained at 25°C and 60%. Before gas determination, both the door and the food hopper of each chamber were kept open. Fifteen 1-year old healthy male NZW rabbits (3.14 $\pm$ 0.14 kg BW) were each housed in an indoor cage (60 $\times$ 50 $\times$ 35 cm in dimensions). Both the chambers and the cages were placed in a temperature-regulated room (24–26°C) with a natural light-dark cycle (approximately 13 h of light and 11 h of dark). Both the sheep and the rabbits were fed the same diet consisting of only alfalfa hay (18.5% CP, 46.0% NDF and 33.0% ADF) and had *ad libitum* access to fresh drinking water during the whole feeding experiment of 23 days for the sheep and 24 days for the rabbits. The feeding experiment consisted of 15 days for acclimation, 7 days for sample collection, and 1 day for gas measurement.

Measurement of Feed Intake and Digestibility

At the beginning of the experiment, the rabbits were blocked by BW (5 blocks, 3 rabbits per block), and each block was transferred to an indoor metabolism cage. The amount of feed offered and refused was recorded daily, and all feces were collected using fecal collection plates at 8:30 AM daily from individual sheep and blocks of rabbits during the 7 days of sample collection. Daily feed and orts samples were pooled by experimental unit (individual sheep and rabbit blocks). Fecal output was weighed, and 100 g wet feces were added to 10 mL 10% hydrochloric acid to preserve the samples for nitrogen analysis. All the feed and the fecal samples were dried in a forced-air oven at 65°C for 72 h, then ground

through a 1 mm screen, and stored in sealed plastic containers at 4°C until analysis. Standard analysis methods (Official Methods of Analysis [AOAC], 2015) were used for analysis for dry matter (DM, method 930.15), CP (method 990.03), and ADF (method 973.18). NDF contents were analyzed following the procedure of Van Soest et al. (1991) without sodium sulfite and amylase added. The BW was weighted before measuring gas production at day 22 of the feeding experiment.

Determination of Gas Production

Gas production was determined using open circuit chambers (1.16 m^3 interior volume each). Each chamber was completely airtight but received a continuous air flow at 8.0 m^3 h^{-1}. Total air flow was recorded using a flow meter (model number: SY-LWD-B-20; Shi Yi Automation Equipment Co., Ltd., Hangzhou, China), and concentration of CH_4 and CO_2 was determined using a gas detector (model number: Photoacoustic Gas Monitor INNOVA 1412; Innova AirTech Instruments A/S, Ballerup, Denmark). The alfalfa hay diet was provided to the animals twice daily at 09:00 and 16:00 using a food hopper that was reloaded outside of the chambers via a lid without opening the whole chamber. After the sample collection period, gas production from each sheep was directly determined for 24 h at day 23 of the feeding experiment. Each block of the rabbits was transferred to a chamber at day 23 of the feeding experiment for 24 h acclimation before continuous gas measurement for 24 h at day 24. CH_4 production was expressed as CH_4 yield per kg of $BW^{0.75}$, DMI, digestible NDF intake, and digestible ADF intake, while CO_2 production was expressed as CO_2 yield per kg of $BW^{0.75}$.

Collection of Ruminal and Cecal Samples

Ruminal content samples (about 50 mL each) were taken from individual sheep through their rumen cannula immediately after completing the gas determination. All the rabbits were sacrificed following euthanasia that was administered by a licensed animal technician following the procedures described by Yang et al. (2016). A quiet environment was provided to individual rabbit on a table with a slight angle to avoid stress and minimize pressure on the diaphragm. Rabbits were intravenous injected phenobarbital sodium (Sigma, Saint Louis, MO, United States) with a dose of 100 mg/kg BW. Once losing toe pinch and leg withdrawal reflex, each rabbit received ear intravenous injection of 20 mL of air. Then, cecal content samples were immediately collected (Yang et al., 2016). Briefly, each cecum was delineated into its proximal, middle, and distal segments, which were tied at the boundaries with a nylon string to prevent the cecal digesta from moving longitudinally. Each of the three cecal segments was cut separated, and its digesta content was squeezed into one 50-ml sterile Falcon tubes within 30 min of decease. One composite sample was prepared for each rabbit by combining about the same amount of digesta from each cecal segment. After immediate pH measurement using a pH meter (PB-10; Sartorius, Göettingen, Germany), the cecum samples were stored in liquid nitrogen and transported to the laboratory. Approximately 20 g of each rumen content and cecal content sample were freeze-dried for 30 h using a freeze-dryer (model number: BETA 1-8 LD; Martin Christ Gefriertrocknungsanlagen GmbH, Osterode, Germany). Each of the freeze-dried samples was crushed into fine particles manually and stored at −80°C until further analysis.

Measurement of Volatile Fatty Acid (VFA) Concentrations

An aliquot of each ruminal and cecal sample was subjected to analysis for VFAs as described by Yang et al. (2015). Briefly, approximately 2 g of each ruminal content sample and 1 g of each cecal content sample were added to 5 mL and 3 mL sterile phosphate buffered saline (PBS, pH 7.0), respectively, and mixed, and the mixture was centrifuged at 13,000 × *g* at 4°C for 15 min. To 1 mL of each supernatant were added 20 μL of 85% orthophosphate acid and centrifuged again as described above to obtain the final supernatant. The VFAs concentration in the supernatant was determined using a gas chromatograph (model number: GC-2010; Shimadzu Corp., Kyoto, Japan) against external standards purchased from Aladdin (China, Shanghai).

Measurement of Microbial Enzyme Activity

The activities of CMCase, MCCase, xylanase, and pectinase of each sample was determined essentially as described previously (Wang et al., 2015), using carboxymethyl cellulose sodium (Sigma-Aldrich, Saint Louis, MO, United States), microcrystalline cellulose (Sigma-Aldrich), beechwood xylan (Sigma-Aldrich), and pectin from citrus peel (Fluka, St. Louis, MO, United States) as respective substrates, according to the dinitrosalicylic acid (DNS) method expounded by Bailey et al. (1992). Briefly, approximately 0.5 g of each freeze-dried ruminal content sample or cecal content sample was vortexed in 6 mL sterile PBS (pH 7.0), and the sample suspension was then sonicated (20 kHz, 195 W, 10 min) using a JY92-IIN Ultrasonic Cell Mixer (Ningbo Scientz, Ningbo, China) and centrifuged at 12,000 × *g* at 4°C for 10 min. Then, 0.2 mL of the supernatant of each sample was mixed with 0.2 mL corresponding substrates (0.01 g mL^{-1} in phosphate buffer, pH 6.6) and then incubated at 39°C for 30 min. The enzyme activity was expressed as μmoL of glucose (for CMCase and MCCase), xylose (for xylanase), or D-galacturonic acid (for pectinase) released min^{-1} g^{-1} of the freeze-dried samples or their microbial crude protein (MCP).

Measurement of Microbial Crude Protein

Approximately 0.5 g of each freeze-dried ruminal content sample or cecal content sample was vortexed in 6 mL sterile PBS buffer solution (pH 7.0) to get the microbes in the gut fluid and most of the microbes adhering to the feed particles, and the suspension was then centrifuged at 408 × *g* for 5 min to remove protozoa and remain feed particles. Then, 1 mL of each supernatant was centrifuged at 25,000 × *g* at 4°C for 20 min. The supernatants were discarded, and the pelleted microbial cells were suspended in 3 mL of 0.25 N sodium hydroxide and heated in boiling water for 10 min to lyse the microbial cells. The cell lysate samples were centrifuged at 25,000 × *g* for 30 min, and the supernatants were subjected to protein assay with bovine serum albumin as the standard using the Coomassie brilliant blue (CBB) method

(Makkar et al., 1982). The content of MCP was expressed as mg g^{-1} freeze-dried ruminal content or cecal content samples.

DNA Extraction and Real-Time PCR Quantification

Metagenomic DNA was extracted from 0.1 g each of freeze-dried ruminal content sample or cecal content sample using the CTAB (cetyltrimethylammonium bromide) method but with bead-beating (Gagen et al., 2010). The quality of the DNA extracts was evaluated using agarose (1%) electrophoresis, while the DNA concentration was determined using the Qubit dsDNA BR Assay Kit (Invitrogen Corporation, United States) on a Qubit 2.0 fluorometer (Invitrogen Corporation, United States). Standards for qPCR assay were prepared for individual groups of targeted microbes or targeted genes (the primers were listed in Supplementary Table S1) using cloning of PCR amplicons with a pGEM®T Easy kit (Promega, Shanghai, China) following the method of Koike et al. (2007). The abundance of each species or group of microbes was quantified using real-time PCR as described previously (Liu et al., 2014) and expressed as $\log_{10}$ copies of 16S rRNA gene (or 18S rRNA gene in the case of protozoa, and ITS1 in the case of fungi) per g of freeze-dried ruminal content or cecal content samples.

Analysis of Microbiota

One amplicon library each was separately prepared for archaea and bacteria from each of the metagenomic DNA samples using the primers M86F/M448R and 515F/806R, respectively. All amplicons were pooled in equal molar ratio and sequenced using the 2 × 250 bp paired-end protocol on an Illumina MiSeq system. The raw sequences were de-multiplexed, quality-filtered, and analyzed using QIIME (v 1.9.0) (Caporaso et al., 2010). Briefly, bases from each sequencing read with a Q score less than 25 were trimmed off, then the paired reads (R1 and R2) were merged to form single sequences using the fastq-join script (Aronesty, 2011). Sequences shorter than 352 bp for archaea and 281 bp for bacteria were discarded, and the primers were further trimmed off. Chimera checking was performed using the ChimeraSlayer algorithm (Haas et al., 2011). The quality-checked sequences were clustered into species-equivalent operational taxonomic units (OTUs) by comparison to the Greengenes database 13.5 (DeSantis et al., 2006) using the open-reference OTU picking option (pick_open_reference_otus.py). The OTUs were taxonomically classified by comparison to the Greengenes database 13.5. Minor OTUs were filtered out if they were each represented by less than 0.005% of the total sequences (Bokulich et al., 2013) or appeared in less than 60% of each experimental animal species. The sequences of each sample were rarefied to the same number (46,609 sequences/sample for archaea and 19,170 sequences/sample for bacteria) before diversity analysis. Alpha diversity measurements including Chao1 richness estimate, Shannon diversity index, and observed number of OTUs were calculated for each sample. The microbiota were compare as beta diversity using the distance matrices generated from weighted UniFrac analysis (Lozupone and Knight, 2005) and principal coordinates analysis (PCoA). The raw sequence data were deposited in the Sequence Read Archive of NCBI under accession no. SRP108266.

Statistical Analysis

Statistical analysis of the data was performed using one-way ANOVA, with means separation using *t*-test at the level of significance of 0.05 using the SAS software package (SAS Institute Inc., 2000). Pearson correlation coefficients were calculated to examine the correlation between animal performances and relative abundance of microbial groups. The data were expressed as Mean ± SD in the Tables.

RESULTS

Feed Digestibility and Gas Yields

The two species of animals used in this study consumed a similar amount of feed (DM) per unit of $BW^{0.75}$ daily (**Table 1**). However, the sheep had a higher apparent digestibility of DM, NDF, and ADF than the rabbits ($P < 0.05$). Each of the sheep emitted substantially more CH_4 than each rabbit per day per unit of $BW^{0.75}$, DMI, digestible NDF, or digestible ADF (**Figures 1A–D**). Per unit of $BW^{0.75}$, however, the sheep emitted less CO_2, resulting in a 6.4 times higher CH_4 to CO_2 ratio than the rabbits (**Figures 1E,F**).

Fermentation Characteristics

The main fermentation characteristics, including pH, concentrations of VFA, and their molar proportions, in the sheep rumen and the rabbit cecum are presented in **Table 2**. The pH in the sheep rumen was more than 1 unit higher ($P < 0.01$) than that in the rabbit cecum. No significant difference ($P = 0.16$) in total VFA concentration was observed between the two digestive organs. However, a much lower propionate concentration was seen in the rabbit cecum than in the sheep rumen. The two digestive organs also differed ($P \leq 0.01$) in molar proportions of VFA, with the rabbit cecum having a higher value for acetate and butyrate but a lower value for propionate. The acetate to propionate (A: P) ratio in the rabbit cecum was more than threefold greater than that in the sheep rumen.

Microbial Crude Protein Yields and Enzymes Activity

The rabbit cecal content had a higher ($P < 0.01$) concentration of MCP than the sheep rumen content (**Table 3**). A higher ($P < 0.01$) activity of CMCase, MCCase, and pectinase was observed in the rabbit cecum than in the sheep rumen either per g content or mg MCP. Xylanase activity was similar between the two digestive organs (**Table 3**).

Abundance of Select Microbes and Genes Involved in Hydrogen Metabolism

The total bacterial population ($\log_{10}$ 16S rRNA gene copies/g sample) was larger in the rabbit cecum than in the sheep rumen (**Figure 2A**). The abundance of *Ruminococcus albus*, *R. flavefaciens*, *Fibrobacter succinogenes*, and *Butyrivibrio*

TABLE 1 | Body weight, feed dry matter intake, and digestibility of the sheep (n = 5) and the rabbits (n = 15).

Parameter	Units	Sheep	NZW rabbits	*P*-value
Body weight	kg	63.91 ± 6.18	3.14 ± 0.13	<0.01
	$kg^{0.75}$	22.59 ± 1.65	2.36 ± 0.075	<0.01
Dry matter intake	g/d/kg of $BW^{0.75}$	63.42 ± 3.34	67.66 ± 15.12	0.56
Apparent digestibility	%			
Dry matter		68.91 ± 1.76	57.76 ± 7.97	0.02
NDF		60.05 ± 3.09	41.35 ± 11.83	<0.01
ADF		59.28 ± 3.15	38.26 ± 13.55	<0.01

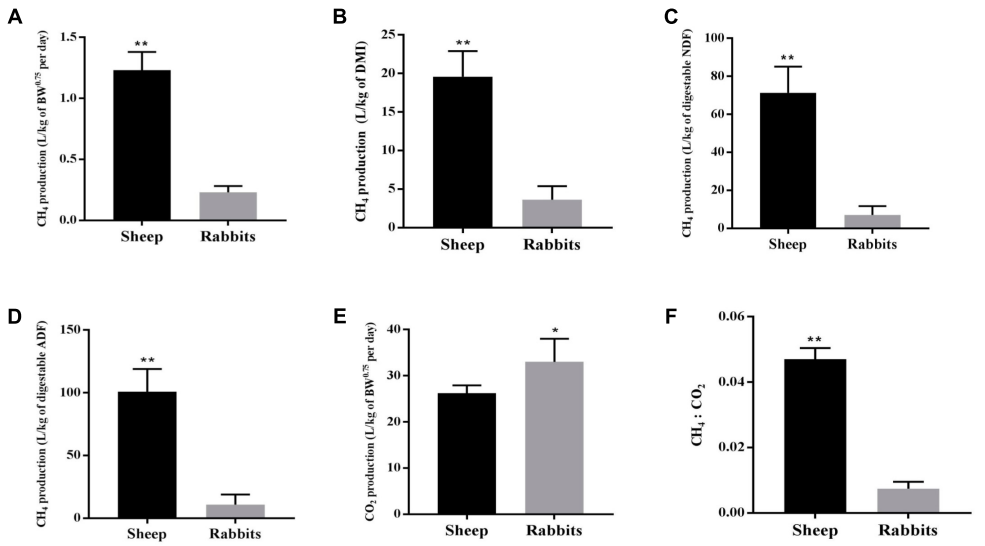


FIGURE 1 | Yields of CH_4 and CO_2 and their ratio of the sheep (n = 5) and the rabbits (n = 15). CH_4 yields per unit of metabolic body weight **(A)**, dry matter intake **(B)**, digestible NDF **(C)**, and digestible ADF **(D)**; CO_2 yields per unit of metabolic body weight **(E)** and CH_4 : CO_2 ratio **(F)** of the sheep and the rabbits. $^{**}P < 0.01$; $^{*}P < 0.05$.

fibrisolvens was greater in the sheep rumen than in the rabbit cecum. The same holds true for the abundance of fungi and protozoa. The abundance of *Clostridium* Cluster XIVa was similar between the two digestive organs, while that of *Clostridium* Cluster IV was greater in the rabbit cecum than in the sheep rumen (**Figure 2B**). The copy number of *mcr*A gene per g sample was greater in the sheep rumen than in the rabbit cecum, while that of *fhs* gene and *frd*A gene was smaller (**Figure 2C**). The sheep rumen had a greater abundance of RCC methanogens and non-RCC methanogens (including *Methanobrevibacter*, *Methanomicrobium*, *Methanobacterium*, *Methanomicrococcus,* and *Methanosphaera*) than the rabbit cecum. The archaea : bacteria ratio differed between the two different digestive organs, 0.089 in the sheep rumen and 2.30E-05 in the rabbit cecum.

Diversity, Species Richness, and Composition of Archaeal Microbiota

The Chao1 richness estimate was similar between the two digestive organs, but the rabbit cecum had a lower Shannon diversity index and Simpson index of diversity than the sheep rumen (**Table 4A**). The archaeal microbiota of the two digestive organs clustered separately along the PC1 that explained greater than 84% of total variation. The rabbit cecal archaeal microbiotas clustered relatively tightly, while those of the sheep rumen quite scattered along PC2 that explained less than 10% total variation (**Figure 3A**).

TABLE 2 | Fermentation characteristics of the sheep rumen (n = 5) and the rabbit cecum (n = 15).

Measurements	Units	Sheep rumen	Rabbit cecum	*P*-value
pH		7.11 ± 0.17	5.82 ± 0.25	<0.01
Concentration	mmol/kg			
Total VFA		67.15 ± 10.54	55.70 ± 12.97	0.16
Acetate		51.73 ± 7.90	46.37 ± 10.78	0.40
Propionate		9.12 ± 1.50	2.39 ± 0.45	<0.01
Butyrate		6.31 ± 1.65	6.95 ± 1.94	0.59
Molar proportion	%			
Acetate		77.07 ± 1.79	83.26 ± 1.54	<0.01
Propionate		13.58 ± 0.97	4.33 ± 0.26	<0.01
Butyrate		9.34 ± 1.35	12.41 ± 1.61	0.01
Acetate/Propionate		5.70 ± 0.50	19.34 ± 1.07	<0.01

TABLE 3 | Microbial crude protein (MCP) and enzyme activities in the sheep rumen (n = 5) and the rabbit cecum (n = 15).

Measurements	Units	Sheep rumen	Rabbit cecum	*P*-value
MCP	mg/g content	24.01 ± 2.78	30.76 ± 2.03	<0.01
Absolute enzyme activities	U/g content			
CMCase[1]		2.00 ± 0.30	4.20 ± 0.48	<0.01
MCCase[2]		2.06 ± 0.16	5.86 ± 0.43	<0.01
Xylanase		0.050 ± 0.02	0.052 ± 0.01	0.88
Pectinase		8.80 ± 1.15	20.40 ± 3.29	<0.01
Specific enzyme activities	U/mg MCP			
CMCase[1]		0.086 ± 0.018	0.14 ± 0.015	<0.01
MCCase[2]		0.088 ± 0.011	0.19 ± 0.015	<0.01
Xylanase		0.0021 ± 0.001	0.0017 ± 0.001	0.33
Pectinase		0.37 ± 0.038	0.66 ± 0.069	<0.01

[1]*Carboxymethyl cellulase.* [2]*Microcrystalline cellulose cellulase.*

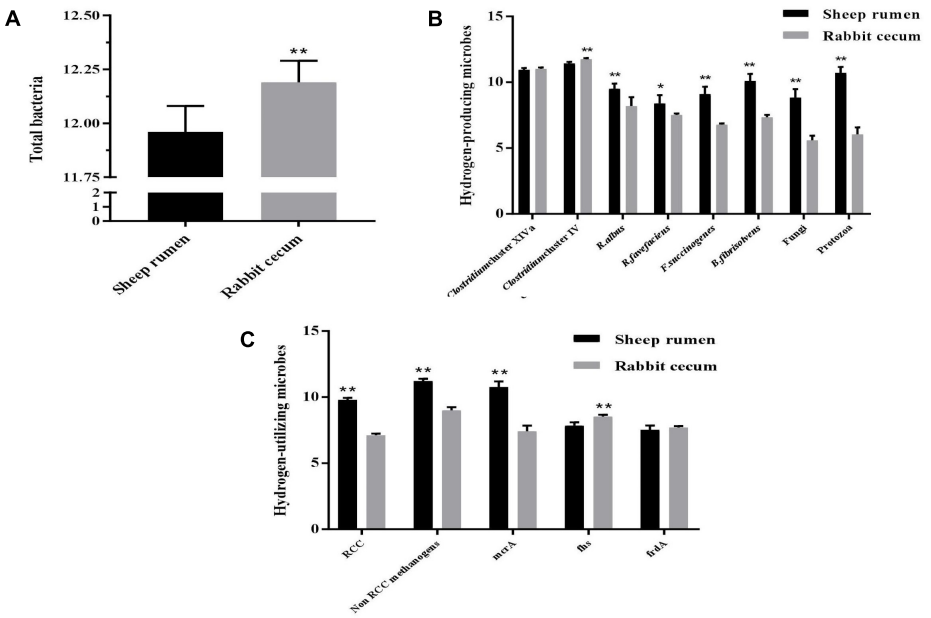


FIGURE 2 | Abundance of select microbes in the sheep rumen (n = 5) and the rabbit cecum (n = 15). **(A)** Total bacteria; **(B)** Select hydrogen-producing microbes; **(C)** Select hydrogen-utilizing microbes. Abundance was expressed as log10 copies of target gene/g freeze-dried digesta sample. Target gene: 16S rRNA gene for bacteria and methanogens (except for *mcr*A); ITS1 for fungi; 18S rRNA for protozoa; *mcr*A, methyl-CoA reductase α subunit gene; *fhs*, Formyltetrahydrofolate synthetase gene; *frd*A, Fumarate reductase gene α subunit gene. $^{**}P < 0.01$; $^{*}P < 0.05$.

Approximately 99.7% of the sequences obtained from both the digestive organs were assigned to known archaeal genera (**Figure 4**). More than 95% of the archaeal sequences from the rabbit cecum were assigned to the genus *Methanobrevibacter*, while the archaeal sequences from the sheep rumen were assigned to *Methanobrevibacter* (68.3%), *Methanosphaera* (17.3%), and unidentified achaeon vadinCA11 (14.2%). The two digestive organs shared 120 archaeal OTUs besides their unique archaeal OTUs. The 30 OTUs unique to the sheep rumen were assigned to *Methanobrevibacter* (8 OTUs), vadinCA11 (10 OTUs), and *Methanosphaera* (8 OTUs), together accounting for 3.6% of total archaeal sequences identified therein. Of the 59 OTUs only found in the rabbit cecum, 53 were classified to *Methanobrevibacter*, and these 53 *Methanobrevibacter* OTUs accounted for only 1.1% of total archaeal sequences identified in the rabbit cecum. A significant portion of the archaeal sequences was assigned to known species, with 36.3% assigned to ****M. thaueri*, 14.6% to *M. woesei*, and 10.8% to *M. millerae* for the sheep rumen sequences, while for the rabbit cecal sequences, 74.6% to *M. woesei*, and 13.9% to *M. thaueri*.

Diversity, Species Richness, and Composition of Bacterial Microbiota

The number of OTUs, Chao1 richness estimate, Shannon diversity index, and Simpson index of diversity in the sheep rumen were all significantly greater than those in the rabbit cecum (**Table 4B**). When compared using weighted UniFrac

distance, the rabbit cecal bacterial microbiotas were separated, as a tight cluster, from those of the sheep rumen along the PC1 that explained greater than 83% of total variation (**Figure 3B**). The sheep rumen bacterial microbiotas of the five sheep exhibited considerable scattering along PC2, but it only explained less than 6% of total variation.

Almost all the sequences obtained from both the sheep rumen and the rabbit cecum were assigned to known bacterial phyla, with *Firmicutes* (48.7 vs. 56.1%) and *Bacteroidetes* (47.4 vs. 36.1%) being represented by more sequences than other phyla (**Figure 5A**). Seven bacterial phyla were identified in both the sheep rumen and the rabbit cecum, which included, in addition to the above two predominant phyla, *Actinobacteria*, *Proteobacteria*, *Tenericutes*, *Verrucomicrobia*, and *Synergistetes*. However, another five bacterial phyla, i.e., *Spirochaetes*, *Chloroflexi*, *Fibrobacteres*, *Planctomycetes*, and candidate phylum SR1 were only found in the sheep rumen. Of the common bacterial phyla, the relative abundance of *Verrucomicrobia* was significantly greater in the rabbit cecum than in the sheep rumen (5.7 vs. 0.6%, $P < 0.01$).

In total, 840 and 728 OTUs were respectively observed in the sheep rumen and the rabbit cecum, and only 51 OTUs were found in both digestive organs. These common OTUs were assigned, at the lowest taxonomic rank, to the candidate order RF39 (3 OTUs), *Clostridiales* (23 OTUs), *Bacteroidales* (2 OTUs), candidate family S24-*7* (1 OTUs), *Ruminococcaceae* (7 OTUs), *Lachnospiraceae* (4 OTUs), *Ruminococcus* (3 OTUs), *Oscillospira* (3 OTUs), *Blautia* (3 OTUs), and *Bacteroides* (1 OTUs). These common OTUs represented 9.6% and 16.9% of the total bacterial sequences in the sheep rumen and the rabbit cecum, respectively.

The OTUs found in the sheep rumen and the rabbit cecum was assigned to 49 and 45 lowest possible taxa, respectively. However, only 25 (for the sheep rumen) and 24 (for the rabbit cecum) of them are recognized genera. The two digestive organs shared 18 common taxa, leaving 31 taxa being unique in the sheep rumen and 27 taxa found only in the rabbit cecum (**Figure 5B**). The predominant taxa (each represented by >1.0% of total bacterial sequences in at least 3 of the 5 experiment units) common to the sheep rumen and the rabbit cecum included *Bacteroidales* (11.1 vs. 12%, $P = 0.83$), *Lachnospiraceae* (4.9 vs. 5.3%, $P = 0.82$), and *Ruminococcus* (1.4 vs. 1.9%, $P = 0.24$). A small number of taxa had different relative abundance between the sheep rumen and the rabbit cecum, with *Ruminococcaceae* (10.1 vs. 24.1%, $P < 0.01$), *Clostridiales* (12 vs. 18%, $P = 0.03$), *Oscillospira* (0.1 vs.1.2%, $P < 0.01$), *Blautia* (0.4 vs. 1.7%, $P = 0.01$), and *Clostridium* (0.5 vs. 1.7%, $P = 0.01$) being less predominant, while *Christensenellaceae* (2.2 vs. 0.1%, $P < 0.01$) and candidate family S24-7 (11.3 vs. 2.1%, $P = 0.02$) being more predominant in the sheep rumen than in the rabbit cecum. The major unique taxa (with a relative abundance of >1.0% in at least 3 of the 5 experiment units) in the sheep rumen included *Prevotella* (21.3%), *Butyrivibrio* (9%), *Succiniclasticum* (3%), candidate family BS11 (1.7%), candidate family [*Paraprevotellaceae*] (1.4%), and *Mogibacterium* (1.3%). *Bacteroides* (16.9%), *Akkermansia* (5.7%), *Rikenellaceae* (2.9%), and candidate family [*Barnesiellaceae*] (2%) were the major taxa unique to the rabbit cecum.

TABLE 4 | Alpha diversity measurements of archaeal **(A)** and bacterial **(B)** microbiota of the sheep rumen ($n = 5$) and the rabbit cecum ($n = 5$ blocks).

Measurements	Sheep rumen	Rabbits cecum	*P*-value
(A)			
Number of sequences per sample	50546 ± 2681	110748 ± 14572	
Observed species	134.70 ± 8.90	138.72 ± 23.99	0.73
Chao1	142.47 ± 7.88	156.00 ± 14.44	0.10
Shannon index	3.20 ± 0.33	2.31 ± 0.26	<0.01
Simpson	0.80 ± 0.08	0.65 ± 0.04	0.01
(B)			
Number of sequences per sample	34284 ± 7253	32452 ± 2983	
Observed species	721.80 ± 36.31	647.38 ± 12.15	<0.01
Chao1	768.51 ± 32.43	687.28 ± 20.27	<0.01
Shannon index	7.38 ± 0.25	6.60 ± 0.10	<0.01
Simpson index	0.98 ± 0.01	0.96 ± 0.01	<0.01

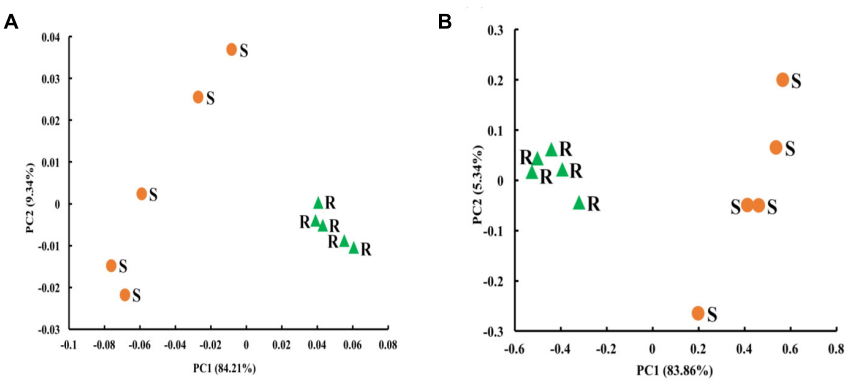


FIGURE 3 | Principal coordinates analysis (PCoA) plots in the sheep rumen ($n = 5$) and the rabbit cecum ($n = 5$ blocks). **(A)** Archaeal microbiota; **(B)** Bacterial microbiota. The PCoA were based on distance matrices generated using weighted UniFrac. S, sheep rumen samples; R, rabbits cecal samples.

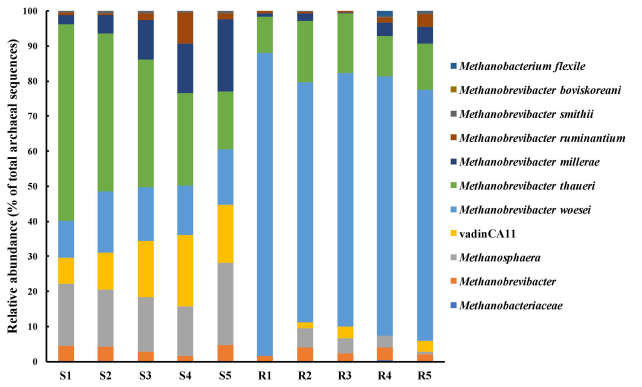


FIGURE 4 | Comparison of the relative abundance of species and genera of methanogens identified in the two digestive organs. S1–S5, the five sheep; R1–R5, the five blocks of rabbits.

Pearson Correlation Between Chemical Parameters and Dominant Bacterial Taxa

Pearson correlation coefficients were calculated to reveal correlations between the animal phenotypic data and the predominant bacterial taxa (**Figure 6A**). The relative abundance of *Butyrivibrio*, *Prevotella*, *Succiniclasticum*, *Mogibacterium*, *Christensenellaceae*, candidate family [*Mogibacteriaceae*], and candidate family S24-7 appeared to be positively correlated ($P < 0.05$) with both CH_4 yield and feed digestibility, whereas that of *Oscillospira*, *Ruminococcaceae*, *Clostridiales*, candidate genus [*Ruminococcus*], *Blautia*, *Clostridium*, *Bacteroides*, *Rikenellaceae* and *Akkermansia* was negatively ($P < 0.05$) correlated with these two measurements. A positive correlation was also seen between A:P ratio and the relative abundance of some taxa, including *Oscillospira*, *Ruminococcaceae*, *Clostridiales*, candidate genus [*Ruminococcus*], *Blautia*, *Clostridium*, *Bacteroides*, *Akkermansia*, and *Rikenellaceae*.

Pearson Correlation Between the Archaeal Taxa and Bacterial Taxa in Relative Abundance

Pearson correlation coefficients were calculated to reveal correlations between the relative abundance of archaea and bacteria (**Figure 6B**). The relative abundance of *M. thaueri* appeared to be positively correlated ($P < 0.05$) with *Succiniclasticum*, *Mogibacterium*, candidate family [*Mogibacteriaceae*], *Prevotella*, and CH_4 yields, and was negatively ($P < 0.05$) correlated with *Akkermansia*. The relative abundance of *Methanosphaera* and unidentified achaeon vadinCA11 appeared to be positively correlated ($P < 0.05$) with *Butyrivibrio*, *Succiniclasticum*, *Christensenellaceae*, candidate family [*Mogibacteriaceae*], BS11, *Prevotella*, S24-7, as well as CH_4 yields, whereas were negatively ($P < 0.05$) correlated with the relative abundance of *Oscillospira*, *Ruminococcaceae*, candidate genus [*Ruminococcus*], *Blautia*, *Clostridium*, *Bacteroides*, and *Akkermansia*. The opposite correlations of *M. woesei* with both CH_4 yields and those bacteria were showed.

DISCUSSION

Both ruminants and non-ruminant herbivores emit CH_4, but the former emits much more CH_4 than the latter (Franz et al., 2010, 2011; Cabezas Garcia, 2017). It has been speculated that such difference in CH_4 emission is probably attributable primarily to the differences in the microbiota of the rumen and the hindgut of non-ruminant herbivores (Yang et al., 2016). However, the microbiological peculiarity for the different CH_4 emissions by these two groups of herbivores is largely unknown. Identification of these responsible microbes and the relationship to CH_4 emission and the fermentation characteristics of the rumen and the cecum will help understand the factors that affect CH_4 production in the rumen and develop dietary strategies to effectively mitigate CH_4 emission from ruminants. In the present study, we comparatively characterized the microbiota and the fermentation characteristics in the rumen of sheep and the cecum of rabbits when fed the same diet. This approach allowed us to quantitatively determine and compare CH_4 production by a representative species of ruminants and non-ruminants on the basis of feed intake, feed digestibility, and metabolic $BW^{0.75}$. This approach overcomes the limitation of using the same ruminant animal in which CH_4 emission from the rumen and large intestines cannot be independently determined, and the two digestive organs received different fermentation substrates.

The rabbits produced no more than ¼ of the amount of CH_4 produced by the sheep when compared on per unit of $BW^{0.75}$, DMI, digestible NDF or ADF. Franz et al. (2010, 2011) proposed a linear relationship between BW and CH_4 production by both ruminants and non-ruminants. However, the magnitude of different CH_4 outputs between the sheep and the rabbits probably suggests physiological and microbiological peculiarities of these two digestive organs. First, the pH inside the rabbit cecum was nearly 1.3 units lower than that in the sheep rumen. It is well documented that methanogenesis is inhibited at low pH, as exemplified by no CH_4 production at pH 5.5 or below in *in-vitro* cultures inoculated with rumen fluid from roughage-fed cows (Russell, 1998). Thus, the lower pH in the rabbit cecum (pH 5.8) than in the sheep rumen (pH 7.1) is probably a major chemical factor attributable to the low CH_4 output from the rabbits. The rumen receives a large volume of saliva (about 1.31 L of saliva is secreted from one parotid gland per day for an adult sheep), which buffers the acidity from VFA (McDougall, 1948), while cecum receives no saliva. The lack of saliva secretion to the cecum is probably one of the reasons for the lower pH in the rabbit cecum than in the sheep rumen. Second, the Eh in the rabbit cecum ranges from −160 to −210 mV (Kimsé et al., 2010; Michelland et al., 2010), while the rumen Eh ranges from −268 to −318 mV (Mathieu et al., 1996). We did not analyze the Eh in the present study, but it should be within the above range. The tubiform and small in diameter of the rabbit cecum may explain its relatively higher Eh. Apparently, the Eh of the rabbit cecum is not optimal for hydrogenotrophic methanogenesis, which requires -238 mV (Cord-Ruwisch et al., 1988). Indeed, CH_4 production in a *Methanothermobacter*

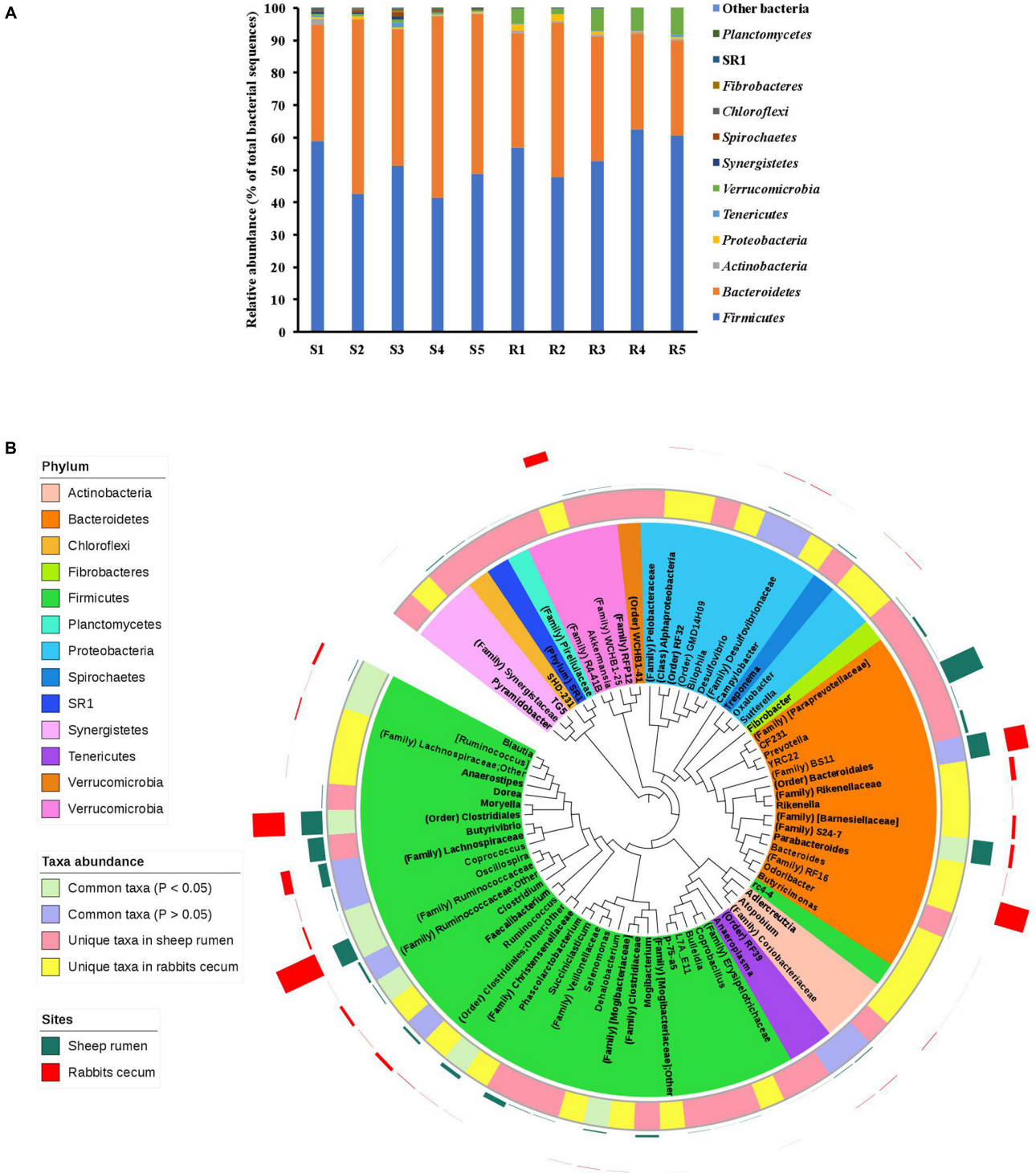


FIGURE 5 | Comparison of the bacterial microbiota in the sheep rumen and the rabbit cecum. **(A)** The relative abundance of bacterial phyla; **(B)** A neighbor-joining phylogenetic tree of the representative sequences of the OTUs that were graphically displayed in the Interactive Tree of Life. S1–S5, the five sheep; R1–R5, the five blocks of rabbits.

thermautotrophicus culture was suppressed at Eh higher than −200 mV (Hirano et al., 2013). Future research using *in vitro* cultures of both digestive organs is warranted to verify if Eh is a primary factor determining the different CH_4 production in the rumen and the cecum. The Eh of *in vitro* cultures can be regulated using bioelectrochemical systems that can control Eh

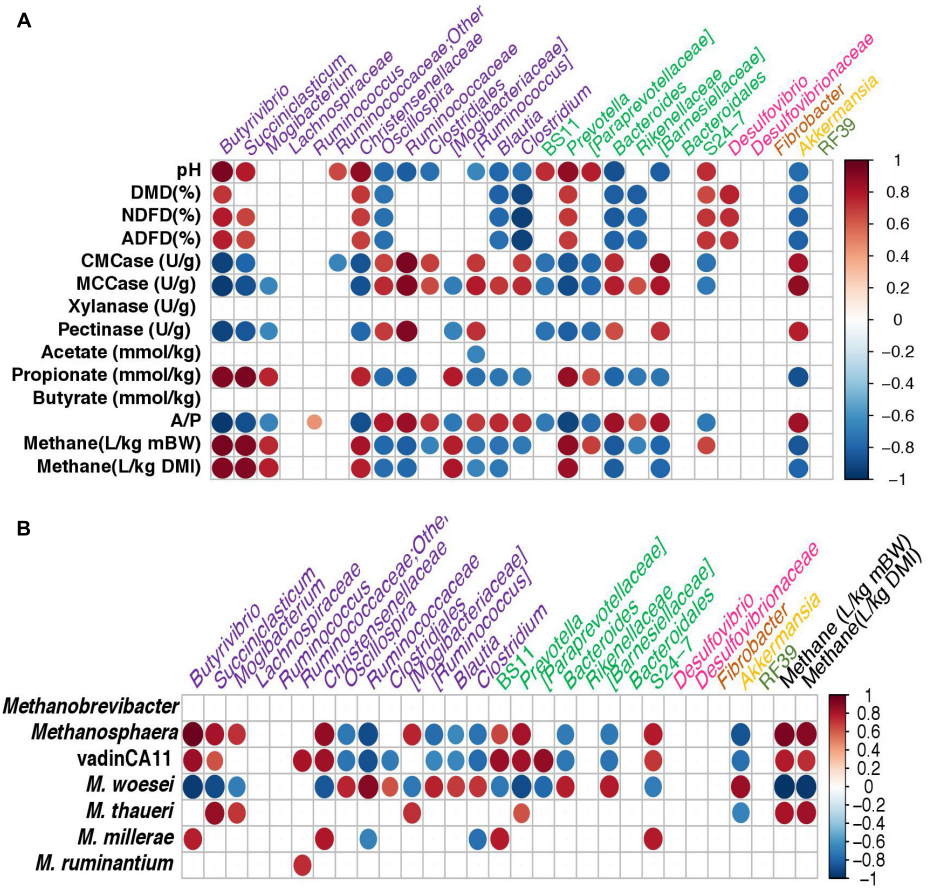


FIGURE 6 | Pearson correlation between **(A)** chemical parameters and the relative abundance of bacterial taxa; **(B)** Archaeal taxa and bacterial taxa in relative abundance for the sheep rumen (n = 5) and the rabbit cecum (n = 5 blocks) in combination. Taxa were included in the matrix only if they were found in at least 3 of the five sheep or five rabbit blocks and they were each represented by at least 1.0% of the total bacterial sequences. Size of the filled circle reflects the strength of the correlations. Only significant correlation ($P < 0.05$) were shown with colors. The scale colors denote the direction of correlation (1 to −1). The taxa assigned to different phyla were color coded: purple, *Firmicutes;* green, *Bacteroidetes;* pink, *Proteobacteria;* brown, *Fibrobacteres;* yellow, *Verrucomicrobia;* and olive, *Tenericutes*.

without using oxidative and reducing agents (Hirano et al., 2013). Furthermore, digesta passage rate through the rumen has been found reversely correlated with CH_4 production therein (Janssen, 2010; Goopy et al., 2014; Shi et al., 2014). The cecum is a tubiform tract, while the rumen is a large sac. Such structural difference can causes a faster digesta passage rate through the rabbit cecum than through the sheep rumen, contributing to the less CH_4 output from the rabbits than from the sheep. This premise is consistent with the high passage rate and low CH_4 production by kangaroos, a group of tubiform foregut fermenters (Von Engelhardt et al., 1978).

In the present study, we analyzed the diversity and structure of the archaeal microbiota and quantified the abundance of methanogens to understand the archaeal underpinning of the different CH_4 yields between the two digestive organs. The rabbit cecum had a lower abundance of RCC methanogens, non-RCC methanogens, and total methanogens (as quantified as *mcr*A gene copies/g sample) than the sheep rumen. This is consistent with the finding in sheep (Popova et al., 2013), reindeer (Wedlock et al., 2013), and Chinese roe deer (Li et al., 2014), in which a greater abundance of methanogens was found in the rumen than in the cecum. The low abundance of methanogens in the rabbit cecum may be explained partially by the low pH, probably a higher *Eh*, and a greater passage. These three factors might have directly decreased methanogenesis in the rabbit cecum. Although the abundance of methanogens in the rumen does not necessarily linearly correlate to CH_4 output (Danielsson et al., 2012; Patra and Yu, 2014), the greater abundance of methanogens in the sheep rumen than in the rabbit cecum corroborates the more CH_4 produced by the former than by the latter.

The two digestive organs each harbored a distinct archaeal microbiota, with *M. woesei* being the dominant species in the rabbit cecum, whereas the sheep rumen containing *M. thaueri* as the most predominant known species followed by *M. millerae* and *M. woesei*. Based on a literature search of the Pubmed, only one study has analyzed the archaeal microbiota in the cecum of rabbits, and *M. woesei* was represented by more

cloned 16S rRNA gene sequences than other species (Kušar and Avguštin, 2010). *M. ruminantium*, one of the two species (*M. ruminantium* and *M. Olleyae*) in the *Methanobrevibacter* RO clade (Janssen and Kirs, 2008; Kittelmann et al., 2013), was found in both the digestive organs at low relative abundance and low correlation with CH_4 yield. The RO clade was found associated with low CH_4 yield in the rumen (Danielsson et al., 2012). The dominance of *M. woesei*, which has only been reported in the chicken cecum (Saengkerdsub et al., 2007) other than rabbit cecum, is of further research interest. It is also interesting to note that *M. thaueri* and *M. millerae*, two of the four species (*M. smithii*, *M. gottschalkii*, *M. Millerae*, and *M. thaueri*) in the *Methanobrevibacter* SGMT clade (Janssen and Kirs, 2008; Kittelmann et al., 2013), were more predominant in the sheep rumen than in the rabbit cecum. The *Methanobrevibacter* SGMT clade, which possesses methyl coenzyme M reductase isozymes Mcr I and Mcr II and are competitive at high hydrogen concentrations (Leahy et al., 2010), has been reported to have a positive association with CH_4 emissions from ruminants (Tapio et al., 2017). The positive correlation between *M. thaueri* and CH_4 yield was found in the present study. It is not known if the differential predominance of these methanogen species is one explanation of the different CH_4 production seen between the two animal species. Several studies revealed a strong correlation between CH_4 yields and archaea: bacteria ratio (Wallace et al., 2014), and a similar finding was found in the present study. However, the previous study indicated that it was gene expression rather than gene abundance of methanogens that was strongly correlated with CH_4 yields from sheep (Shi et al., 2014). Metatranscriptomic studies will help determine the contribution of each methanogen species to the overall CH_4 yield in these two digestive organs.

Methane is produced by methanogens, but other members of the microbiota can determine or profoundly affects the rate and yield of methanogenesis (Kittelmann et al., 2014; Danielsson, 2016). In the present study, we characterized the bacterial microbiota for diversity and structure and quantified the population of microbes that produce hydrogen, including anaerobic fungi, protozoa, select hydrogen-producing bacteria, and acetogens to help understand the role and significance of these microbes in determining the different CH_4 yields between the two digestive organs. Expectedly, the sheep rumen and the rabbit cecum differed in the communities of the above microbes. Different microbiota were also reported between the rumen and the cecum of growing bulls (Popova et al., 2017), reindeer (Wedlock et al., 2013), and Chinese roe deer (Li et al., 2014). Such difference may be attributable to the combined effect of a host of factors, including pH, saliva (present in the rumen but not in cecum), passage rate, nutrients (lack of non-structural carbohydrates and protein due to digestion and absorption in the foregut), mixing, Eh (higher in the cecum), mucosa and antimicrobial peptides (both present in the rabbit cecum but not in the sheep rumen).

Hydrogen-producing microbes provide the reducing power for hydrogenotrophic methanogenesis, and indeed, sheep producing more hydrogen also produced more CH_4 than those that produced less hydrogen (Kittelmann et al., 2014). As determined by qPCR, the sheep rumen did have a greater abundance of *R. albus, R. flavefaciens*, *B. fibrisolvens*, fungi, and protozoa, all of which can produce hydrogen during feed fermentation, than the rabbit cecum. Combined analysis of the bacterial microbiota also revealed correlations between several animal phenotypic measurements and individual bacterial groups. Among the bacterial genera whose relative abundance was strongly and positively correlated with CH_4 yield, *Prevotella* and related bacteria, *Butyrivibrio*, and *Succiniclasticum* are unique and/or predominant hydrogen-producing bacteria in the rumen (Leahy et al., 2013). The bacterial genera whose relative abundance appeared to be negatively correlated with CH_4 yield are either acetogens (e.g., *Blautia*) (Müller and Frerichs, 2013), butyrate producers (e.g., *Oscillospira*) (Gophna et al., 2017), or succinate producers (e.g., *Bacteroides*) (Song et al., 2015). In the rumens of high methane-emitting sheep, members of *Ruminococcaceae* and *Lachnospiraceae* were found at higher relative abundance, while the rumens of low methane-emitting sheep were enriched with *Erysipelotrichaceae*, especially *Sharpea* spp. (Kamke et al., 2016). However, the low correlation between CH_4 yield and the relative abundance of *Ruminococcus* and *Lachnospiraceae* were found when comparing the two different the digestive organs in the present study. The relationship between the abundance of *F. succinogenes* and CH_4 production cannot be explained also, nor the contribution of *Clostridium* cluster IV. It should be noted that the correlation of a few bacterial genera with CH_4 yield might be due to their occurrence in only one of the two digestive organs, such as *Mogibacterium* and *Succiniclasticum* that were only detected in the rumen, and *Akkermansia* that was only detected in the large intestines. Nevertheless, the positive correlation between CH_4 yield and the relative abundance of *Mogibacterium*, which does not ferment carbohydrate (Nakazawa et al., 2015), is consistent with its ability to produce phenylacetate, a metabolite that is needed for the degradation of cellulose by some *R. albus* strains (Morrison et al., 1990).

Constant hydrogen disposal is essential for sustained fermentation in the rumen and the large intestines (Moss et al., 2010). Thus, alternative hydrogen utilization pathways must exist in the rabbit cecum. We analyzed two genes, *fhs* and *frd*, involved in two different [H]-utilizing pathways to understand the alternative hydrogen utilization potential in the two digestive organs. The rabbit cecum had a higher abundance of *fhs*, which encodes formyltetrahydrofolate synthetase, a key enzyme in the homoacetogenesis, consistent with the lower CH_4 production and higher molar proportion of acetate therein. A strong positive correlation was also found between A: P ratio and some bacterial taxa, including some taxa of *Clostridiales*, *Lachnospiraceae*, *Ruminococcaceae*, and *Blautia*, all of which contain known acetogens (Yang et al., 2016). All these results suggest acetate is not just a hydrogen donor, but a hydrogen sink. Indeed, homoacetogenesis might be predominant in the rabbit cecum than in the sheep rumen. The dominance of homoacetogenesis has been reported in some tubiform gut ecosystems, such as rabbits cecum (Yang et al., 2016), kangaroos foregut (Gagen et al., 2010), and termite hindgut

(Ottesen and Leadbetter, 2011). It has been suggested that pH might determine the predominance of hydrogen disposal pathways, with relatively neutral pH favoring methanogens and acidic pH favoring acetogens (Gibson et al., 1990). The relatively lower pH in the cecum of rabbits probably suppress hydrogenotrophic methanogenesis, allowing homocaetogenesis to increase. It is interesting to note that more CO_2 was produced from the rabbits than from the sheep per kg of $BW^{0.75}$. It is likely that less CO_2 is consumed during homoacetogenesis in the cecum than during methanogenesis in the sheep rumen. This premise is consistent with the less thermodynamic feature of homoacetogenesis than hydrogenotrophic methanogenesis and the higher *Ks* of the former (Kohn and Boston, 2000). The lower pH in the rabbit cecum could also decrease the CO_2 solubility and thus more CO_2 emission.

Surprisingly, significantly greater activities of CMCase, MCCasse, and pectinase were detected in the rabbit cecum than in the sheep rumen, suggesting an enrichment of fibrolytic and pectinolytic microbes in the rabbit cecum. Along with the significantly different bacterial microbiota structures between the sheep rumen and the rabbit cecum, our findings indicate that rabbit cecum probably harbors novel and uncharacterized cellulolytic bacteria and glycoside hydrolases. These novel microbes and enzymes can be identified in future studies using functional metagenomics and transcriptomics.

CONCLUSION

The present study demonstrates that different methane production between the sheep and the rabbits can be explained by the different physiological environments of their respective digestive organs and the microbiota residing therein. Lower abundance of hydrogen-producing microbes (bacteria, fungi, and protozoa) and methanogens, and increased homoacetogenesis as an alternative hydrogen utilization pathway in the rabbit cecum might result in lower CH_4 yield from the rabbits. The cecum of rabbits is potentially a rich resource to fibrolytic bacteria and hence novel cellulolytic enzymes. Future studies using functional approaches, such as functional metagenomics and transcriptomics, will help reveal the potential and functionality of metabolic pathways involved in fiber digestion, methanogenesis, and acetogenesis and help develop new strategies to achieve effective CH_4 mitigation for ruminal livestock.

AUTHOR CONTRIBUTIONS

JW, LM, and JL conceived and designed the study. LM performed both animal feeding and laboratory experiments, analyzed the sequencing data, interpreted the data, prepared the figures and tables, and wrote the manuscript. ZY and BY helped analyzing the sequencing data. XH and YL participated in the animal feeding experiments. ZY and JW helped interpret the data and write and revise the paper. All authors read and approved the final manuscript.

FUNDING

This research was partially supported by the National Natural Science Foundation of China (award number: 31372337). LM's tenure at The Ohio State University was supported by a scholarship from Zhejiang University.

ACKNOWLEDGMENTS

The authors thank Bo He, Chunlei Yang, Dan Feng, and Shanshan Wang for their assistance with feeding and caring of the animals and sample collection. The authors are also grateful to Dr. Lingling Wang for assistance in sequence data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00575/full#supplementary-material

REFERENCES

Aronesty, E. (2011). *ea-utils: Command-line Tools for Processing Biological Sequencing Data*. Available at: https://github.com/ExpressionAnalysis/ea-utils

Attwood, G., and McSweeney, C. (2008). Methanogen genomics to discover targets for methane mitigation technologies and options for alternative H2 utilisation in the rumen. *Aust. J. Exp. Agric.* 48, 28–37. doi: 10.1071/EA07203

Bailey, M. J., Biely, P., and Poutanen, K. (1992). Interlaboratory testing of methods for assay of xylanase activity. *J. Biotechnol.* 23, 257–270. doi: 10.1016/0168-1656(92)90074-J

Bhatta, R., Enishi, O., and Kurihara, M. (2007). Measurement of methane production from ruminants. *Asian Aust. J. Anim. Sci.* 20, 1305–1318. doi: 10.5713/ajas.2007.1305

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Cabezas Garcia, E. H. (2017). *Methane Production in Dairy Cows*. Doctoral thesis, Swedish University of Agricultural Sciences, Umeå.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Cord-Ruwisch, R., Seitz, H. J., and Conrad, R. (1988). The capacity of hydrogenotrophic anaerobic bacteria to compete for traces of hydrogen depends on the redox potential of the terminal electron acceptor. *Arch. Microbiol.* 149, 350–357. doi: 10.1007/BF00411655

Danielsson, R. (2016). *Methane Production in Dairy Cows*. Doctoral thesis, Swedish University of Agricultural Sciences, Uppsala.

Danielsson, R., Schnürer, A., Arthurson, V., and Bertilsson, J. (2012). Methanogenic population and CH4 production in Swedish dairy cows fed different levels of forage. *Appl. Environ. Microbiol.* 78, 6172–6179. doi: 10.1128/AEM.00675-12

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

Franz, R., Soliva, C. R., Kreuzer, M., Hummel, J., and Clauss, M. (2011). Methane output of rabbits (*Oryctolagus cuniculus*) and guinea pigs (*Cavia porcellus*) fed a hay-only diet: implications for the scaling of methane production with body mass in non-ruminant mammalian herbivores. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 158, 177–181. doi: 10.1016/j.cbpa.2010.10.019

Franz, R., Soliva, C. R., Kreuzer, M., Steuer, P., Hummel, J., and Clauss, M. (2010). Methane production in relation to body mass of ruminants and equids. *Evol. Ecol. Res.* 12, 727–738.

Furness, J. B., Cottrell, J. J., and Bravo, D. M. (2015). Comparative gut physiology symposium: comparative physiology of digestion. *J. Anim. Sci.* 93, 485–491. doi: 10.2527/jas.2014-8481

Gagen, E. J., Denman, S. E., Padmanabha, J., Zadbuke, S., Al Jassim, R., Morrison, M., et al. (2010). Functional gene analysis suggests different acetogen populations in the bovine rumen and tammar wallaby forestomach. *Appl. Environ. Microbiol.* 76, 7785–7795. doi: 10.1128/AEM.01679-10

Gibson, G. R., Cummings, J. H., Macfarlane, G. T., Allison, C., Segal, I., Vorster, H. H., et al. (1990). Alternative pathways for hydrogen disposal during fermentation in the human colon. *Gut* 31, 679–683. doi: 10.1136/gut.31.6.679

Goopy, J. P., Donaldson, A., Hegarty, R., Vercoe, P. E., Haynes, F., Barnett, M., et al. (2014). Low-methane yield sheep have smaller rumens and shorter rumen retention time. *Br. J. Nutr.* 111, 578–585. doi: 10.1017/S0007114513002936

Gophna, U., Konikoff, T., and Nielsen, H. B. (2017). Oscillospira and related bacteria–From metagenomic species to metabolic features. *Environ. Microbiol.* 19, 835–841. doi: 10.1111/1462-2920.13658

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi: 10.1101/gr.112730.110

Hirano, S., Matsumoto, N., Morita, M., Sasaki, K., and Ohmura, N. (2013). Electrochemical control of redox potential affects methanogenesis of the hydrogenotrophic methanogen *Methanothermobacter thermautotrophicus*. *Lett. Appl. Microbiol.* 56, 315–321. doi: 10.1111/lam.12059

IPCC (2014). *IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of Intergovernmental Panel on Climate Change* [Core Writing Team, eds R. K. Pachauri and L. A. Meyer]. Geneva: Geneva Press, 151.

Janssen, P. H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *Anim. Feed Sci. Technol.* 160, 1–22. doi: 10.1016/j.anifeedsci.2010.07.002

Janssen, P. H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* 74, 3619–3625. doi: 10.1128/AEM.02812-07

Joblin, K. N. (1999). Ruminal acetogens and their potential to lower ruminant methane emissions. *Aust. J. Agr. Res.* 50, 1307–1314. doi: 10.1071/AR99004

Johnson, K. A., and Johnson, D. E. (1995). Methane emissions from cattle. *J. Anim. Sci.* 73, 2483–2492. doi: 10.2527/1995.7382483x

Kamke, J., Sandra, K., Priya, S., Yang, L., Michael, T., Siva, G., et al. (2016). Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a sharpea-enriched microbiome characterised by lactic acid formation and utilisation. *Microbiome* 4:56. doi: 10.1186/s40168-016-0201-2

Kimsé, M., Monteils, V., Bayourthe, C., and Gidenne, T. (2010). A new method to measure the redox potential (Eh) in rabbit caecum: relationship with pH and fermentation pattern. *World Rabbit Sci.* 17, 63–70. doi: 10.4995/wrs.2009.659

Kittelmann, S., Pinares-Patiño, C. S., Seedorf, H., Kirk, M. R., Ganesh, S., McEwan, J. C., et al. (2014). Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS One* 9:e103171. doi: 10.1371/journal.pone.0103171

Kittelmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I., et al. (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* 8:e47879. doi: 10.1371/journal.pone.0047879

Kohn, R. A., and Boston, R. C. (2000). "The role of thermodynamics in controlling rumen metabolism," in *Modelling Nutrient Utilization in Farm Animals*, ed. J. P. McNamara (Wallingford, OX: CABI Press), 11–24.

Koike, S., Yabuki, H., and Kobayashi, Y. (2007). Validation and application of real-time polymerase chain reaction assays for representative rumen bacteria. *J. Anim. Sci.* 78, 135–141. doi: 10.1111/j.1740-0929.2007.00417.x

Kušar, D., and Avguštin, G. (2010). Molecular profiling and identification of methanogenic archaeal species from rabbit caecum. *FEMS Microbiol. Ecol.* 74, 623–630. doi: 10.1111/j.1574-6941.2010.00980.x

Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., et al. (2010). The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLoS One* 5:e8926. doi: 10.1371/journal.pone.0008926

Leahy, S. C., Kelly, W. J., Ronimus, R. S., Wedlock, N., Altermann, E., and Attwood, G. T. (2013). Genome sequencing of rumen bacteria and archaea and its application to methane mitigation strategies. *Animal* 7, 235–243. doi: 10.1017/S1751731113000700

Li, Z., Zhang, Z., Xu, C., Zhao, J., Liu, H., Fan, Z., et al. (2014). Bacteria and methanogens differ along the gastrointestinal tract of Chinese roe deer (*Capreolus pygargus*). *PLoS One* 9:e114513. doi: 10.1371/journal.pone.0114513

Liu, J., Wang, J. K., Zhu, W., Pu, Y. Y., Guan, L. L., and Liu, J. X. (2014). Monitoring the rumen pectinolytic bacteria *Treponema saccharophilum* using real-time PCR. *FEMS Microbiol. Ecol.* 87, 576–585. doi: 10.1111/1574-6941.12246

Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* 1125, 171–189. doi: 10.1196/annals.1419.019

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Makkar, H. P. S., Sharma, O. P., Dawra, R. K., and Negi, S. S. (1982). Simple determination of microbial protein in rumen liquor. *J. Dairy Sci.* 65, 2170–2173. doi: 10.3168/jds.S0022-0302(82)82477-6

Mathieu, F., Jouany, J. P., Senaud, J., Bohatier, J., Bertin, G., and Mercier, M. (1996). The effect of Saccharomyces cerevisiae and *Aspergillus oryzae* on fermentations in the rumen of faunated and defaunated sheep; protozoal and probiotic interactions. *Reprod. Nutr. Dev.* 36, 271–287. doi: 10.1051/rnd:19960305

McDougall, E. I. (1948). Studies on ruminant saliva. 1. the composition and output of sheep's saliva. *Biochem. J.* 43, 99–109. doi: 10.1042/bj0430099

Michelland, R. J., Combes, S., Monteils, V., Cauquil, L., Gidenne, T., and Fortun-Lamothe, L. (2010). Molecular analysis of the bacterial community in digestive tract of rabbit. *Anaerobe* 16, 61–65. doi: 10.1016/j.anaerobe.2009.05.002

Morrison, M., Mackie, R. I., and Kistner, A. (1990). 3-Phenylpropanoic acid improves the affinity of *Ruminococcus albus* for cellulose in continuous culture. *Appl. Environ. Microbiol.* 56, 3220–3222.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature* 463, 747–756. doi: 10.1038/nature08823

Müller, V., and Frerichs, J. (2013). "Acetogenic bacteria," in *eLS*, ed. J. Battista (Chichecster, NJ: John Wiley and Sons Ltd.). doi: 10.1002/9780470015902.a0020086.pub2

Nakazawa, F., Sato, M., Poco, S. E., Hashimura, T., Ideka, T., Kalfas, S., et al. (2015). "Mogibacterium," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. B. W. William (New York, NY: The Editorial Board Press), 1–5.

Official Methods of Analysis [AOAC] (2015). *Official Methods of Analysis*. Gaithersburg, MD: Association of Official Analytical Chemists Press.

Ottesen, E. A., and Leadbetter, J. R. (2011). Formyltetrahydrofolate synthetase gene diversity in the guts of higher termites with different diets and lifestyles. *Appl. Environ. Microbiol.* 77, 3461–3467. doi: 10.1128/AEM.02657-10

Patra, A. K., and Yu, Z. T. (2014). Combinations of nitrate, saponin, and sulfate additively reduce methane production by rumen cultures in vitro while not adversely affecting feed digestion, fermentation or microbial communities. *Bioresour. Technol.* 155, 129–135. doi: 10.1016/j.biortech.2013.12.099

Piattoni, F., Demeyer, D., and Maertens, L. (1996). In vitro study of the age-dependent caecal fermentation pattern and methanogenesis in young rabbits. *Reprod. Nutr. Dev.* 36, 253–261. doi: 10.1051/rnd:19960303

Popova, M., McGovern, E., McCabe, M. S., Martin, C., Doreau, M., Arbre, M., et al. (2017). The structural and functional capacity of ruminal and cecal microbiota in growing cattle was unaffected by dietary supplementation of linseed oil and nitrate. *Front. Microbiol.* 8:937. doi: 10.3389/fmicb.2017.00937

Popova, M., Morgavi, D. P., and Martin, C. (2013). Methanogens and methanogenesis in the rumens and ceca of lambs fed two different high-grain-content diets. *Appl. Environ. Microbiol.* 79, 1777–1786. doi: 10.1128/AEM.03115-12

Poulsen, M., Schwab, C., Jensen, B. B., Engberg, R. M., Spang, A., Canibe, N., et al. (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat. Commun.* 4:1428. doi: 10.1038/ncomms2432

Russell, J. B. (1998). The importance of pH in the regulation of ruminal acetate to propionate ratio and methane production in vitro. *J. Dairy Sci.* 81, 3222–3230. doi: 10.3168/jds.S0022-0302(98)75886-2

Russell, J. B., and Rychlik, J. L. (2001). Factors that alter rumen microbial ecology. *Science* 292, 1119–1122. doi: 10.1126/science.1058830

Saengkerdsub, S., Anderson, R. C., Wilkinson, H. H., Kim, W. K., Nisbet, D. J., and Ricke, S. C. (2007). Identification and quantification of methanogenic archaea in adult chicken ceca. *Appl. Environ. Microbiol.* 73, 353–356. doi: 10.1128/AEM.01931-06

SAS Institute Inc. (2000). *SAS User's Guide: Statistics. Version 9.2*. (Cary, NC: SAS Institute Inc.)

Shi, W., Moon, C. D., Leahy, S. C., Kang, D., Froula, J., Kittelmann, S., et al. (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* 24, 1517–1525. doi: 10.1101/gr.168245.113

Song, Y., Liu, C., and Finegold, S. M. (2015). "Bacteroides," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. B. W. William (New York, NY: The Editorial Board Press), 1–24.

St-Pierre, B., and Wright, A. D. G. (2013). Diversity of gut methanogens in herbivorous animals. *Animal* 7, 49–56. doi: 10.1017/S1751731112000912

Tapio, I., Snelling, T. J., Strozzi, F., and Wallace, R. J. (2017). The ruminal microbiome associated with methane emissions from ruminant livestock. *J. Anim. Sci. Biotechnol.* 8:7. doi: 10.1186/s40104-017-0141-0

Van Soest, P. V., Robertson, J. B., and Lewis, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J. Dairy Sci.* 74, 3583–3597. doi: 10.3168/jds.S0022-0302(91)78551-2

Von Engelhardt, W., Wolter, S., Lawrenz, H., and Hemsley, J. A. (1978). Production of methane in two non-ruminant herbivores. *Comp. Biochem. Physiol. Part A Physiol.* 60, 309–311. doi: 10.1016/0300-9629(78)90254-2

Wallace, R. J., Rooke, J. A., Duthie, C. A., Hyslop, J. J., Ross, D. W., McKain, N., et al. (2014). Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci. Rep.* 4:5892. doi: 10.1038/srep05892

Wallace, R. J., Rooke, J. A., McKain, N., Duthie, C. A., Hyslop, J. J., Ross, D. W., et al. (2015). The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16:839. doi: 10.1186/s12864-015-2032-0

Wang, J. K., He, B., Du, W., Luo, Y., Yu, Z. T., and Liu, J. X. (2015). Yeast with surface displayed xylanase as a new dual purpose delivery vehicle of xylanase and yeast. *Anim. Feed Sci. Technol.* 208, 44–52. doi: 10.1016/j.anifeedsci.2015.07.002

Wedlock, D. N., Janssen, P. H., Leahy, S. C., Shu, D., and Buddle, B. M. (2013). Progress in the development of vaccines against rumen methanogens. *Animal* 7, 244–252. doi: 10.1017/S1751731113000682

Yang, B., He, B., Wang, S. S., Liu, J. X., and Wang, J. K. (2015). Early supplementation of starter pellets with alfalfa improves the performance of pre- and postweaning Hu lambs. *J. Anim. Sci.* 93, 4984–4994. doi: 10.2527/jas.2015-9266

Yang, C. L., Mi, L., Hu, X. L., Liu, J. X., and Wang, J. K. (2016). Investigation into host selection of the cecal acetogen population in rabbits after weaning. *PLoS One* 11:e0158768. doi: 10.1371/journal.pone.0158768

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Mi, Yang, Hu, Luo, Liu, Yu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 October 2017
doi: 10.3389/fmicb.2017.01997

Amplicon-Based Sequencing of Soil Fungi from Wood Preservative Test Sites

Grant T. Kirker[1], Amy B. Bishell[1], Michelle A. Jusino[2], Jonathan M. Palmer[2], William J. Hickey[3] and Daniel L. Lindner[2]*

[1] FPL, United States Department of Agriculture-Forest Service (USDA-FS), Durability and Wood Protection, Madison, WI, United States, [2] NRS, United States Department of Agriculture-Forest Service (USDA-FS), Center for Forest Mycology Research, Madison, WI, United States, [3] Department of Soil Science, University of Wisconsin-Madison, Madison, WI, United States

Soil samples were collected from field sites in two AWPA (American Wood Protection Association) wood decay hazard zones in North America. Two field plots at each site were exposed to differing preservative chemistries via in-ground installations of treated wood stakes for approximately 50 years. The purpose of this study is to characterize soil fungal species and to determine if long term exposure to various wood preservatives impacts soil fungal community composition. Soil fungal communities were compared using amplicon-based DNA sequencing of the internal transcribed spacer 1 (ITS1) region of the rDNA array. Data show that soil fungal community composition differs significantly between the two sites and that long-term exposure to different preservative chemistries is correlated with different species composition of soil fungi. However, chemical analyses using ICP-OES found levels of select residual preservative actives (copper, chromium and arsenic) to be similar to naturally occurring levels in unexposed areas. A list of indicator species was compiled for each treatment-site combination; functional guild analyses indicate that long-term exposure to wood preservatives may have both detrimental and stimulatory effects on soil fungal species composition. Fungi with demonstrated capacity to degrade industrial pollutants were found to be highly correlated with areas that experienced long-term exposure to preservative testing.

Keywords: amplicon sequencing, environmental impacts, DNA, soil fungal communities, wood decay fungi, wood preservatives

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Seung Gu Shin, Pohang University of Science and Technology, South Korea
Stefan Fränzle, Technische Universität Dresden, Germany

****Correspondence:***
Grant T. Kirker
gkirker@fs.fed.us

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *26 April 2017*
Accepted: *28 September 2017*
Published: *18 October 2017*

Citation:
Kirker GT, Bishell AB, Jusino MA, Palmer JM, Hickey WJ and Lindner DL (2017) Amplicon-Based Sequencing of Soil Fungi from Wood Preservative Test Sites. Front. Microbiol. 8:1997. doi: 10.3389/fmicb.2017.01997

INTRODUCTION

Wood preservation as an industry has a long-standing history of using persistent chemicals to prevent colonization of treated wood by decay fungi and insects. Chemically treated wood in contact with soil has the potential to expose the surrounding soil and its inhabitants to these chemicals through chemical migration. Previous studies have shown that wood preservatives can migrate from treated wood, but environmental exposure is typically very low and varies based on preservative chemistry, environmental exposure and soil composition (Lebow et al., 2004). The effects of prolonged wood preservation chemical exposure on soil fungal communities is not fully understood (Bhattacharya et al., 2002). Soil fungal communities are key drivers of soil geochemical processes and assist in cycling of carbon, nitrogen, and other nutrients (Tedersoo et al., 2014). Classical biodiversity studies based on fungal fruiting body surveys are limited, as uncultivable fungi can constitute a major proportion of the total micro-biota of a given location (Gams, 2007). Thus,

DNA-based identification of fungal communities has been a valuable tool to assess fungal diversity (Schoch et al., 2012). Coupling DNA-based identification methods and next-generation sequencing has proven to be a powerful tool to identify fungi within environmental samples and subsequently compare these samples under different experimental conditions (Daniel, 2005).

Previous studies have found that soil fungal diversity is in fact very homogeneous in deeper soil horizons, consisting mostly of ectomycorrhizal and other root associated fungi (Dickie et al., 2002). However, the litter layer, found in the first few inches of soil strata, holds the key components of soil metabolic capabilities (Voriskova and Baldrian, 2013). The rate at which biomass is turned over through biochemical breakdown and subsequent nutrient cycling has been found to vary considerably with respect to soil fungal diversity and structure (Abrego and Salcedo, 2013). Therefore, the effects of wood preservative migration into surrounding soil communities has the potential to impact fungal species with varying nutrient cycling strategies.

Environmental impacts of wood preservatives are a major concern of the industry; end users have been a driving force behind the ever changing spectrum of preservatives used to protect wood in service (Schultz et al., 2007). Early preservative systems relied on broad spectrum toxins that restricted the growth of deteriorative agents with little or no consideration of the impacts of these compounds on the surrounding environment (Connell, 1991). More recent preservative systems have been more targeted in an attempt to minimize broad environmental impacts (Schultz and Nicholas, 2002). First generation preservation systems consisted of mostly creosote, some highly toxic metals such as mercury, cadmium and arsenic and relied on broad spectrum control of microbes, including bacteria, which have been shown to contribute to wood permeability and preservative breakdown (Schultz et al., 2007). Chemical migration of wood preservatives in use has also been a driver of change based on consumer concerns, which has led to some voluntary restrictions on their use by the wood preservation industry (Hingston et al., 2001). The potential negative health effects of these toxic metal compounds greatly outweighed their efficacy as preservatives; mercury and cadmium are both highly water soluble and arsenic in high dosages is toxic to most life forms (Hughes, 2002). Newer preservative formulations have much reduced risks associated with long-term exposure and consist of copper combined with an organic co-biocide. These have been shown to have good activity against most wood decay fungi (Freeman and Mcintyre, 2008). A persistent problem that can severely impact the service life of copper treated lumber is the presence of copper tolerant fungi, which often arise either due to genetic predisposition of fungi to detoxify copper or as a result of long-term exposure to a copper rich environment (Clausen and Green, 2003). Several known wood decay fungi are classified as copper tolerant and have been studied in order to understand mechanisms of copper tolerance (Green and Clausen, 2005).

While there is substantial literature depicting changes in soil fungal biota due to metal contaminates (Kandeler et al., 1996; Giller et al., 1998; Bhattacharya et al., 2002; Rajapaksha et al., 2004; Oliveira and Pampulha, 2006), little is known about changes in soil fungal biodiversity as a specific result of prolonged wood preservative exposure and few studies have employed amplicon based sequencing to compare soil fungal communities under after exposure. Prior studies have shown that preservatives can migrate from wood in ground contact (Lebow et al., 2004), but these concentrations are much less than the amounts present in the wood and are typically only within 5–10 cm of the treated material. Kirker et al. (2012) found differences in the fungal diversity and colonization patterns of ammoniacal copper quaternary type C (ACQ-C) treated pine in field studies, but those differences became less pronounced after longer exposure (12 months). The primary objective of this study is to characterize soil fungi from sites exposed to different wood preservative types and to analyze differences in fungal species composition due to long-term preservative exposure. Simultaneously, the data generated will also serve as baseline data for continued monitoring of the soil fungal communities from different climates. A secondary objective of this study is to assess fungal diversity in copper exposed soils and compare to non-copper exposed soils to screen for presence of copper tolerant fungi.

MATERIALS AND METHODS

Field Sampling

Field sites are located in Saucier, Mississippi and Madison, Wisconsin, USA. The Saucier site is located in the Harrison Experimental Forest in southern Mississippi. This geographic region is classified as high decay hazard for wood in service (zone 5-severe) according to both the American Wood Protection Association (AWPA) (Association, 2007) and the Scheffer index (Scheffer, 1971). This soil region is classified as coastal plain and the dominant soil type is sandy clay with surrounding overstory being predominantly loblolly pine (*Pinus taeda L.*). Untreated southern pine sapwood will typically decay to the point of structural failure after 1 year in ground exposure in this environment. The Saucier, Mississippi test site has been utilized by the United States Department of Agriculture Forest Service (USDA-FS) Forest Products Laboratory (FPL) as a preservative test site since 1940 and contains many of the early testing sites used to develop long-lasting and effective wood protectants. Samples were obtained on April 10, 2014 from 3 test plots at the field site. The mixed plot was a 50 year-old plot containing chromated copper arsenate (CCA), creosote and pentachlorophenol treated posts. The copper plot was a 30 year-old plot containing wooden stakes treated with basic solutions of copper salts. The control samples were obtained from a nearby, unused plot of forest within the Harrison Experimental Forest. Four field replicate soil samples per site-treatment combination were transported to the laboratory and frozen at −30°C for 4 days before processing.

The Madison, WI site has been used by FPL since 1950 for routine testing of experimental and existing preservative systems in a moderate decay hazard (AWPA, zone 2). The soil type at Madison is darker and contains more silt and clay as compared to the Saucier site and is subject to more freeze thaw

cycles, in addition to a solid winter freeze. The Madison, WI site is mostly prairie/savannah type ecosystem with continuously decreasing overstory containing mostly mixed hardwoods [red maple (*Acer rubrum*), mixed oaks (*Quercus* spp.), and some black locust (*Robinia pseudoacacia*)]. In comparison to the Saucier site, untreated southern pine sapwood typically fails due to decay within 2 years at this more northern site. Samples were obtained on May 9, 2014 from 3 test plots at the Madison field site. The mixed plot was an approximately 60 year-old plot containing CCA, creosote and pentachlorophenol treated posts approximately 10 cm in diameter. The copper plot was a 40 year-old plot containing 60 cm (2 ft) × 120 cm (4 ft) wooden stakes treated with various copper solutions. The control samples were obtained from a nearby, unused plot of forest within the FPL test site approximately 1 km from the treated plots.

Samples were taken from the soil at both sites using 5 g Terra Core (En Novative Technologies, Dexter, MI, USA) disposable soil samplers. Samples were taken from four locations approximately 10 feet apart within each test plot and 3 plugs were combined within each treatment replicate. Four field replicate soil samples were transported to the laboratory and frozen at −30°C for 3 days before processing.

DNA Isolation and Amplicon Based Sequencing

A 0.25 g aliquot of each hand mixed sample was extracted using the MoBio Power Soil DNA Isolation Kit (Carlsbad, CA, USA) following manufacturer's instructions. The samples were eluted in 100 µl C6 elution buffer and cleaned using the MoBio Powerclean Pro DNA Clean-up Kit following manufacturer's instructions. Samples were eluted in 100 µl DC5 elution buffer then quantified by spectrophotometer and diluted to 10 ng/µl in Tris EDTA pH 8. DNA samples were amplified in triplicate using 25 ng template or water controls and ITS1F and ITS2 primers (De Gannes et al., 2013) with Illumina adapters for the MiSeq platform with 24 unique identifiers on the reverse primers. Phusion Hot Start Flex DNA Polymerase (New England Biolabs, Ipswich, MA, USA) in HF buffer was used according to manufacturer's instructions for PCR with the following program: 4 min at 94°C, followed by 30 cycles of 30 s at 94°C, 60 s at 50°C and 90 s at 72°C and a final extension of 10 min at 72°C. Check gels were run on one of each PCR replicate and 400–500 bp products were confirmed. The three PCR replicates were combined and cleaned using Agencort AMPure XP beads following manufacturer instructions. Each sample was quantified using the Quant-it DNA Assay Kit (high sensitivity, ThermoFisher, Waltham, MA, USA) following the microplate procedure with a Synergy H1 multimodal plate reader (Biotek, Winooski, VT, USA). All samples were then normalized to 10 nM and combined in equal amounts. This pooled sample was submitted to the University of Wisconsin-Madison Biotechnology Center—DNA Sequencing Facility for 250 paired Illumina MiSeq sequencing using paired reads of 250 base pairs.

Sequencing Data Analysis

Sequencing data were processed using the AMPtk v0.4.0 pipeline (Palmer, 2015 -https://github.com/nextgenusfs/amptk). Briefly, overlapping 2 × 250 base pair Illumina MiSeq reads were merged using USEARCH v8.1.1831 (Edgar and Flyvbjerg, 2015), forward and reverse primers were removed from the merged reads, and the reads were trimmed or padded with N's to a set length of 250 base pairs. Because ITS sequences are of variable length they require extra processing steps in comparison to 16S reads that are nearly identical in length. The average length of known ITS1 sequences in public databases is ∼250 bp; ranging in size from ∼150 bp up to >600 bp, thus we aren't able to recover *all* full length sequences with PE 2 × 250 bp sequencing. And it is recommended to use reads that are truncated to the same length when clustering with UPARSE pipeline, padding/trimming to 250 bp was used to maximize the number of reads passing the quality filters of AMPtk as well as provide enough information to assign taxonomy to OTUs. Processed reads were quality trimmed based on accumulation of expected errors less than 1.0 (Edgar and Flyvbjerg, 2015) and clustered using the UPARSE algorithm using default parameters (singletons removed, 97% OTU radius). An OTU table was generated by mapping the original reads to the OTUs using VSEARCH 1.9.1 (Rognes et al., 2016) and the OTU table was subsequently filtered to eliminate "index-bleed" between samples by setting read counts to zero if the number of reads mapped was less than 0.5% of the sum of all read counts for each OTU. Index-bleed (also called index-crossover, index-hopping, barcode-mismatch, etc.) is a phenomenon in NGS sequencing where a small number of reads are mis-assigned to the wrong sample group during multiplex sequencing (combining multiple samples on a single NGS sequencing run). Thus to filter out low read counts that could have arisen from sample mis-assignment (index-bleed), AMPtk provides a filter to clean an OTU table. Taxonomy was assigned using a combination of UTAX and global alignment (USEARCH Edgar, 2010) to the UNITE v7.0 database (Abarenkov et al., 2010) and non-fungal OTUs were removed prior to downstream data processing. BIOM data from AMPtk pipeline was further analyzed using METACOMET (Wang et al., 2016) and PHINCH (Bik and Pitch Interactive, 2014) to visually compare treatment and sites at each taxonomic level.

Fungal Community Analyses

All distance-matrix based community analyses were performed on a presence absence (binary) OTU matrix, using the Raup-Crick distance metric as calculated by the raupcrick function in the vegan package (Oksanen et al., 2015) of R (R Core Team, 2013). This metric is robust to common issues in large data sets such as an abundance of zeros, and variation in alpha diversity among samples (Chase et al., 2011). To visualize fungal communities, we performed ordinations using non-parametric multi-dimensional scaling (NMDS), implemented by the metaMDS function in the vegan package (Oksanen et al., 2015) of R (Team, 2016). Permutational multivariate analysis of variance (PERMANOVA) tests were used to test for site and treatment effects. PERMANOVA was calculated by the Adonis function in the vegan package (Oksanen et al., 2015).

We also tested for multivariate dispersion among groups using the betadisper function in the vegan package of R. Finally, we performed indicator species analyses to identify specific fungal OTUs associated with each treatment group at each site using the multipatt function in the indicspecies package (Cáceres and Legendre, 2009) in R.

Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES) Methods

Twenty four soil samples were dried at 105°C for 2 days, then immediately capped and transferred to a desiccator. Samples were weighed in a dry room to 4 decimal places and directly into a tared Teflon digestion vessel-Anton Paar (Ashland, VA) HVT50 vessels. Two milliliters of 70% HNO_3 (Sigma-Aldrich (Milwaukee, WI) ACS-grade) was added, using an acid-resistant bottle-top pipette, to each digestion vessel. Samples were pre-digested for 15 min. at room temperature. In a fume hood, 5 ml 18 MΩ H_2O was added, and further digested for 10 min, then vessels were sealed. Digested solutions were cooled, then filtered through glass microfiber filter (Whatman 934-AH) into a 50-ml volumetric flask and brought to volume using 18 MΩ H_2O.

A Horiba (Edison, NJ) ULTIMA II high resolution spectrometer was used for the ICP-OES analysis. The instrument was equipped with a solid state, water cooled 40 MHz radio frequency source, a Czerny Turner monochrometer with 1 meter focal length, nitrogen purged optical bench at 6 L/min, holographic grating (2400 grooves/mm) with resolution <5 pm in the 120–320 nm range (1st order) and <10 pm in the 320–800 nm range (2nd order), and a vertical torch with radial viewing. Power level was 1,000 W. Gas flows were: Plasma gas 12 L/min, sheath gas 0.2 L/min, nebulizer gas 0.28 L/min. Sample introduction used a Conical U-series concentric glass nebulizer (1 ml/min) and a Tracey spray chamber with helix connection and TruFlow sample monitor (Glass Expansion, Pocasset, MA).

Standards were prepared from 1,000 ppm stock solutions without serial dilution. Standard curves were obtained using external standards for Ca 396.847, Ca 422.673, Cr 205.552, Cr 267.716, Cu 224.700, Cu 324.750, Cu 327.396, K 766.455, Mg 279.079, Mg 280.270, Mg 285.213, and P178.229 nm. Each standard curve was checked using a separate calibration solution (Inorganic Ventures (Christiansburg, VA) ICP-MS 71A multi-element standard), then the best line chosen for quantitation of samples. The calibrations, each with a correlation coefficient $r^2 > 0.9999$, were checked at the start, middle and end of the ICP sequence, with pass falling within 10% of target. Ten replicate readings of a blank were also run at the start, middle and end of the ICP sequence, then used to determine Limit of Quantitation (LOQ). Acquisition was done in Max mode peak shape, with 0.5 s integration time and 3 replicate readings averaged for the reported measurement. A 20/15 um slit combination was used for all lines. The instrument electronics were turned on ∼30 min. prior to start in order to minimize instrumental drift. Prior to each measurement, the sample intake was rinsed 20 s with 5% HNO3 (v/v) at high pump speed, followed by a 20 s rinse with analyte solution at high pump speed, then a 20 s plasma equilibration period at normal pump speed. Prior to calibration, a profile acquisition was taken to visually inspect peak shape and manually set background correction points at peak base.

RESULTS

A total of 6,663 OTUs were recovered from the 24 soil samples. Thirty-one protist OTUs were omitted from the data set. A total of 3,110 OTUs were identified only to the Kingdom level and 3,448 OTUs were classified to the phylum level. A total of 1,881 OTUs were classified as Ascomycetes (Phylum = Ascomycota), 1,040 OTUs were classified as Basidiomycetes (Phylum = Basidiomycota), 147 were classified as Zygomycetes (Phylum = Zygomycota) and 233 were classified as Glomeromcyetes (Phylum = Glomeromycota), 4 OTUS were classified as Blastocladiomycota, which are recently split from the Chytridomycota (James et al., 2006), 78 OTUs were classified as Chytridomycota, and 55 OTUs were classified as Rozellomycota, which represent another lineage recently split from the Chytridomycota, These are typically amoeboid microfungi and are almost exclusively identified through environmental sequencing. The distribution of the different phyla recovered from the samples is shown in **Supplemental Figure 1**.

A total of 25 fungal classes were identified from the soil samples. The number of representative OTUs of the class Archeorhizomycetes was noticeably higher in several of the preservative exposed samples. In addition to several unidentified classes, representatives of class Agaricomycetes, Archeorhizomycetes, Sordariomycetes, Eurotiomyctes, Leotiomycetes, and Dothidiomycetes were all abundant in our sampling. Their relative abundances and distributions across the samples are shown in **Supplemental Figure 2**.

A total of 55 fungal orders were classified. Order Agaricales was abundant in our sampling and widespread across both site and treatment. Order Mortierellales was also abundant and widespread, although more abundant at the MS site. Order Russulales exhibited a patchy distribution and was found in unexposed, mixed treatments and copper treated plots intermittently and much less prevalent at the WI site. Both within the class Sordariomycetidae, Order Hypocreales was more abundant in MS while order Sordariales was more prevalent in WI. Both of these classes contain ascomycete fungi typically associated with soft rot as well as several important plant pathogens. Order Thelophorales was more prevalent in MS but did not appear to be affected by treatment. Relative abundances and distributions of the orders are shown in **Supplemental Figure 3**.

A total of 206 fungal families were classified. The 10 most common families were Mortierrellaceae, Hygrophoraceae, Russullaceae, Clavariaceae, Tricholomataceae, and Thelephoraceae with four additional unidentified families. Representatives of the Hygrophoraceae were sparsely distributed but more prevalent in WI. Family Russulaceae was more prevalent in the MS unexposed areas. Family Tricholomataceae was sparsely distributed among mixed preservative and copper treated sites in MS. Family Clavariaceae was sparsely distributed

in WI among all the treatments and less common in MS. Class Archeorhizomycetes was more prevalent in the preservative exposed sites and far less common in the unexposed areas. Family Mortierrellaceae was highly abundant and widespread in MS and widespread but less abundant in WI. Relative abundances and distributions of the Families are shown in **Supplemental Figure 4**.

A total of 1,136 OTUs were classified to the genus level with an additional 509 OTUs classified to species. The genus *Mortierella* was highly abundant and widespread in MS but also widespread in WI with lower abundance. Four additional fungal unidentified genera were sparsely distributed. The genus *Hygrocybe* was more abundant in the WI soil samples especially one unexposed sample. The genus *Lactarius* was widely distributed and found in both sites (MS and WI). Relative abundances and distributions of the genera are shown in **Supplemental Figure 5**.

Mortierella humulis was the only widespread, abundant fungus identified to species in the study and was highly prevalent in MS, but sparsely distributed in WI.

Fungal Community Analyses

In order to compare fungal species composition between sites and treatments, PERMANOVA was used to analyze the taxonomy data. Due to large differences in community composition between the MS and WI sites, data sets were analyzed separately.

Sites were significantly different from each other with respect to fungal species composition ($P < 0.0001$) (**Figure 1**). Chemical analysis showed richer nutrient profiles in the MS site and lower pH of soil (**Table 1**). Highly correlated OTUs that contributed to the differences between sites along each NMDS axis were compiled in PERMANOVA and the five OTUs with the highest and lowest degrees of correlation for each site and axis of the NMDS ordination are listed in **Table 2**. These represent outliers from the shared core fungal groups within the total pool of OTUs and help to define differences between the sites.

Significant treatment and dispersion effects were seen in MS (PERMANOVA on treatment r-squared = 0.96, $p < 0.0001$, $F = 125$ Mississippi betadisper $F = 61.871$ $p < 0.001$) and WI (PERMANOVA on treatment r-squared = 0.39, $p = 0.09$, $F = 2.85$ betadisper $F = 7.98$, $p < 0.001$). 2-Dimensional results from the Non-metric Multidimensional Scaling (NMDS) analysis are presented in **Figure 2** showing treatment differences in both MS (**Figure 2A**) and WI (**Figure 2B**).

Indicator Species Analysis

Thirteen indicator species were found in the mixed preservative MS plot (**Table 3**); of those, 6 were highly correlated (corr Val = 1). Twenty-three indicator species were found in the mixed preservative WI plot (**Table 4**); of those, 8 were highly correlated (corr Val = 1). A total of 14 indicator species were identified with the copper treatments in MS (**Table 5**), three of which were highly correlated with copper treatments (Corr_val = 1) and the remainder were still correlated, but to a lesser extent (Corr_val = 0.894). A total of 7 indicator species were identified for copper exposed soils in WI (**Table 6**). Two species (*Orbillia* sp. and *Rhytidhysteron rufulum*) had the highest degree of association (Corr_val = 1), while the remaining species were less correlated (Corr_value = 0.894–0.0866). Twenty-five indicator species were found in the unexposed MS plot (**Table 7**); of those, 6 were highly correlated (corr Val = 1). A total of four indicator species were identified for the WI unexposed soil (**Table 8**).

Functional Guilds Analysis

Of the total 6,668 OTUs, 1,330 (20%) of the OTUs were classified using Funguild (Nguyen et al., 2016), which assigns functional guild information to taxonomic sequence data.

DISCUSSION

This study marks our first amplicon-based DNA sequence analysis of soil fungi under long-term wood preservative exposure. Characterization of soil fungi by metabarcode analysis gives a prediction of which fungi are present in a given location and can give highly detailed information about the decay potential of a given site. This study provided baseline data about our field sites that can be used to assist in future field exposure decisions.

Community Analysis

Unsurprisingly, the fungal soil communities differed between our two field sites that were separated by 995 miles and 13° of latitude. Overall, a greater number of OTUs were found in the WI site, presumably due to the more neutral soil pH and richer soil nutrient composition (see **Table 1**). Differences between treatments were also noted for both sites with greater difference between treatments in the harsher decay hazard climate (MS); treatments were also different in WI but with much greater variability.

Indicator Species

Indicator species were determined for each site treatment combination and notable observations were the low diversity of wood saprobes present in the soil, which did not appear to be influenced by treatment history but did differ between test sites. Several ectomycorrhizal fungi were found to be associated with both mixed and Cu preservatives exposure, which agrees with previous literature (Meharg and Cairney, 2000; Fomina et al., 2005; Colpaert, 2008), suggesting that ectomycorrhizal fungi are able to persist and might be beneficial remediators for preservative exposed soils, and that long-term preservative exposure has little effect on the abundance and prevalence of this group of fungi.

Thirteen indicator species were found in the mixed preservative MS plot (**Table 2**); of those, 6 were highly correlated (corr Val = 1). *Cenococcum geophilum* is a mycorrhizal species associated with a diverse list of hosts with wide distribution and is routinely identified from soil, based on morphology of colonized roots (Douhan and Rizzo, 2005). Herpotrichiellaceae is a family of loculoascomycetes with black yeast anamorphs (Untereiner et al., 1995). *Ceriporiopsis* sp. include lignin degrading white-rot fungi studied for enzymes that can be used in biopulping (Ferraz et al., 2003). *Neofusicoccum grevillea* is a pathogen on *Frivillea aurea* causing a leaf spot (Sakalidis et al., 2013). The genus

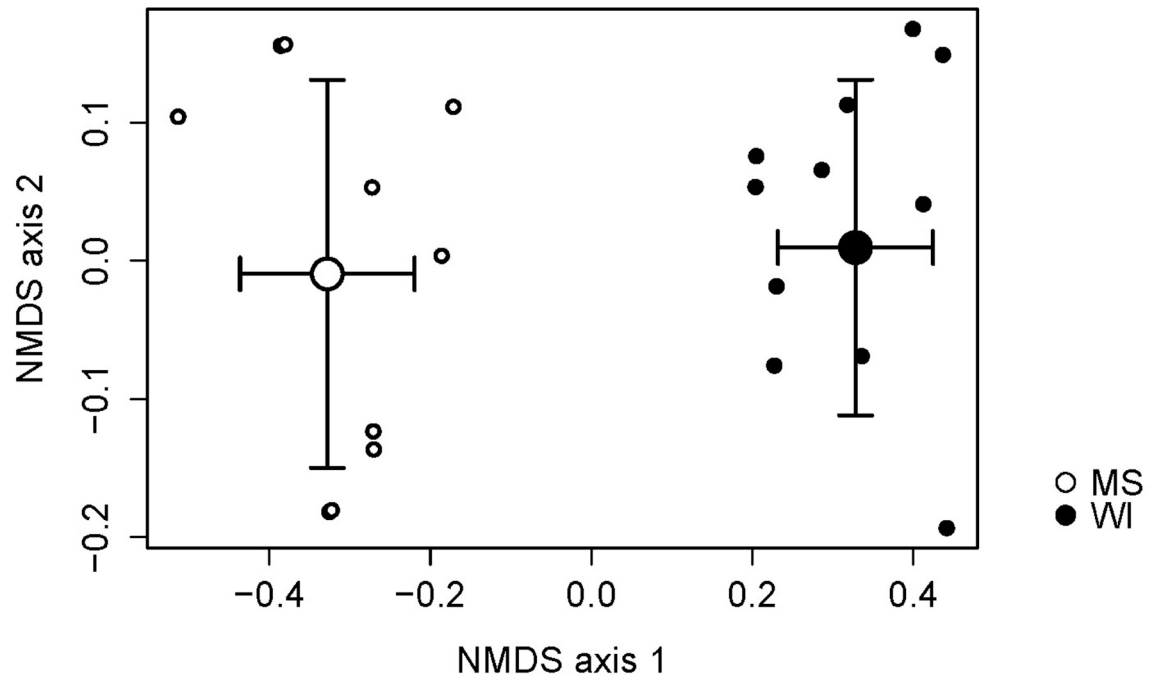


FIGURE 1 | Non-metric multidimensional scaling (NMDS) plot indicating the effects of site on the fungal species composition. Sites were distinctly different and necessitated separate analyses of treatments.

TABLE 1 | Results of chemical analysis of soil samples.

SampleID	Sample#	Site	Treat	Rep	Descr	Lat	Long	pH	Ca	Cr	Cu	K	Mg	P	As
WIA1	1	WI	A	1	Mixed	43°02′39″N	89°33′18″W	7	2.88	0.02	0.03	1.60	2.94	0.55	0.01
WIA2	2	WI	A	2	Mixed	43°02′39″N	89°33′18″W	6.15	2.44	0.02	0.03	1.56	2.85	0.55	0.01
WIA3	3	WI	A	3	Mixed	43°02′39″N	89°33′18″W	6.82	2.52	0.02	0.03	1.74	3.20	0.57	0.01
WIA4	4	WI	A	4	Mixed	43°02′39″N	89°33′18″W	6.23	2.76	0.01	0.03	1.32	2.62	0.56	0.01
WIB1	5	WI	B	1	Cu	43°02′40″N	89°33′09″W	6.18	3.32	0.01	0.03	1.52	2.54	0.64	0.01
W IB2	6	WI	B	2	Cu	43°02′40″N	89°33′09?W	6.15	2.55	0.01	0.03	1.33	2.59	0.57	0.01
WIB3	7	WI	B	3	Cu	43°02′40″N	89°33′09″W	6.09	3.36	0.01	0.03	1.16	3.01	0.48	0.01
WIB4	8	WI	B	4	Cu	43°02′40″N	89°33′09″W	6.5	2.84	0.01	0.03	1.34	2.57	0.43	0.01
WIC1	9	WI	C	1	Control	43°02′40″N	89°33′04″W	6.85	99.89	0.00	0.02	0.76	54.12	0.67	nd
WIC2	10	WI	C	2	Control	43°02′40″N	89°33′04″W	6.8	101.35	0.00	0.02	0.71	54.66	0.78	nd
WIC3	11	WI	C	3	Control	43°02′40″N	89°33′04″W	6.47	2.91	0.02	0.03	1.48	3.02	0.58	0.01
WIC4	12	WI	C	4	Control	43°02′40″N	89°33′04″W	6.3	2.89	0.02	0.03	1.77	2.85	0.65	0.01
MSA1	13	MS	A	1	Mixed	30°37′44″N	89°02′46″W	4.59	0.89	0.01	0.01	0.57	0.59	0.05	0.01
MSA2	14	MS	A	2	Mixed	30°37′44″N	89°02′46″W	4.95	0.25	0.01	0.01	0.23	0.29	0.12	0.00
MSA3	15	MS	A	3	Mixed	30°37′44″N	89°02′46″W	4.99	2.68	0.00	0.01	0.58	0.51	0.24	0.01
MSA4	16	MS	A	4	Mixed	30°37′44″N	89°02′46″W	4.94	0.38	0.00	0.01	0.28	0.30	0.12	0.02
MSB1	17	MS	B	1	Cu	30°37′44″N	89°02′47″W	4.82	0.36	0.00	0.01	0.21	0.26	0.11	nd
MSB2	18	MS	B	2	Cu	30°37′44″N	89°02′47″W	5.22	0.40	0.00	0.01	0.16	0.20	0.08	0.00
MSB3	19	MS	B	3	Cu	30°37′44″N	89°02′47″W	4.97	1.48	0.00	0.33	0.37	0.45	0.29	nd
MSB4	20	MS	B	4	Cu	30°37′44″N	89°02′47″W	5.16	0.09	0.00	0.02	0.18	0.24	0.07	nd
MSC1	21	MS	C	1	Control	30°37′42″N	89°33′04″W	4.83	0.64	0.00	0.01	0.28	0.31	0.25	0.00
MSC2	22	MS	C	2	Control	30°37′42″N	89°33′04″W	5.33	0.24	0.00	0.01	0.21	0.20	0.15	nd
MSC3	23	MS	C	3	Control	30°37′42″N	89°33′04″W	5.14	0.27	0.01	0.01	0.30	0.30	0.22	0.00
MSC4	24	MS	C	4	Control	30°37′42″N	89°33′04″W	5.08	0.60	0.00	0.01	0.28	0.28	0.18	0.00

pH was determined by benchtop pH meter and soil elemental concentrations were determined by ICP-OES and are expressed in mg/kg of soil. WI soils were richer in macro-nutrients and also with higher pH values than those in MS. Residual preservatives were not elevated in any of the plots compared to untreated areas with only one exception (MSB3-elevated copper).

TABLE 2 | Highly correlated OTUs contributing to differences between sites MS and WI.

OTU	Corr	Taxonomy	OTU	Corr	Taxonomy
MS NMDS-axis 1 top 5 positively correlated OTUS			**WI NMDS-axis 1-top 5 positively correlated OTUs**		
OTU_357	0.91972	*Phialocephala humicola*	OTU_66	0.961668	*Mortierella humilis*
OTU_1091	0.91972	*Wickerhamomyces subpelliculosus*	OTU_99	0.961668	*Saccharomycopsis crataegensis*
OTU_1476	0.871663	*Sebacinales*	OTU_148	0.961668	*Penicillium copticola*
OTU_1531	0.871663	*Glomus proliferum*	OTU_199	0.961668	*Mortierella* spp.
OTU_2139	0.871663	*Tomentella amyloapiculata*	OTU_284	0.961668	*Kodamaea ohmeri*
MS NMDS-axis 1 top 5 negatively correlated OTUS			**WI NMDS-axis 1-top 5 negatively correlated OTUs**		
OTU_644	−0.88078	*Pseudozyma prolifica*	OTU_6907	−0.96167	*Tremella subalpina*
OTU_3456	−0.88078	*Genea cephalonicae*	OTU_7079	−0.96167	*Batcheloromyces sedgefieldii*
OTU_523	−0.91972	*Hygrocybe virginea*	OTU_7086	−0.96167	*Clydaea vesicula*
OTU_876	−0.91972	*Anthostomella leucospermi*	OTU_7117	−0.96167	*Mycosphaerella elongata*
OTU_1781	−0.91972	*Cladophialophora chaetospira*	OTU_7260	−0.96167	Unidentified
MS NMDS-axis 2 top 5 positively correlated OTUS			**WI NMDS-axis 2-top 5 positively correlated OTUs**		
OTU_570	0.803391	*Glomus proliferum*	OTU_135	0.89658	*Populocrescentia forlicesenensis*
OTU_950	0.803391	*Claroideoglomus luteum*	OTU_994	0.851367	*Buckleyzyma aurantiaca*
OTU_2713	0.784669	*Resinicium furfuraceum*	OTU_3609	0.851367	*Scytalidium lignicola*
OTU_4984	0.751979	*Ceramothyrium podocarpi*	OTU_1506	0.766713	*Clavaria falcata*
OTU_145	0.705732	*Umbelopsis swartii*	OTU_2178	0.766713	*Glomus proliferum*
MS NMDS-axis 2 top 5 negatively correlated OTUS			**WI NMDS-axis 2-top 5 negatively correlated OTUs**		
OTU_2638	−0.87761	*Protrudomyces lateralis*	OTU_128	−0.66493	*Mortierella hyalina*
OTU_246	−0.88503	*Mortierella echinula*	OTU_816	−0.66493	*Hebeloma erebium*
OTU_568	−0.88503	*Umbilicaria iberica*	OTU_3820	−0.66493	*Lactifluus subiculatus*
OTU_1984	−0.88503	*Periconia macrospinosa*	OTU_349	−0.74489	*Polyscytalum algarvense*
OTU_2630	−0.88503	*Tremella samoensis*	OTU_546	−0.74489	*Hirsutella rhossiliensis*

i.e., Core groups of fungal species that can be used to distinguish between the two sites.

Neopaxillus is in the crepidotaceae family in the agaricales, *N. dominicanus* is a new species found in the Dominican Republic in 2011 (Vizzini et al., 2012).

Twenty-three indicator species were found in the mixed preservative WI plot (**Table 3**); of those, 8 were highly correlated (corr Val = 1). Of the highly correlated, *Clavicorona taxophila* is a saprotroph in the Clavariaceae family including mushroom-forming fungi with *Clavicorona* producing coralloid sporocarps (Birkebak et al., 2013). It is a small coral fungus that is currently red-listed in the UK as threatened (Evans et al., 2006). *Nothojafnea cryptotricha* is a mycorrhiza of eucalypts (Warcup, 1990). *Archaeorhizomyces borealis* is a pine associated species of the soil-inhabiting genus commonly isolated from environmental samples (Menkis et al., 2014). They constitute a significant component of the rhizosphere in fungal DNA community analyses, comprising up to one third of the total fungal community (Porter et al., 2008). *Ganoderma lucidum* belongs to a genus of white rot fungi that are considered one of the most important medicinal fungi worldwide and is an ingredient in many health products supposedly having anti-cancer, anti-aging, and antimicrobial functions (Paterson, 2006). *Podospora intestinacea* is a coprophilous fungus commonly isolated form dung and produces a perithecial resting structure that can persist in soil or dung (Watling and Richardson, 2010). *Flagelloscypha citrispora* is a basidiomycete that occurs on rotten logs and stumps (Reid, 1964). Species of the *Coniothyrium* genus have been used as biocontrol for *Sclerotinia* as they are mycoparasites of mycelium and sclerotia; this treatment has been used in sunflower production (Whipps and Gerlagh, 1992). *Operculomyces laminatus* is a chytrid formerly known as *Rhizophlyctis harderi* (Powell et al., 2011).

A total of 14 indicator species were identified with the copper exposure in MS (**Table 4**). The highest correlations between species and copper treatment were for *Hygrocybe virginiana*, *Anthostomella leucospermi*, and *Cladophialophora chaetospira*.

A total of 7 indicator species were identified for copper exposed soils in WI (**Table 5**). *Orbillia* sp. is an operculate ascomycete belonging to the family Orbilliaceae which are often nematophagous (Ahrén et al., 1998); members of this genus have been isolated from forest soils in Argentina (Allegrucci et al., 2009). It produces an orange cup shaped fruiting body and has been observed on copper treated field stakes from our Wisconsin field site. Efforts are currently underway to confirm these fruiting bodies as being *Orbillia* spp. *Rhytidhysteron rufulum* is a saprophyte or weak parasite on a variety of plants most commonly associated with pan-tropical environments (Murillo et al., 2009).

Twenty-five indicator species were found in the unexposed MS plot (**Table 6**); of those, 6 were highly correlated (corr Val = 1). *Arthrinium arundinis* is a plant pathogen reported to cause kernel blight of barley (Crous and Groenewald, 2013). *Cryptococcus cuniculi* was previously described as an anamorphic yeast that was isolated from rabbit feces and is now considered a basidiomycetous yeast (Findley et al., 2009). *Oliveonia pauxilla*

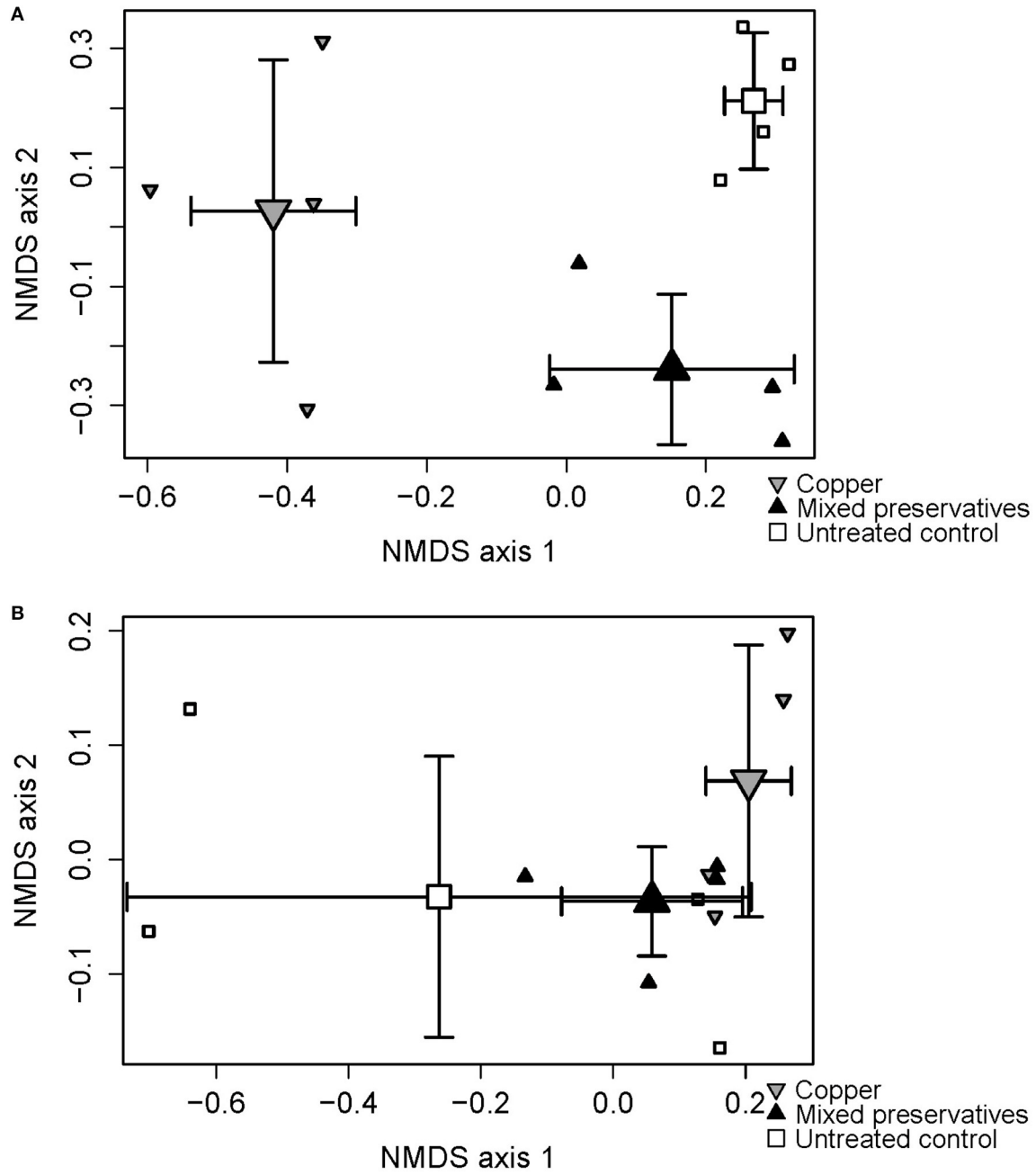


FIGURE 2 | Non-metric multidimensional scaling (NMDS) ordination plot displaying differences between treatments with respect to fungal species composition at **(A)** MS test site and **(B)** WI test site.

is a holobasidiomycete fungus and a saprobe on dead leaf litter (Warcup and Talbot, 1962). *Craterocolla cerasi* is a non-culturable fungus in the sebacinaceae group of the Hymenomycetidae family with fungi involved in various mycorrhizal associations (Weiss et al., 2004). *Flagelloscypha citrispora* is a wood saprobe also found in the mixed preservative plot in WI described above.

A total of four indicator species were determined for the WI unexposed soil (**Table 7**). A group containing Neotropical ascomycetes (*Coniochaetales* sp.) was correlated (0.866), and has a related species, *Coniochaetales lignaria*, that has been evaluated for its ability to break down furans and other phenolics that are the result of lignocellulose pretreatment for hydrolysis (López et al., 2004). *Septoglomus* sp. was also found to be correlated with unexposed soils in WI. The *Septoglomus* genus contains mostly mycorrhizal species that are widely distributed (Rydlová et al., 2015). This isolate clustered with *Septoglomus constrictum*, which has been isolated in other studies associated with heavy metal contaminated mine spoil (Shetty et al., 1994). *Geotrichum*

TABLE 3 | Indicator species for soil exposed to mixed preservative chemistries in MS for 50 years.

OTU_ID	Corr_val	*p*-value	Species_ID
OTU_46	1	0.010**	*Cenococcum geophilum*
OTU_416	1	0.010**	*Herpotrichiellaceae* sp.
OTU_463	1	0.010**	*Ceriporiopsis* sp.
OTU_1356	1	0.010**	*Neofusicoccum grevillea*
OTU_2038	1	0.010**	*Orbilia* sp.
OTU_2783	1	0.010**	*Neopaxillus dominicanus*
OTU_246	0.894	0.034*	*Mortierella echinula*
OTU_568	0.894	0.034*	*Pleosporales* sp.
OTU_1223	0.894	0.034*	*Ramariopsis corniculata*
OTU_1299	0.894	0.035*	*Ramariopsis* sp.
OTU_1980	0.894	0.031*	*Botryobasidium* sp.
OTU_1984	0.894	0.034*	*Periconia macrospinosa*
OTU_2630	0.894	0.034*	*Tremella samoensis*

*Indicates significance at the 0.05 level, and **indicates significance at the 0.005 level.

TABLE 4 | Indicator species for soil exposed to mixed preservative chemistries in WI for 60 years.

OTU_ID	Corr_val	*p*-value	Species_ID
OTU_55	1	0.006**	*Clavicorona taxophila*
OTU_378	1	0.006**	*Nothojafnea cryptotricha*
OTU_578	1	0.006**	*Archaeorhizomyces borealis*
OTU_1242	1	0.006**	*Ganoderma lucidum*
OTU_1288	1	0.006**	*Podospora intestinacea*
OTU_1882	1	0.006**	*Flagelloscypha citrispora*
OTU_2588	1	0.006**	*Coniothyrium* sp.
OTU_3787	1	0.006**	*Operculomyces laminatus*
OTU_282	0.894	0.036*	*Inocybe* sp.
OTU_807	0.894	0.027*	*Fellomyces lichenicola*
OTU_905	0.894	0.027*	*Lactarius chrysorrheus*
OTU_931	0.894	0.029*	*Oliveonia pauxilla*
OTU_1118	0.894	0.029*	*Glomus* sp.
OTU_1382	0.894	0.029*	*Lophiostoma corticola*
OTU_1442	0.894	0.029*	*Myrothecium leucotrichum*
OTU_1938	0.894	0.026*	*Penicillium coprobium*
OTU_3470	0.894	0.026*	*Kodamaea ohmeri*
OTU_18	0.866	0.048*	*Ramariopsis helvola*
OTU_1226	0.866	0.050*	*Lasiodiplodia crassispora*
OTU_1639	0.866	0.050*	*Sarocladium mycophilum*
OTU_2694	0.866	0.048*	*Mortierellales* sp.
OTU_2801	0.866	0.048*	*Pachyphlodes citrinus*
OTU_3788	0.866	0.048*	*Cladophialophora* sp.

*Indicates significance at the 0.05 level, and **indicates significance at the 0.005 level.

candidum was also correlated (0.866) with untreated WI soils and has been widely studied for its production of lipases (Sugihara et al., 1990), phenol oxidases (Assas et al., 2000) and a host of other enzymes that can effectively break down a wide range of chemical classes (Sun et al., 2008). *Ophiostoma piliferum* is a sapstain fungus associated with bark beetles (Klepzig, 1998). The overall low diversity of the unexposed WI soils could be attributed to habitat. The samples were taken from a grassy prairie area with virtually no canopy and very little coarse woody debris—a driver of fungal diversity (Zellweger et al., 2015)—in the vicinity.

TABLE 5 | Indicator species for MS for soils exposed to copper treatments for ~30 years.

OTU_ID	Corr_val	*P*-value	BLAST result
OTU_523	1	0.009**	*Hygrocybe virginea* var. *ochraceopallida*
OTU_876	1	0.009**	*Anthostomella leucospermi*
OTU_1781	1	0.009**	*Cladophialophora chaetospira*
OTU_95	0.894	0.045*	*Cladophialophora* sp.
OTU_521	0.894	0.029*	*Helotiales* sp.
OTU_644	0.894	0.031*	*Pseudozyma prolifica*
OTU_898	0.894	0.029*	Dothidiomycetes sp.
OTU_996	0.894	0.045*	*Phialophora europaea*
OTU_1669	0.894	0.045*	*Sordariales* sp.
OTU_1981	0.894	0.036*	*Leccinum talamancae*
OTU_2235	0.894	0.038*	*Cladophilaophora* sp.
OTU_1390	0.866	0.047*	*Dactylaria* sp.
OTU_2283	0.866	0.047*	*Glomus diaphanum*
OTU_4748	0.866	0.047*	*Tremellodendropsis tuberosa*

*Indicates significance at the 0.05 level, and **indicates significance at the 0.005 level.

TABLE 6 | Indicator species for WI soils exposed to copper treatments for 40 years.

OTU_ID	Corr_val	*p*-value	Species ID
OTU_332	1	0.007**	*Orbilia* sp.
OTU_3106	1	0.007**	*Rhytidhysteron rufulum*
OTU_472	0.894	0.026*	*Tuber* sp.
OTU_1506	0.866	0.046*	*Clavaria falcate*
OTU_2178	0.866	0.046*	*Glomus* mycorrhizal symbiont of Marchantia foliacea
OTU_2190	0.866	0.046*	*Glomus* sp.
OTU_2313	0.866	0.046*	*Metarhizium anisopliae*

*Indicates significance at the 0.05 level, and **indicates significance at the 0.005 level.

Funguild Analysis

Relative proportions of fungal guilds were generally conserved across sites and preservative exposures. Pie charts showing proportions of guilds classified for each treatment group are shown in **Figure 3**. Wood saprobes compromised a relatively small portion of the total fungi detected in each sample group, likely owing to the fact that isolations where made from soil and not wood. Higher proportions of ECM fungi were noted at the MS site, but this is likely due to the predominant pine overstory which is heavily dependent on ECM fungi for growth and survival (Svenson et al., 1991) and predominately lower pH soils.

Analysis at the functional guild level showed that wood saprobes also make up a relatively small proportion of the total soil fungal community (7.29% of the OTUs characterized) and that saprophytic fungi, many of which are capable of inducing

TABLE 7 | Indicator species for unexposed soils in MS.

OTU_ID	Corr_val	*p*-value	Species_ID
OTU_330	1	0.008**	Ericoid mycorrhizal sp.
OTU_1273	1	0.008**	*Arthrinium arundinis*
OTU_1618	1	0.008**	*Cryptococcus cuniculi*
OTU_2246	1	0.008**	*Oliveonia pauxilla*
OTU_2485	1	0.008**	*Craterocolla cerasi*
OTU_3377	1	0.008**	*Flagelloscypha citrispora*
OTU_20	0.894	0.036*	*Lactarius cf. subserifluus*
OTU_412	0.894	0.033*	*Cladophialophora* sp.
OTU_672	0.894	0.042*	*Cladophialophora* sp.
OTU_688	0.894	0.042*	*Dioszegia rishiriensis*
OTU_1226	0.894	0.031*	*Lasiodiplodia crassispora*
OTU_1447	0.894	0.033*	*Mucoromycotina* sp.
OTU_1515	0.894	0.042*	*Limonomyces culmigenus*
OTU_7398	0.894	0.036*	*Herpotrichiellaceae* sp.
OTU_131	0.866	0.049*	*Cenococcum geophilum*
OTU_339	0.866	0.049*	*Scleroderma polyrhizum*
OTU_462	0.866	0.049*	*Lambiella caeca*
OTU_700	0.866	0.049*	*Hyaloscyphaceae* sp.
OTU_1089	0.866	0.049*	Glomeromycetes sp.
OTU_1102	0.866	0.049*	*Pulvinula constellation*
OTU_1564	0.866	0.049*	Dothideomycetes sp.
OTU_2047	0.866	0.049*	*Glomus* sp.
OTU_2629	0.866	0.049*	*Mortierellales* sp.

*Indicates significance at the 0.05 level, and **indicates significance at the 0.005 level.

TABLE 8 | Indicator species for unexposed soils in WI.

OTU_ID	Corr_val	*p*-value	Species_ID
OTU_372	0.866	0.049*	*Coniochaetales* sp.
OTU_928	0.866	0.049*	*Septoglomus* sp.
OTU_1201	0.866	0.049*	*Geotrichum candidum*
OTU_3016	0.866	0.049*	*Ophiostoma piliferum*

*Indicates significance at the 0.05 level.

soft rot fungal degradation patterns in wood, comprise a large portion. A total of 67 OTUs were classified as animal pathogens, which include *Cryptococcus*, *Beauvaria*, *Metarhizium*, *Fusarium*, *Cordyceps,* and others. Two additional OTUs were classified as animal symbionts, both belonging to the genus *Rhodotorula*. A total of 233 OTUs were classified as Arbuscular Mycorrhizae mostly contained in the genus *Glomus*. A total of 41 OTUs were classified as dung associated saprotrophs mostly *Penicillium*, *Chaetomium,* or *Podospora*.

A total of 147 OTUs were classified as Ectomycorrhizal fungi and included the genera *Lactarius*, *Russula*, *Amanita*, *Tuber* and several other commonly found fleshy terrestrial fungi. Several representatives within this group were found to be indicator species in the community analysis, *Lactarius chrysorrheus*- (mixed preservative associated in WI), *Tuber* sp. (Cu associated in WI), and *Leccinum talamanaceae* (Cu associated in MS). Ectomycorrhizal fungi have been found to possess abilities to biotransform and sequester toxins, such as arsenic in the form of arsenobetaine (Nearing et al., 2015) and other persistent heavy metals contaminating soil (Khan et al., 2000). Previous works to bio-augment and remediate treated wood waste and soil have revolved primarily around preservative tolerant species of polyporoid fungi (De Groot and Woodward, 1999; Illman and Yang, 2004) and have not fully explored the roles of these terrestrial fungi as potential remediators. Ectomycorrhizal fungi have been studied for their role in remediating heavy metal contaminated soil and several species that are known to be metal tolerant were detected in our study (*Cenococcum geophilum*, *Amanita* spp., and *Pisolothus tinctorus*). An important exception would be *H. virginiana*, which was found to be highly associated with Cu treatments in MS, but is actually classified as an undefined saprotroph since *Hygrocybe* species are considered to be neither saprotrophic nor mycorrhizal (Seitzman et al., 2010) but symbionts of mosses. A total of 17 OTUs were classified as endophytes and included many common plant pathogens (*Phialophora*, *Cladosporium*, *Periconia macrospinosa*). Representatives of this guild were found as indicator species in both unexposed (*Herpotrichiellaceae* sp.), mixed preservative exposed (*P. macrospinosa*) and Cu exposed soils (*Phialophora europaea*). A total of 590 OTUs were classified as saprophytes and included an extremely wide range of species from the entire spectrum of fungal taxonomy (lower fungi, such as Mortierella to higher basidiomycetes, such as *Peniophorella pallida* and *Lycoperdon perlatum*).

A total of 97 OTUs were classified as wood saprobes, which would include all of the typical wood decay fungi. **Table 8** shows a breakdown of those fungi classified as wood saprobes by Funguild and their decay traits (soft, brown or white) with their occurrence across treatments.

A total of 14 brown rots, 38 soft rots and 37 white rots were identified and characterized from the soil samples (an additional 8 OTUs, containing 4 species of *Psathyrella*, *Crepidotus* sp. and *Simocybe* spp., were not characterized according to decay traits). Overall, a relatively low number of known wood decay fungi were obtained from our sampling, but these results will serve as important baseline data as we begin to focus on fungi that colonize and persist in solid wood and also highlights the importance of saprophytic fungi in the decay process. At both sites, higher abundance of wood rot fungi was noted for the exposed areas compared to unexposed areas. A relatively small number of wood saprobes were recovered in our sampling. Many of these species are well studied wood rot fungi or at least the genera, except for *Trechispora* spp. which is more common on decaying wood debris on the forest floor than solid wood. The less recognized species could be more difficult to culture or less likely to show fruiting in the field or even occur in later stages of decay. Our results could lead to more studies on these lesser emphasized wood rot fungi, especially in the preservative tolerance field. Proximity to remaining field stakes in the plot was not taken into consideration in this study, but could be a significant source of variation in these types of studies. Future studies to address these effects would provide much needed information on whether or not nutrient availability (i.e., untreated wood stakes) stimulates below ground fungal activity. Future studies will include more detailed information on wood preservative residues

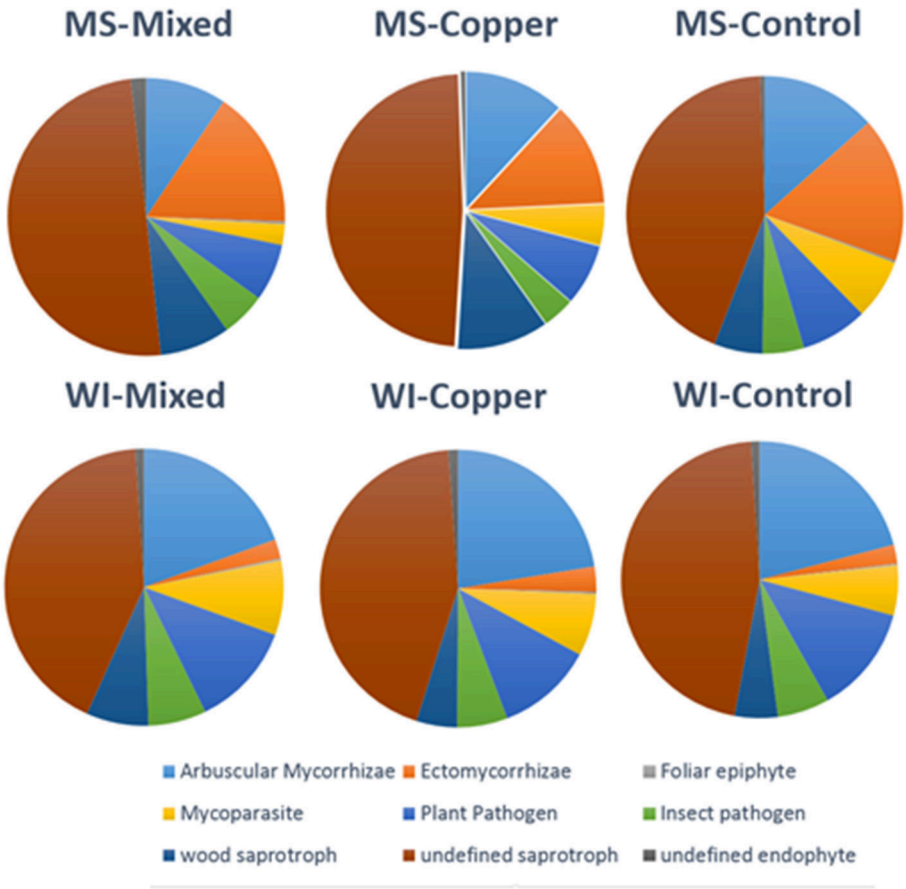


FIGURE 3 | Distribution of major fungal guilds by site and preservative exposure of all OTUs classified using Funguild (*N* = 1,330 out of 6,668).

and breakdown products in the soil and would be an obvious next step in correlating long term preservative breakdown with fungal community composition. Another important consideration is whether present soil fungi directly correlate to wood decay. Continued studies will investigate transfer of soil fungi into untreated pine stakes and eventually preservative treated pine in field exposure.

Additional Copper Tolerant Fungi

A secondary objective of this study was to screen for the presence of copper tolerant basidiomycete fungi in copper exposed sites. Several fungal genera and species were identified with published accounts of copper tolerance. These species were not abundant enough to be classified as indicator species but still exhibited patterns of presence consistent with prior literature. The abundance and distribution of these fungal OTUs by can be found in **Table 9**.

The brown rot are most typically associated with copper tolerance and several taxa were found in copper exposed soils belonging to that group. *Postia guttulata* was only detected in preservative exposed soils in WI along with *Antrodia carbonica*. *Antrodia carbonica* has previously been documented as copper tolerant and *P. guttulata* is in the same genus as *Postia placenta*, which can also exhibit copper tolerance (De Groot and Woodward, 1999). The lack of brown rot detected in untreated areas in MS cannot be explained, however fungi characterized as white rot and soft rot fungi were detected. There were several brown rots only found in preservative treated areas in MS. *Fibroporia radiculosa* is a very well-known cooper tolerant brown rot fungus (Clausen and Jenkins, 2011) and was detected in the mixed preservative post plots in MS. This fungus is also common in our copper treated samples exposed to soil in MS, but wasn't as prevalent as expected in these soil samples. *Dacryobolus sudans* and *Coniophora* sp. were also detected in the mixed preservative plots. *Coniophora puteana* is a common preservative tolerant fungus, especially on creosote treated members and *Dacryobolus sudans* has been shown to at least colonize PAH contaminated soils (Tornberg et al., 2003) also typically associated with creosote contaminated soils. *Leucogryaphana olivasceans* was also abundant in copper plots in MS, but no prior literature describing copper tolerance of this species was found. *Serpula himantoides* was detected in both mixed preservative and copper exposed plots in MS and is closely related to *S. lacrymans*, which has published records of copper tolerance (Watkinson and Eastwood, 2012).

TABLE 9 | Breakdown of Basidiomycete fungi classified as wood saprobes by Funguild highlighting the effects due to treatment on wood decay fungal diversity.

Species	MS			WI			
	Mixed	Copper	Control	Mixed	Copper	Control	Decay mode
Trechispora sp.	0	0	0	3,783	301	1601	White Rot
Trechispora invisitata	0	3,131	0	0	0	0	White Rot
Trechispora sp.	0	0	0	2,927	0	0	White Rot
Trechispora sp.	0	0	0	1,292	11	0	White Rot
Ganoderma applanatum	0	0	0	111	223	446	White Rot
Ceriporiopsis sp.	264	0	0	0	0	0	White Rot
Heterobasidion sp.	92	16	3	0	0	0	White Rot
Trechispora sp.	0	0	0	56	38	0	White Rot
Trechispora sp.	0	0	0	40	4	0	White Rot
Ganoderma lucidum	9	18	10	1	0	0	White Rot
Bjerkandera sp.	0	37	0	0	0	0	White Rot
Sistotrema sp.	25	1	3	0	0	4	White Rot
Trechispora sp.	1	0	28	0	0	0	White Rot
Gymnopilus liquiritiae	0	10	0	0	0	0	White Rot
Bjerkandera adusta	0	0	0	1	0	5	White Rot
Trechispora sp.	5	0	0	0	0	0	White Rot
Irpex sp.	0	1	2	0	0	1	White Rot
Trechispora sp.	1	1	2	0	0	0	White Rot
Perenniporia sp.	0	0	0	0	0	3	White Rot
Ganoderma sp.	0	0	0	0	1	2	White Rot
Perenniporia subacida	1	1	1	0	0	0	White Rot
Phlebia sp.	0	3	0	0	0	0	White Rot
Trechispora sp.	1	1	1	0	0	0	White Rot
Ischnoderma resinosum	0	0	0	0	0	2	White Rot
Rigidoporus sp.	0	1	1	0	0	0	White Rot
Trechispora sp.	0	0	0	2	0	0	White Rot
Schizophyllum sp.	0	0	2	0	0	0	White Rot
Trechispora sp.	0	0	1	0	0	0	White Rot
Ganoderma lucidum	0	0	0	0	1	0	White Rot
Sistotrema sp.	0	1	0	0	0	0	White Rot
Daedaleopsis confragosa	0	0	0	0	0	1	White Rot
Polyporus squamosus	0	0	0	1	0	0	White Rot
Phlebia radiata	0	0	0	0	0	1	White Rot
Trichaptum sp.	0	1	0	0	0	0	White Rot
Bjerkandera sp.	0	1	0	0	0	0	White Rot
Trechispora sp.	1	0	0	0	0	0	White Rot
Ganoderma sp.	0	1	0	0	0	0	White Rot
Scytalidium lignicola	0	1,610	0	0	0	0	Soft Rot
Cladosporium herbarum	0	0	0	397	102	998	Soft Rot
Humicola sp.	0	0	0	25	0	1,086	Soft Rot
Trichoderma sp.	71	179	106	277	17	13	Soft Rot
Trichoderma sp.	360	87	75	31	7	0	Soft Rot
Scytalidium sp.	26	45	197	0	0	0	Soft Rot
Scytalidium sp.	74	79	0	3	0	0	Soft Rot
Scytalidium sp.	34	75	0	0	0	0	Soft Rot
Trichoderma crassum	20	47	26	0	1	0	Soft Rot
Phoma sp.	0	0	0	27	1	21	Soft Rot
Chalara sp.	0	0	41	0	0	0	Soft Rot

(Continued)

TABLE 9 | Continued

Species	MS			WI			
	Mixed	Copper	Control	Mixed	Copper	Control	Decay mode
Stilbella fimetaria	0	0	0	27	0	0	Soft Rot
Alternaria infectoria	0	0	0	3	5	16	Soft Rot
Xylaria sp.	11	1	0	0	0	0	Soft Rot
Chalara sp.	0	11	0	0	0	0	Soft Rot
Scytalidium sp.	10	0	0	0	0	0	Soft Rot
Xylaria sp.	0	2	8	0	0	0	Soft Rot
Scytalidium sp.	8	1	0	0	0	0	Soft Rot
Trichoderma sp.	0	0	0	7	0	0	Soft Rot
Scytalidium sp.	0	7	0	0	0	0	Soft Rot
Coniothyrium sp.	0	0	0	1	1	5	Soft Rot
Trichoderma hamatum	2	0	4	0	0	0	Soft Rot
Trichoderma sp.	1	1	2	0	0	1	Soft Rot
Coniothyrium sp.	0	0	0	0	0	3	Soft Rot
Scytalidium sp.	0	0	3	0	0	0	Soft Rot
Scytalidium sp.	0	0	0	3	0	0	Soft Rot
Coniothyrium sp.	0	0	0	0	2	0	Soft Rot
Xylaria sp.	2	0	0	0	0	0	Soft Rot
Phoma sp.	0	0	0	0	0	2	Soft Rot
Cytospora diatrypelloidea	2	0	0	0	0	0	Soft Rot
Xylaria sp.	0	0	0	0	2	0	Soft Rot
Pyrenochaeta sp.	0	1	0	0	0	0	Soft Rot
Cytospora sp.	0	1	0	0	0	0	Soft Rot
Trichoderma sp.	0	0	1	0	0	0	Soft Rot
Chalara sp.	0	0	0	0	1	0	Soft Rot
Xylaria cornu-damae	0	0	0	1	0	0	Soft Rot
Chaetomium aureum	0	1	0	0	0	0	Soft Rot
Trichoderma sp.	0	0	0	1	0	0	Soft Rot
Psathyrella lyckebodensis	0	0	0	18	13	0	NULL
Psathyrella sp.	0	0	0	6	3	1	NULL
Crepidotus sp.	1	1	0	0	0	1	NULL
Psathyrella romellii	0	0	0	0	2	0	NULL
Psathyrella candolleana	0	1	0	1	0	0	NULL
Crepidotus sp.	0	0	0	0	0	1	NULL
Simocybe sp.	1	0	0	0	0	0	NULL
Psathyrella tenuicula	1	0	0	0	0	0	NULL
Postia guttulata	0	0	0	957	22	0	Brown Rot
Leucogyrophana olivascens	0	589	0	0	0	0	Brown Rot
Serpula himantioides	3	69	0	0	0	0	Brown Rot
Dacryobolus sudans	9	0	0	0	0	0	Brown Rot
Coniophora olivacea	0	5	0	0	0	0	Brown Rot
Sarcoporia sp.	0	0	0	3	0	1	Brown Rot
Ditiola radicata	0	3	0	0	0	0	Brown Rot
Coniophora sp.	3	0	0	0	0	0	Brown Rot
Coniophora sp.	0	2	0	0	0	0	Brown Rot
Gloeophyllum trabeum	0	0	0	0	0	2	Brown Rot
Tapinella panuoides	0	2	0	0	0	0	Brown Rot
Antrodia carbonica	0	0	0	1	0	0	Brown Rot
Fibroporia radiculosa	1	0	0	0	0	0	Brown Rot
Oxyporus latemarginatus	0	0	0	0	0	1	Brown Rot

At both sites, higher abundance of wood decay fungi was noted for the treated areas compared to untreated forested areas. Note that counts in the cells are total reads, not number of OTUs. A relatively small number of small wood saprobes were recovered in our sampling.

There were several unique white rot fungal OTUs only present in the copper exposed areas in MS: *Trechispora,* both sp. and *T. invisitata*, *Heterobasidion* sp., *Ganoderma lucidum*, *Gymnopilus liquirtiae*, *Phlebia* sp., *Sistotrema* sp., and *Trichaptum* sp. Several OTUs identified as separate OTUs of *Trechispora* sp. were also found in WI, but were not limited to copper or mixed preservative exposed soils. *Ceriporiopsis* sp. was detected only in mixed preservative exposed soils in MS. *C. subvermispora* has been screened its ability to break down recalcitrant compounds in soil and liquid culture such as poly chlorinated biphenyls (PCBs) (Valentín et al., 2013), azoles (Woo et al., 2010) and pentachlorophenol (Machado et al., 2005) and is one of four fungi patented for use in mycoremediation of preservative treated wood waste (Lamar et al., 1995).

Most of the wood saprobes classified in the funguild analysis were classified as soft rot fungi. Many of these were cosmopolitan in their distribution among the different treatment exposures (ex. *Trichoderma* sp., *Scytalidium* sp., and *Humicola* sp.). Notable unique occurrences of soft rot fungi include: *Scytalidium lignicola* (detected only in MS copper treated), *Stilbella fimetaria* (detected only in WI mixed treatments). *Chaetomium aureum* was detected solely in copper treatments in MS and members of this genus have been used to pre-inoculate preservative treated wood blocks to initiate preservative breakdown (Duncan and Deverall, 1964).

CONCLUSION

Amplicon-based sequence analysis provides a valuable tool for detailed characterization of soil fungal communities. The results of this study suggest that long-term wood preservative exposure can impact soil community species composition and alter functional guild composition. Both site and treatment had significant effects on soil fungal community composition. Chemical analysis indicated little to no difference in metal concentrations compared to untreated soils, suggesting either leaching or complete breakdown of any remaining preservative residues in these studies. Exposure to the mixed treatment stimulated an increase in terrestrial ectomycorrhizal fungi, suggesting this guild of fungi may be more adapted to breakdown of residual wood preservatives than the wood decay fungi frequently used for bioremediation explorations. Wood decay fungi were often very specific to one site or the other, suggesting that local decay biota are present at a site, which may challenge the notion of generalized laboratory decay tests to determine efficacy of wood protectants.

AUTHOR CONTRIBUTIONS

GK was responsible for design of the experiments, data gathering and analysis, and preparation of the manuscript. WH assisted in the planning stages of the experiments, offered technical expertise for metagenomic analyses and provided input in the preparation of the manuscript. AB assisted in the planning stages of the experiments, processed soil samples for analysis and assisted in preparation of the manuscript. JP provided technical expertise on amplicon sequencing data analysis and assisted in preparation of the manuscript. MJ provided ecological statistics of sequencing data and assisted in preparation of the manuscript. DL provided assistance on design of the experiment and manuscript preparation.

ACKNOWLEDGMENTS

The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service. The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin. This article was written and prepared by U.S. Government employees on official time, and it is therefore in the public domain and not subject to copyright. Chemical analysis was performed by Kolby Hirth, Analytical Chemistry and Microscopy Laboratory (ACML) at the Forest Products Laboratory. Special thanks to Stan Lebow, Wood Durability and Protection work unit, for editorial input on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01997/full#supplementary-material

Supplementary Figure S1 | Taxonomic diversity of samples at the Phylum level.

Supplementary Figure S2 | Taxonomic diversity of samples at the Class level.

Supplementary Figure S3 | Taxonomic diversity of samples at the Order level.

Supplementary Figure S4 | Taxonomic diversity of samples at the Family level.

Supplementary Figure S5 | Taxonomic diversity of samples at the Genus level.

REFERENCES

Abarenkov, K., Henrik Nilsson, R., Larsson, K. H., Alexander, I. J., Eberhardt, U., Erland, S., et al. (2010). The UNITE database for molecular identification of fungi–recent updates and future perspectives. *New Phytol.* 186, 281–285. doi: 10.1111/j.1469-8137.2009.i03160.x

Abrego, N., and Salcedo, I. (2013). Variety of woody debris as the factor influencing wood-inhabiting fungal richness and assemblages: is it a question of quantity or quality? *For. Ecol. Manage.* 291, 377–385. doi: 10.1016/j.foreco.2012.11.025

Ahrén, D., Ursing, B. M., and Tunlid, A. (1998). Phylogeny of nematode-trapping fungi based on 18S rDNA sequences. *FEMS Microbiol. Lett.*, 158, 179–184. doi: 10.1016/S0378-1097(97)00519-3

Allegrucci, N., Cabello, M. N., and Arambarri, A. M. (2009). Diversity of saprotrophic anamorphic ascomycetes from native forests in Argentina: an updated review. *Darwiniana* 47, 108–124.

Assas, N., Marouani, L., and Hamdi, M. (2000). Scale down and optimization of olive mill wastewaters decolorization by Geotrichum candidum. *Bioproc. Eng.* 22, 503–507. doi: 10.1007/s004499900093

Association, A. W. P. (2007). *Book of Standards*. Birmingham, AL: AWPA.

Bhattacharya, P., Mukherjee, A. B., Jacks, G., and Nordqvist, S. (2002). Metal contamination at a wood preservation site: characterisation and experimental studies on remediation. *Sci. Total Environ.* 290, 165–180. doi: 10.1016/S0048-9697(01)01073-7

Bik, H. M., and Pitch Interactive (2014). Phinch: an interactive, exploratory data visualization framework for–Omic datasets. *bioRxiv*. doi: 10.1101/009944

Birkebak, J. M., Mayor, J. R., Ryberg, K. M., and Matheny, P. B. (2013). A systematic, morphological and ecological overview of the Clavariaceae (Agaricales). *Mycologia* 105, 896–911. doi: 10.3852/12-070

Cáceres, M. D., and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90, 3566–3574. doi: 10.1890/08-1823.1

Chase, J. M., Kraft, N. J., Smith, K. G., Vellend, M., and Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* 2, 1–11. doi: 10.1890/ES10-00117.1

Clausen, C. A., and Green, F. (2003). Oxalic acid overproduction by copper-tolerant brown-rot basidiomycetes on southern yellow pine treated with copper-based preservatives. *Int. Biodeterior. Biodegrad.* 51, 139–144. doi: 10.1016/S0964-8305(02)00098-7

Clausen, C. A., and Jenkins, K. M. (2011). *Chronicles of Fibroporia radiculosa (= Antrodia radiculosa) TFFH 294*. General Technical Report FPLGTR-204. U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 5, Madison, WI.

Colpaert, J. V. (2008). "Heavy metal pollution and genetic adaptations in ectomycorrhizal fungi," in *Stress in Yeasts and Filamentous fungi. British Mycological Society Symposia Series Vol. 27*, eds S. Avery, M. Stratford, and P. van West (Cambridge, MA: Academic Press), 157–173.

Connell, M. (1991). Industrial wood preservatives-the history, development, use, advantages and future trends.," in *The Chemistry of Wood Preservation*, ed R. Thompson (London: The Royal Society of Chemistry), 16–33.

Crous, P. W., and Groenewald, J. Z. (2013). A phylogenetic re-evaluation of Arthrinium. *IMA Fungus* 4, 133–154. doi: 10.5598/imafungus.2013.04.01.13

Daniel, R. (2005). The metagenomics of soil. *Nat. Rev. Microbiol.* 3, 470–478. doi: 10.1038/nrmicro1160

De Gannes, V., Eudoxie, G., and Hickey, W. J. (2013). Prokaryotic successions and diversity in composts as revealed by 454-pyrosequencing. *Bioresour. Technol.* 133, 573–580. doi: 10.1016/j.biortech.2013.01.138

De Groot, R. C., and Woodward, B. (1999). Using copper-tolerant fungi to biodegrade wood treated with copper-based preservatives. *Int. Biodeterior. Biodegrad.* 44, 17–27. doi: 10.1016/S0964-8305(99)00047-5

Dickie, I. A., Xu, B., and Koide, R. T. (2002). Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytol.* 156, 527–535. doi: 10.1046/j.1469-8137.2002.00535.x

Douhan, G., and Rizzo, D. (2005). Phylogenetic divergence in a local population of the ectomycorrhizal fungus *Cenococcum geophilum*. *New Phytol.* 166, 263–271.doi: 10.1111/j.1469-8137.2004.01305.x

Duncan, C. G., and Deverall, F. J. (1964). Degradation of wood preservatives by fungi. *Appl. Microbiol.* 12, 57–62.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., and Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, btv401. doi: 10.1093/bioinformatics/btv401

Evans, S., Henrici, A., and Ing, B. (2006). *The Red Data List of Threatened British Fungi*. Report by the British Mycological Society (BMS), Working With the Joint Nature Conservation Committee.

Ferraz, A., Córdova, A. M., and Machuca, A. (2003). Wood biodegradation and enzyme production by Ceriporiopsis subvermispora during solid-state fermentation of *Eucalyptus grandis*. *Enzyme Microb. Technol.* 32, 59–65. doi: 10.1016/S0141-0229(02)00267-3

Findley, K., Rodriguez-Carres, M., Metin, B., Kroiss, J., Fonseca, A., Vilgalys, R., et al. (2009). Phylogeny and phenotypic characterization of pathogenic Cryptococcus species and closely related saprobic taxa in the Tremellales. *Eukaryotic Cell* 8, 353–361. doi: 10.1128/EC.00373-08

Fomina, M., Alexander, I., Colpaert, J., and Gadd, G. (2005). Solubilization of toxic metal minerals and metal tolerance of mycorrhizal fungi. *Soil Biol. Biochem.* 37, 851–866. doi: 10.1016/j.soilbio.2004.10.013

Freeman, M. H., and Mcintyre, C. R. (2008). A comprehensive review of copper-based wood preservatives. *Forest Products J.* 58:6.

Gams, W. (2007). Biodiversity of soil-inhabiting fungi. *Biodivers. Conserv.* 16, 69–72. doi: 10.1007/s10531-006-9121-y

Giller, K. E., Witter, E., and Mcgrath, S. P. (1998). Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. *Soil Biol. Biochem.* 30, 1389–1414. doi: 10.1016/S0038-0717(97)00270-8

Green, F., and Clausen, C. A. (2005). Copper tolerance of brown-rot fungi: oxalic acid production in southern pine treated with arsenic-free preservatives. *Int. Biodeterior. Biodegrad.* 56, 75–79. doi: 10.1016/j.ibiod.2005.04.003

Hingston, J., Collins, C., Murphy, R., and Lester, J. (2001). Leaching of chromated copper arsenate wood preservatives: a review. *Environ. Pollut.* 111, 53–66. doi: 10.1016/S0269-7491(00)00030-0

Hughes, M. F. (2002). Arsenic toxicity and potential mechanisms of action. *Toxicol. Lett.* 133, 1–16. doi: 10.1016/S0378-4274(02)00084-X

Illman, B. L., and Yang, V. W. (2004). "Bioremediation and degradation of CCA-treated wood waste," in *Proceedings of Environmental Impacts of Preservative-Treated Wood* (Gainesville, FL), 8–11.

James, T. Y., Letcher, P. M., Longcore, J. E., Mozley-Standridge, S. E., Porter, D., Powell, M. J., et al. (2006). A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* 98, 860–871. doi: 10.1080/15572536.2006.11832616

Kandeler, F., Kampichler, C., and Horak, O. (1996). Influence of heavy metals on the functional diversity of soil microbial communities. *Biol. Fertil. Soils* 23, 299–306. doi: 10.1007/BF00335958

Khan, A., Kuek, C., Chaudhry, T., Khoo, C., and Hayes, W. (2000). Role of plants, mycorrhizae and phytochelators in heavy metal contaminated land remediation. *Chemosphere* 41, 197–207. doi: 10.1016/S0045-6535(99)00412-9

Kirker, G. T., Prewitt, M. L., Schultz, T. P., and Diehl, S. V. (2012). Community analysis of preservative-treated southern pine (*Pinus* spp.) using terminal restriction fragment length polymorphism (T-RFLP) analysis. Part 1: fungal field study. *Holzforschung* 66, 521–527. doi: 10.1515/hf.2011.171

Klepzig, K. D. (1998). Competition between a biological control fungus, Ophiostoma piliferum, and symbionts of the southern pine beetle. *Mycologia* 90, 69–75. doi: 10.2307/3761013

Lamar, R. T., Dietrich, D. M., and Glaser, J. A. (1995). *Solid Phase Bioremediation Methods Using Lignin-Degrading Fungi*. United States patent US. 5,476,788. The United States Of America As Represented By The Secretary Of Agriculture, The United States Of America As Represented By The Administrator of the Epa.

Lebow, S., Foster, D., and Evans, J. (2004). Long-term soil accumulation of chromium, copper, and arsenic adjacent to preservative-treated wood. *Bull. Environ. Contam. Toxicol.* 72, 225–232. doi: 10.1007/s00128-003-9055-y

López, M., Nichols, N., Dien, B., Moreno, J., and Bothast, R. (2004). Isolation of microorganisms for biological detoxification of lignocellulosic hydrolysates. *Appl. Microbiol. Biotechnol.* 64, 125–131. doi: 10.1007/s00253-003-1401-9

Machado, K. M. G., Matheus, D. R., Monteiro, R. T. R., and Bononi, V. L. R. (2005). Biodegradation of pentachorophenol by tropical basidiomycetes in soils contaminated with industrial residues. *World J. Microbiol. Biotechnol.* 21, 297–301. doi: 10.1007/s11274-004-3693-z

Meharg, A. A., and Cairney, J. W. G. (2000). Ectomycorrhizas—extending the capabilities of rhizosphere remediation? *Soil Biol. Biochem.* 32, 1475–1484. doi: 10.1016/S0038-0717(00)00076-6

Menkis, A., Urbina, H., James, T. Y., and Rosling, A. (2014). Archaeorhizomyces borealis sp. nov. and a sequence-based classification of related soil fungal species. *Fungal Biol.* 118, 943–955. doi: 10.1016/j.funbio.2014.08.005

Murillo, C., Albertazzi, F. J., Carranza, J., Lumbsch, H. T., and Tamayo, G. (2009). Molecular data indicate that *Rhytidhysteron rufulum* (ascomycetes, Patellariales) in Costa Rica consists of four distinct lineages corroborated by morphological and chemical characters. *Mycol. Res.* 113, 405–416. doi: 10.1016/j.mycres.2008.09.003

Nearing, M. M., Koch, I., and Reimer, K. J. (2015). Uptake and transformation of arsenic during the vegetative life stage of terrestrial fungi. *Environ. Pollut.* 197, 108–115. doi: 10.1016/j.envpol.2014.12.006

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fun. Ecol.* 20, 241–248. doi: 10.1016/j.funeco.2015.06.006

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'hara, R., et al. (2015). *Vegan: Community Ecology Package. 2015*. R Package Version 2.2–1.

Oliveira, A., and Pampulha, M. E. (2006). Effects of long-term heavy metal contamination on soil microbial characteristics. *J. Biosci. Bioeng.* 102, 157–161. doi: 10.1263/jbb.102.157

Paterson, R. R. M. (2006). *Ganoderma*–a therapeutic fungal biofactory. *Phytochemistry* 67, 1985–2001. doi: 10.1016/j.phytochem.2006.07.004

Porter, T. M., Schadt, C. W., Rizvi, L., Martin, A. P., Schmidt, S. K., Scott-Denton, L., et al. (2008). Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. *Mol. Phylogenet. Evol.* 46, 635–644. doi: 10.1016/j.ympev.2007.10.002

Powell, M. J., Letcher, P. M., and Longcore, J. E. (2011). Operculomyces is a new genus in the order Rhizophydiales. *Mycologia* 103, 854–862. doi: 10.3852/10-293

Rajapaksha, R., Tobor-Kapłon, M., and Bååth, E. (2004). Metal toxicity affects fungal and bacterial activities in soil differently. *Appl. Environ. Microbiol.* 70, 2966–2973. doi: 10.1128/AEM.70.5.2966-2973.2004

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/

Reid, D. A. (1964). Notes on some fungi of Michigan—I."Cyhellaceae." *Pers. Mol. Phylog. Evol. Fungi* 3, 97–154a.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. doi: 10.7717/peerj.2584

Rydlová, J., Sýkorová, Z., Slavíkov,á, R., and Turis, P. (2015). The importance of arbuscular mycorrhiza for *Cyclamen purpurascens* subsp. *immaculatum* endemic in Slovakia. *Mycorrhiza* 25, 599–609. doi: 10.1007/s00572-015-0634-7

Sakalidis, M., Slippers, B., Wingfield, B., Hardy, G. S. J., and Burgess, T. (2013). The challenge of understanding the origin, pathways and extent of fungal invasions: global populations of the *Neofusicoccum parvum–N. ribis* species complex. *Div. Distrib.* 19, 873–883. doi: 10.1111/ddi.12030

Scheffer, T. C. (1971). A climate index for estimating potential for decay in wood structures above ground. *Forest Products J.* 21, 25–31.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6241–6246. doi: 10.1073/pnas.1117018109

Schultz, T. P., and Nicholas, D. D. (2002). Development of environmentally-benign wood preservatives based on the combination of organic biocides with antioxidants and metal chelators. *Phytochemistry* 61, 555–560. doi: 10.1016/S0031-9422(02)00267-4

Schultz, T. P., Nicholas, D. D., and Preston, A. F. (2007). A brief review of the past, present and future of wood preservation. *Pest Manag. Sci.* 63, 784–788. doi: 10.1002/ps.1386

Seitzman, B. H., Ouimette, A., Mixon, R. L., Hobbie, E. A., and Hibbett, D. S. (2010). Conservation of biotrophy in Hygrophoraceae inferred from combined stable isotope and phylogenetic analyses. *Mycologia* 103, 280–290. doi: 10.3852/10-195

Shetty, K., Hetrick, B., Figge, D., and Schwab, A. (1994). Effects of mycorrhizae and other soil microbes on revegetation of heavy metal contaminated mine spoil. *Environ. Pollut.* 86, 181–188. doi: 10.1016/0269-7491(94)90189-9

Sugihara, A., Shimada, Y., and Tominaga, Y. (1990). Separation and characterization of two molecular forms of *Geotrichum candidum* lipase. *J. Biochem.* 107, 426–430. doi: 10.1093/oxfordjournals.jbchem.a123061

Sun, B.-S., Zheng, L.-L., Liu, T.-M., Wang, H., and Xu, M.-X. (2008). Applied reasearch progress in geotrichum candidum [J]. *Food Res. Develop.* 9:51.

Svenson, S. E., Davies, F. T., and Meier, C. E. (1991). Ectomycorrhizae and drought acclimation influence water relations and growth of loblolly pine. *HortScience* 26, 1406–1409.

Team, R. C. (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2014.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., et al. (2014). Global diversity and geography of soil fungi. *Science* 346:1256688. doi: 10.1126/science.1256688

Tornberg, K., Bååth, E., and Olsson, S. (2003). Fungal growth and effects of different wood decomposing fungi on the indigenous bacterial community of polluted and unpolluted soils. *Biol. Fert. Soils* 37, 190–197. doi: 10.1007/s00374-002-0574-1

Untereiner, W. A., Straus, N. A., and Malloch, D. (1995). A molecular-morphotaxonomic approach to the systematics of the Herpotrichiellaceae and allied black yeasts. *Mycol. Res.* 99, 897–913. doi: 10.1016/S0953-7562(09)80748-X

Valentín, L., Oesch-Kuisma, H., Steffen, K. T., Kähkönen, M. A., Hatakka, A., and Tuomela, M. (2013). Mycoremediation of wood and soil from an old sawmill area contaminated for decades. *J. Hazard. Mater.* 260, 668–675. doi: 10.1016/j.jhazmat.2013.06.014

Vizzini, A., Angelini, C., and Ercole, E. (2012). A new Neopaxillus species (Agaricomycetes) from the Dominican Republic and the status of Neopaxillus within the Agaricales. *Mycologia* 104, 138–147. doi: 10.3852/10-345

Voriskova, J., and Baldrian, P. (2013). Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME J.* 7, 477–486. doi: 10.1038/ismej.2012.116

Wang, Y., Xu, L., Gu, Y. Q., and Coleman-Derr, D. (2016). MetaCoMET: a web platform for discovery and visualization of the core microbiome. *Bioinformatics* 32, 3469–3470. doi: 10.1093/bioinformatics/btw507

Warcup, J. (1990). Occurrence of ectomycorrhizal and saprophytic discomycetes after a wild fire in a eucalypt forest. *Mycol. Res.* 94, 1065–1069. doi: 10.1016/S0953-7562(09)81334-8

Warcup, J., and Talbot, P. (1962). Ecology and identity of mycelia isolated from soil. *Trans. Br. Mycol. Soc.* 45, 495–518. doi: 10.1016/S0007-1536(62)80010-2

Watling, R., and Richardson, M. (2010). Coprophilous fungi of the Falkland Islands. *Edinb. J. Bot.* 67, 399–423. doi: 10.1017/S0960428610000156

Watkinson, S. C., and Eastwood, D. C. (2012). "Chapter 5 *Serpula lacrymans*, wood and buildings," in *Advances in Applied Microbiology,* Vol. 78 (London: Academic Press), 121–149.

Weiss, M., Selosse, M.-A., Rexer, K.-H., Urban, A., and Oberwinkler, F. (2004). Sebacinales: a hitherto overlooked cosm of heterobasidiomycetes with a broad mycorrhizal potential. *Mycol. Res.* 108, 1003–1010. doi: 10.1017/S0953756204000772

Whipps, J., and Gerlagh, M. (1992). Biology of Coniothyrium minitans and its potential for use in disease biocontrol. *Mycol. Res.* 96, 897–907. doi: 10.1016/S0953-7562(09)80588-1

Woo, C., Daniels, B., Stirling, R., and Morris, P. (2010). Tebuconazole and propiconazole tolerance and possible degradation by Basidiomycetes: a wood-based bioassay. *Int. Biodeter. Biodegr.* 64, 403–408. doi: 10.1016/j.ibiod.2010.01.009

Zellweger, F., Braunisch, V., Morsdorf, F., Baltensweiler, A., Abegg, M., Roth, T., et al. (2015). Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests. *For. Ecol. Manage.* 349, 36–44. doi: 10.1016/j.foreco.2015.04.008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Kirker, Bishell, Jusino, Palmer, Hickey and Lindner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 November 2017
doi: 10.3389/fmicb.2017.02272

Quantitative Detection of Active Vibrios Associated with White Plague Disease in *Mussismilia braziliensis* Corals

Luciane A. Chimetto Tonon[1, 2, 3], Janelle R. Thompson[2]*, Ana P. B. Moreira[3], Gizele D. Garcia[4], Kevin Penn[2], Rachelle Lim[2], Roberto G. S. Berlinck[1], Cristiane C. Thompson[3] and Fabiano L. Thompson[3]**

[1] Laboratory of Organic Chemistry of Biological Systems, Chemical Institute of São Carlos, University of São Paulo, São Carlos, Brazil, [2] Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States, [3] Laboratory of Microbiology, Institute of Biology, SAGE-COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, [4] Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Julie L. Meyer, University of Florida, United States
Aldo Cróquer, Simón Bolívar University, Venezuela

***Correspondence:**
Luciane A. Chimetto Tonon
luciane.chimetto@gmail.com
Janelle R. Thompson
janelle@mit.edu
Fabiano L. Thompson
fabianothompson1@gmail.com

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *22 August 2017*
Accepted: *03 November 2017*
Published: *17 November 2017*

Citation:
Chimetto Tonon LA, Thompson JR, Moreira APB, Garcia GD, Penn K, Lim R, Berlinck RGS, Thompson CC and Thompson FL (2017) Quantitative Detection of Active Vibrios Associated with White Plague Disease in Mussismilia braziliensis Corals. Front. Microbiol. 8:2272. doi: 10.3389/fmicb.2017.02272

Over recent decades several coral diseases have been reported as a significant threat to coral reef ecosystems causing the decline of corals cover and diversity around the world. The development of techniques that improve the ability to detect and quantify microbial agents involved in coral disease will aid in the elucidation of disease cause, facilitating coral disease detection and diagnosis, identification and pathogen monitoring, pathogen sources, vectors, and reservoirs. The genus *Vibrio* is known to harbor pathogenic strains to marine organisms. One of the best-characterized coral pathogens is *Vibrio coralliilyticus*, an aetilogic agent of White Plague Disease (WPD). We used *Mussismilia* coral tissue (healthy and diseased specimens) to develop a rapid reproducible detection system for vibrios based on RT-QPCR and SYBR chemistry. We were able to detect total vibrios in expressed RNA targeting the 16S rRNA gene at 5.23×10^6 copies/μg RNA and *V. coralliilyticus* targeting the *pyrH* gene at 5.10×10^3 copies/μg RNA in coral tissue. Detection of *V. coralliilyticus* in diseased and in healthy samples suggests that WPD in the Abrolhos Bank may be caused by a consortium of microorganism and not only a single pathogen. We developed a more practical and economic system compared with probe uses for the real-time detection and quantification of vibrios from coral tissues by using the 16S rRNA and pyrH gene. This qPCR assay is a reliable tool for the monitoring of coral pathogens, and can be useful to prevent, control, or reduce impacts in this ecosystem.

Keywords: *Vibrio coralliilyticus*, *Mussismilia braziliensis*, *pyrH* gene, reef health monitoring, marine biology, biodiversity, microbiology

INTRODUCTION

Despite the undeniable importance of coral reefs around the World, they are undergoing massive extinction due to anthropogenic (e.g., overfishing and pollution) and global impacts (e.g., infectious diseases, ocean warming, and acidification; De'ath et al., 2009, 2012). A remarkable reduction in coral cover has been reported in both the Pacific and Caribbean reefs. In the south Atlantic, the

coral reefs in Abrolhos constitute the most extensive and richest reefs in Brazilian waters (Leão and Kikuchi, 2001; Freitas et al., 2011; Osinga et al., 2011). *Mussismilia* species, the main Abrolhos reef builders, have suffered massively from white plague disease (WPD; Francini-Filho et al., 2010; McDole et al., 2012; Garcia et al., 2013). Corals affected by WPD show a pronounced line of bright, white tissue that separates the colored (living) part of the coral from bare, rapidly algal-colonized skeleton (Richardson et al., 2001). A range of different white diseases including WPD (Richardson et al., 1998a, 2001), white band (Aronson and Precht, 2001) and "shutdown reaction" (Antonius, 1981; Bythell et al., 2004) are widespread and collectively referred to as white syndromes (Bythell et al., 2004). On Indo-Pacific scleractinian corals a spreading band of tissue loss exposing the white skeleton is a symptom indicative for the classification of White Syndrome (WS; Willis et al., 2004).

Vibrio coralliilyticus has been identified as the causative agent of white plague disease in several Pacific reefs (Sussman et al., 2008, 2009). In hard corals, infection by *V. coralliilyticus* causes loss of *Symbiodinium* cells from the coenosarc tissue (the live tissue between polyps) and subsequent tissue loss exposing the white skeleton (Sussman et al., 2008, 2009). *V. coralliilyticus* P1 had high proteolytic activity as a result of the secretion of a set of proteases, including a Zinc-metalloprotease that plays an important role in the cleavage of connective tissue and other cellular perturbations (Santos Ede et al., 2011). *Vibrio harveyi* has also been considered as an important causal agent of WPD in tropical stony corals (Woodley et al., 2015). Luna et al. (2010), confirmed through Koch's postulates, the involvement of *V. harveyi* in the development of WPD. The inoculation of *V. harveyi* strains in healthy colonies of *Pocillopora damicornis* induced the disease and tissue lysis. *V. harveyi* and *V. coralliilyticus* are also pathogenic for many marine organisms, such as fishes (e.g., flounders, groupers, sharks, seabream, seabass, and turbots; Gauger et al., 2006), mollusks, and prawns (Nicolas et al., 2002; Alavandi et al., 2006).

Detection of vibrios is an important tool to understand disease ecology. PCR-based diagnostic methods are rapid, specific and sensitive for the detection of vibrios, and it is able to be applied in a management situation, monitoring specific pathogens (Goarant and Merien, 2006; Pollock et al., 2010; Wilson et al., 2013; Ahmed et al., 2015; Garrido-Maestu et al., 2016). The qPCR technologies are based on two categories of fluorescence chemistries (oligonucleotide-specific probes and intercalating dyes). The cost of oligonucleotide probe technologies, including TaqMan and Molecular Beacon, can be very high limiting wide adoption. Intercalating dye technologies, such as SYBR Green, fluoresce as they anneal to the double-stranded DNA (dsDNA) that is synthesized during PCR amplification. Intercalating dyes are less expensive and work with traditional PCR primer sets, negating the time and labor- intensive design of specific probes (Pollock et al., 2011). Several qPCR assays have been described for the detection of pathogenic vibrio species based in Taqman chemistry: targeting *Vibrio aestuarianus* (Saulnier et al., 2009; McCleary and Henshilwood, 2015), *V. harveyi* (Schikorski et al., 2013), *Vibrio tapetis* (Bidault et al., 2015), and *V. coralliilyticus*, targeting the *dnaJ* gene and employing seeded seawater and seeded coral tissue (Pollock et al., 2010). *V. parahaemolyticus, Vibrio vulnificus,* and *V. cholerae* were detected in Galician mussels by using multiplex qPCR (Garrido-Maestu et al., 2016). *Vibrio alginolyticus* was detected in shellfish and shrimp using SYBR Green I chemistry targeting the groEL (Ahmed et al., 2015). The use of SYBR is cheaper for reagents and equipment, but deserve attention to confirm specificity.

Coral associated microbial communities are very diverse (Wegley et al., 2007; Fernando et al., 2015). We have previously shown that a majority of bacterial OTUs have a low prevalence among individual colonies suggesting dynamic assemblages (Fernando et al., 2015) which may be influenced by populations introduced via the water or through ingestion of food particles (Thompson et al., 2014). Such populations may have variable levels of activity including populations that are inactive or dead due to stresses associated with the coral host and populations that are activity metabolizing and growing in the holobiont.

Recently, it has been introduced the idea that the coral meta-organism or holobiont hosts a microbiome with distinct microbial sub-communities, including (1) a ubiquitous and stable core microbiome (consisting of very few symbiotic host-selected microbiota), (2) a microbiome of spatially and/or regionally explicit core microbes each filling functional niches, and (3) a highly variable microbial community that is responsive to biotic and abiotic processes across spatial and temporal scales (Hernandez-Agreda et al., 2016; Sweet and Bulling, 2017). Many biotic and abiotic factors (e.g., algal competition, age of the colony, temperature, pH, nutrients, light, dissolved organic carbon, etc.) can affect the composition of the microbiome in corals (Bourne and Webster, 2013; Hernandez-Agreda et al., 2016; Sweet and Bulling, 2017). For instance, when the dynamic of the microbiome is disrupted in response to stress, pathogenic microbes, and their relationship with the "normal" microbiome of the organism, may influence or drive disease processes in the holobiont. This complex interaction of pathogenic microbes has been named as pathobiome (Vayssier-Taussat et al., 2014; Sweet and Bulling, 2017).

For better understanding of active fraction of the microbiota that may mediate symptoms associated with WS we chose to develop an RT-qPCR protocol. The 16S rRNA gene was selected to quantify total vibrios due to the availability of validated group-specific primers and the expectation that the abundance of 16S rRNA copies in extracted RNA would be indicative of metabolic activity and growth of the targeted group in coral tissue. Since several disease-causing vibrios are closely-related and difficult to resolve via the 16S rRNA gene sequence, we also developed species-specific PCR primers targeting the *pyrH* gene, which resolves species-level differences among the vibrios (Thompson et al., 2005). The *pyrH* gene encodes the housekeeping gene uridine monophosphate kinase (UMP kinase) that participates in pyrimidine biosynthesis catalyzing the conversion of UMP into UDP (Voet and Voet, 2004). This gene has been identified as a highly-expressed colonization factor in *Vibrio* species such as *V. vulnificus* and *V. harveyi* (Kim et al., 2003; Lee et al., 2007; Guerrero-Ferreira and Nishiguchi, 2010). *PyrH* expression by vibrios is expected during coral colonization. Therefore, we tested the hypotheses that (i) diseased corals have a higher proportion

of active vibrio counts (detected by RT-qPCR) than healthy corals (Hp1), and; (ii) Diseased coral has more *Vibrio coralliilyticus* and/or *V. harveyi* than healthy corals (Hp2).

The aim of the present study was to develop a rapid and reproducible tool for the detection of vibrios, *V. coralliilyticus* and *V. harveyi* in coral tissue. We used the method to investigate the presence and activity of vibrio species, including the *V. coralliilyticus* and *V. harveyi* in healthy and diseased coral samples from the Abrolhos Bank. Diseased samples displayed symptoms corresponding to those described by Work and Aeby (2006) as typical of WPD, i.e., tissue loss, distinctly separated from intact tissue and revealing an intact skeleton.

Here we developed a practical and economic system for the real-time detection and quantification of vibrios from coral tissues by using the 16S rRNA and *pyrH* gene.

MATERIALS AND METHODS

Bacterial Strains: Culture Conditions and Genomic DNA Extraction

For use as standards during assay development cultures of *Vibrio* strains including *V. coralliilyticus* (CAIM 616^T; YB1^T; LMG 21349; LMG 10953), *V. neptunius* (INCO17^T; RFT5), *V. tubiashii* (LMG 10936^T), *V. harveyi* (R-246; B-392), *V. parahaemolyticus* (R-241), *V. communis* (R-233^T), and *V. campbellii* (HY01) were grown overnight in Marine Agar at 28°C. Total genomic DNA was extracted by using the Wizard Genomic DNA Kit (Promega, Madison, Wisconsin, USA) according to the manufacturer's instructions.

Site Description and Coral Collection

Abrolhos Bank (AB) is located in the southwestern Atlantic Ocean (SAO). Fragments of *M. braziliensis* (healthy and white plague diseased) were sampled in August 2011 and Februrary 2012 in Parcel dos Abrolhos (17°57′32.7″/ 38°30′20.3″), located off-shore (∼70 km), inside the protected area Abrolhos Marine National Park. Four white plague (D1–D4) plus two healthy samples (H1–H2) were collected in winter (08/18/2011); and five white plague (D5–D9) plus three healthy samples (H3–H5) were collected in summer (02/29/2012; **Table 1**). Sampling was performed by scuba diving with a hammer and a chisel. Fragments were immediately stored in polypropylene tubes, identified, and frozen in liquid nitrogen.

RNA Extraction

The RNA was obtained by macerating 100 mg of coral tissue in small particles, until it turns to dust, in a crucible using pistillate and liquid N_2. Trizol (1 mL) was added and mixed by vortexing for 2 s. Tubes were kept at room temperature for 5 min and then 200 uL of chloroform was added and the solution was mixed for 15 s by shaking hands. Tubes were incubated at room temperature for 3 min and then centrifuged at 2,000 rpm for 15 min. The clear phase was transferred to a clean tube and the purification step was performed with RNeasy® mini kit (Qiagen group) as of step 5 according to the manufacturer's protocols. The samples were kept on ice during extraction procedures.

TABLE 1 | Sample data information used in this study.

Sample number	Sampling place	Sample type	Status	Coral species	Sampling data
D1	Abrolhos bank	RNA	Diseased (white plague)	*Mussismilia brasiliensis*	19/08/2011
D2					19/08/2011
D3					19/08/2011
D4					19/08/2011
D5	Parcel dos Abrolhos				29/02/2012
D6					29/02/2012
D7					29/02/2012
D8					29/02/2012
D9					29/02/2012
H1	Abrolhos bank		Healthy		19/08/2011
H2					19/08/2011
H3	Parcel dos Abrolhos				29/02/2012
H4					29/02/2012
H5					29/02/2012

With this protocol we obtained the average number of 23 μg of RNA (SE = ±9.4) per 100 mg of coral tissue. The samples were stored at −80°C. All samples were resuspended in water electrophoresed on 1% agarose gel (Figure S1), and quantified on a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) to evaluate quality and purity. All samples analyzed are listed (**Table 1**).

cDNA Synthesis

Reverse transcription was performed with the QuantiTect Reverse Transcription Kit—QIAGEN, according to the manufacturer's protocols with minor modifications. A DNAse step was performed and random pentadecamer primers were used to generate the initial cDNA strands (Stangegaard et al., 2006), which were quantified using a NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Primer Design

Primers for *Vibrio* detection were designed in MEGA 5 (Tamura et al., 2011) against a multiple alignment of *Vibrio* sequences from public and personal databases (**Table 2**, Figure S2, Support Material 1). Conserved regions were selected and putative primer combinations were tested *in silico* by using the NCBI tool Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). We used the reliable taxonomic marker *pyrH* as target, which has higher discriminatory power than the 16S rRNA gene and allows the distinction of closely related *Vibrio* species (Thompson et al., 2005; Sawabe et al., 2007). The best combination of primer pairs, indicated through *in silico* tests, were analyzed by PCR.

Predicted amplicon size was verified through conventional PCR using the Amplitaq Gold 360 Master Mix (Applied Biosystems) in a total volume of 25 μL reaction and primers

TABLE 2 | Primer information.

CODE	Sequence 5′-3′	Tm (°C)	Amplicon (bp)	Gene target	Phylogenetic target
567F	GGC GTA AAG CGC ATG CAG GT	61.4	114	16S rRNA (Thompson et al., 2004)	*Vibrio* spp.
680R	GAA ATT CTA CCC CCC TCT ACA G	54.3			
Vc pyrH F	CAA CTG GGC AGA CGC AAT CCG TGA GT	64.8	166	*pyrH* (this study)	*V. coralliilyticus*
Vc pyrH R	CGT AAA TAC GCC ATC AAC TTT TGT C	54.9			
Vh pyrH F	CTT GCA ACG GTA ATG AAC GGT TTG GCA	61.9	171	*pyrH* (this study)	*V. harveyi*
Vh pyrH R	AGT ACC TGC AGA GAA GAT TAC CAC T	57.1			

at 10 μM. The program was 95°C–5 min, 40 cycles (95°C–30 s, range 50°C up to 68°C–1 min, 72°C–1 min), and 72°C–5 min. DNA template of the targets *V. coralliilyticus* (CAIM 616^T; YB1^T; LMG 21349; LMG 10953), *V. harveyi* (R-246; B-392) and phylogenetically closely related *Vibrio* species, *V. communis* (R-233^T), *V. parahaemolyticus* (R-241), *V. campbellii* (HY01), *V. neptunius* (INCO17^T; RFT5), and *V. tubiashii* (LMG 10936^T) were included in each reaction. Primer dimerization was checked in 1% agarose gel. Primer sequences selected for this study were Vc_pyrHF (5′- CAA CTG GGC AGA CGC AAT CCG TGA GT-3′) and Vc_pyrHR (5′-CGT AAA TAC GCC ATC AAC TTT TGT C-3′) to quantify *V. coralliilyticus* and Vh_pyrHF (5′- CTT GCA ACG GTA ATG AAC GGT TTG GCA-3′) and Vh_pyrHR (5′-AGT ACC TGC AGA GAA GAT TAC CAC T-3′) to quantify *V. harveyi*. In addition, 16S rRNA gene primer pair 567F (5′-GGC GTA AAG CGC ATG CAG GT-3′) and 680R (5′-GAA ATT CTA CCCCCC TCT ACA G-3′; Thompson et al., 2004) were applied to quantify the total number of vibrio sequences per sample (**Table 2**).

Quantitative Polymerase Chain Reaction (qPCR)

Total vibrios, *V. coralliilyticus* and *V. harveyi* were quantified by qPCR using primer pairs 567F-−680R, Vc_pyrHF—Vc_pyrHR and Vh_pyrHF—Vh_pyrHR, respectively, in a LightCycler_480 Real-Time PCR system with software v. 1.5.0 (Roche Applied Sciences, Indianapolis, IN, USA) for calculation of crossing point (Cp) values and melting temperature (T_m) analysis. The correct size of the amplicons obtained (114-bp 16S rRNA, 166-bp *pyrH,* 171-bp *pyrH* gene fragments) was confirmed by 1% agarose gel electrophoresis. qPCR reaction mixtures consisted of 10 μL of KAPA SYBR_ FAST 2X Master Mix (KAPABIOSYSTEMS, Woburn, MA, USA), 10 μM of each primer and 1 μL of cDNA template (20 ng/uL). Amplification followed the manufacturer's instructions. Briefly, reactions were subjected to a preincubation step of 95°C for 3 min, followed by 50 cycles of 95°C for 10 s, 58°C (16S primers), 60°C (Vc-*pyrH* primers) or 62°C (Vh-*pyrH* primers) for 20 s and 72°C for 1 s. Each sample was analyzed in triplicate, and Cp values were examined after amplification to verify consistency (i.e., coefficient of variation ≤ 3%). To confirm the specificity of amplification, melting temperatures (T_m) of sample amplicons were confirmed to be within two standard deviations of the mean T_m associated with qPCR standards at concentrations of 10^1–10^7 copies per qPCR (82.65°C ± SD 0.097). The fluorescence history, melting- curves, and peaks can be visualized in Figure S3.

Standards for qPCR were prepared by dilution of genomic DNA. Ten-fold serial dilutions of genomic DNA from *V. coralliilyticus*, *V. harveyi,* and total vibrio represented by *V. neptunius* were used to generate the standard curves (SCs; Figure S4). Genomic DNA concentration was measured by nanodrop method. Gene copy number was determined based on reported genome size and gene frequency within each genome. A genome size of 3.5 Mbp (*V. harveyi*) and 5.68 Mbp (*V. coralliilyticus*) were used, corresponding to a single copy of the *pyrH* gene. Moreover, a genome size of 5 Mbp (*V. neptunius*) was used, corresponding to 10 copies of the 16S rRNA gene.

For calculation of the standard curve Cp values were plotted against Log10 of computed gene copies/uL from genomic DNA added to each qPCR run using at least-squares fit. Confidence intervals for the predicted target concentrations based on measured Cp values were calculated based on the propagation of error in the SC (Harris, 1995). The limit of detection (LOD) was determined based on the uncertainty in the SC as the upper 99th per cent confidence interval of the Cp values of the negative controls or 50 cycles if no signal was apparent (Nshimyimana et al., 2014). For consistency in statistical analysis, the highest LOD used to indicate non-detectable target was selected as the study-wide LOD. The amplification efficiency (E) for each qPCR run was calculated from the slope of the standard curve and was considered consistent in the range of 89–100%. We determined the inhibition impact on qPCR by spiking the positive control DNA diluted 1:100, into an aliquot from each sample before qPCR amplification and by comparing the qPCR results measured for samples with and without spike addition. If the spiked sample was quantified as having <65% of the added amount of positive control (corresponding to both the 95% confidence interval for quantification of the qPCR standard curve and observed variability between technical replicates), then the sample was diluted 10-fold and re-analyzed (Nshimyimana et al., 2014).

To estimate the back-of-the-envelope conversion between copies per μg RNA and cells per ml, we used the average number of ribosomes per vibrio cell (that is 10) as determined by Lee et al. (2009), and the average number of μg of RNA (23 μg) extracted from 100 mg of coral tissue. Considering the density of coral tissue 1.15 g/mL (https://www.aqua-calc.com/page/

density-table), which is close to the seawater 1.02 g/mL (https://www.aqua-calc.com/page/density-table/substance/seawater), we have 100 mg of coral tissue corresponds to 0.08 mL. So, in 1mL we found 287.5 μg RNA. Now, we have the formula: Cells/mL = {[n°qPCR copies/μgRNA x $\bar{x}$ (μgRNA/mL)] / $\bar{x}$ (n° of ribosomes vibrio cell)}.

To verify primer specificity in a closely-related non-target background target genomic DNA from *V. harveyi* or *V. coralliilyticus* were mixed with 10- to 100-fold excess of non-target DNA from *V. campbellii* (HY01), *V. communis* (R-233^T), *V. neptunius* (INCO17^T), *V. tubiashii* (LMG 10936^T), or *E. coli*. All tested DNA were initially diluted to equimolar concentrations (100 ng/ul), then serial dilutions were performed and non-target DNA were kept 10-fold excess compared with target DNA mixed in the solution. Mixtures were quantified by QPCR as described above. If primers were specific for the target, then the observed concentrations would match the expected concentration of target DNA. In contrast, higher than expected observed concentrations would indicate non-specific amplification of closely related targets.

The percentage of the cross-reactivity specificity obtained from the target was calculated according to this equation:

$$\left(\frac{\text{qPCR copies of mixed sample}_{(\text{non-target 1/10 plus target 1/100})} - \text{qPCR copies of non} - \text{target}_{(\text{non-target 1/10})}}{\text{qPCR copies of mixed sample}_{(\text{non-target 1/10 plus target 1/100})}}\right) \times 100$$

Statistical Analyses

We tested the hypotheses that diseased corals have more total vibrios (Hp1) and more *V. coralliilyticus*/*V. harveyi* (Hp2) than health corals by Welch's *t*-test with the GraphPad software (https://www.graphpad.com/). We examined the interaction between health state and sampling season by two-way ANOVA using the OriginPro 8 SR0 software (Origin Lab Corporation). We tested the correlation between the abundance of all vibrios and the abundance of *V. coralliilyticus* using Spearman's Rho Correlation (http://www.socscistatistics.com/tests/Default.aspx).

RESULTS

Primer Specificity

The new primer pairs designed for the *pyrH* gene of *V. coralliilyticus* and *V. harveyi* showed specific match to the targeted species (Figures S5, S6). A single amplicon was observed when the target DNA was added whilst no amplification was observed when closely phylogenetic *Vibrio* species (i.e., *V. neptunius* INCO17^T and *V. tubiashii* LMG 10936^T) were tested in standard PCR (Figure S5). This result confirmed the *in silico* selection of the primer pairs (Vc_pyrHF and Vc_pyrHR) for *V. coralliilyticus* LMG 20984^T, CAIM 616^T (Accession Number GU266292). The same specificity was observed for the primer pairs Vh_pyrHF and Vh_pyrHR tested against *V. harveyi* R-246 (Accession Number EU251625), and closely related species *V. communis* (R-233^T) and *V. parahaemolyticus* (R-241). Only the DNA fragment from the target could be amplified (171 bp) but not from the phylogenetic closely related species (Figure S6). Comparing the 16S rRNA sequence similarity among these *Vibrio* species we found that *V. coralliilyticus* shares up 98% of 16S rRNA sequence similarity with *V. neptunius* and *V. tubiashii*. while *V. harveyi* shares up 99% of 16S rRNA sequence similarity with *V. communis* and *V. parahaemolyticus*. The amplicons obtained for the genes tested were checked through sequencing and the specificity of this assay for the target strains was confirmed.

To determine the specificity of the assay in a background of closely-related non-target DNA qPCR analyses were performed on a mixture of non-target plus target DNA and up to 100% of the detection signal was due to the target DNA amplification. We observed that the values detected, represented basically only target DNA by qPCR (**Figure 1**). The number of copies obtained to the specific target species was very close to the observed in our assay. It means that the primers keep bind to the target species even when we have at least 10 times more concentrated non-target species phylogenetically closely related in the same sample. The percentage of the cross-reactivity specificity calculated showed that 100% of specificity was achieved when the target *V. harveyi* (1/100) was mixed with *V. communis* (1/10) and 80% when mixed with *V. campbellii* (1/10); 100% of the detection signal was achieved when the target *V. coralliilyticus* (1/100) was mixed with *V. tubiashii* (1/10) and 99% when mixed with *V. neptunius* (1/10); 99% when mixing *E. coli* (1/10) with *Vibrio* spp. (1/100) and 94% when *Vibrio* spp. (1/1,000) was mixed with *E. coli* (1/10).

Total Vibrios

Expressed *Vibrio* 16S rRNA gene varied from 134 to 2,750 copies per μg of RNA in healthy corals and from 3,000 to 5,230,000 per μg of RNA in diseased corals (**Figure 2**). Which represents the detection range of 3.85×10^3 to 7.91×10^4 cells/mL in health corals and 8.63×10^4 to 1.5×10^8 cells/mL in diseased corals. The load of total vibrios was significantly higher in diseased corals than in health corals (Welch's *t*-test, *p*-value = 0.0091). Two samples D5 and D7, collected during the summer sampling campaign showed over two orders of magnitude higher *Vibrio* RNA concentration compared to the other diseased or healthy corals. Nevertheless, the interaction between health state (health/disease) and sampling season (summer/winter) by two-way ANOVA was not significant.

V. harveyi and *V. coralliilyticus*

Expressed *V. harveyi* RNA was not detected in the coral samples by RT-qPCR of the *pyrH* gene. On the other hand, *V. coralliilyticus* was detected in all samples analyzed with exception of D3 (**Figure 2**). Expressed *V. coralliilyticus pyrH* gene varied from 658 to 1,160 copies per μg of RNA in healthy corals and from 973 to 5,100 per μg of RNA in diseased corals. The load of *V. coralliilyticus* was significantly higher in diseased corals than in health corals (Welch's *t*-test, $p = 0.0194$). However, the interaction between health state and sampling season by two-way

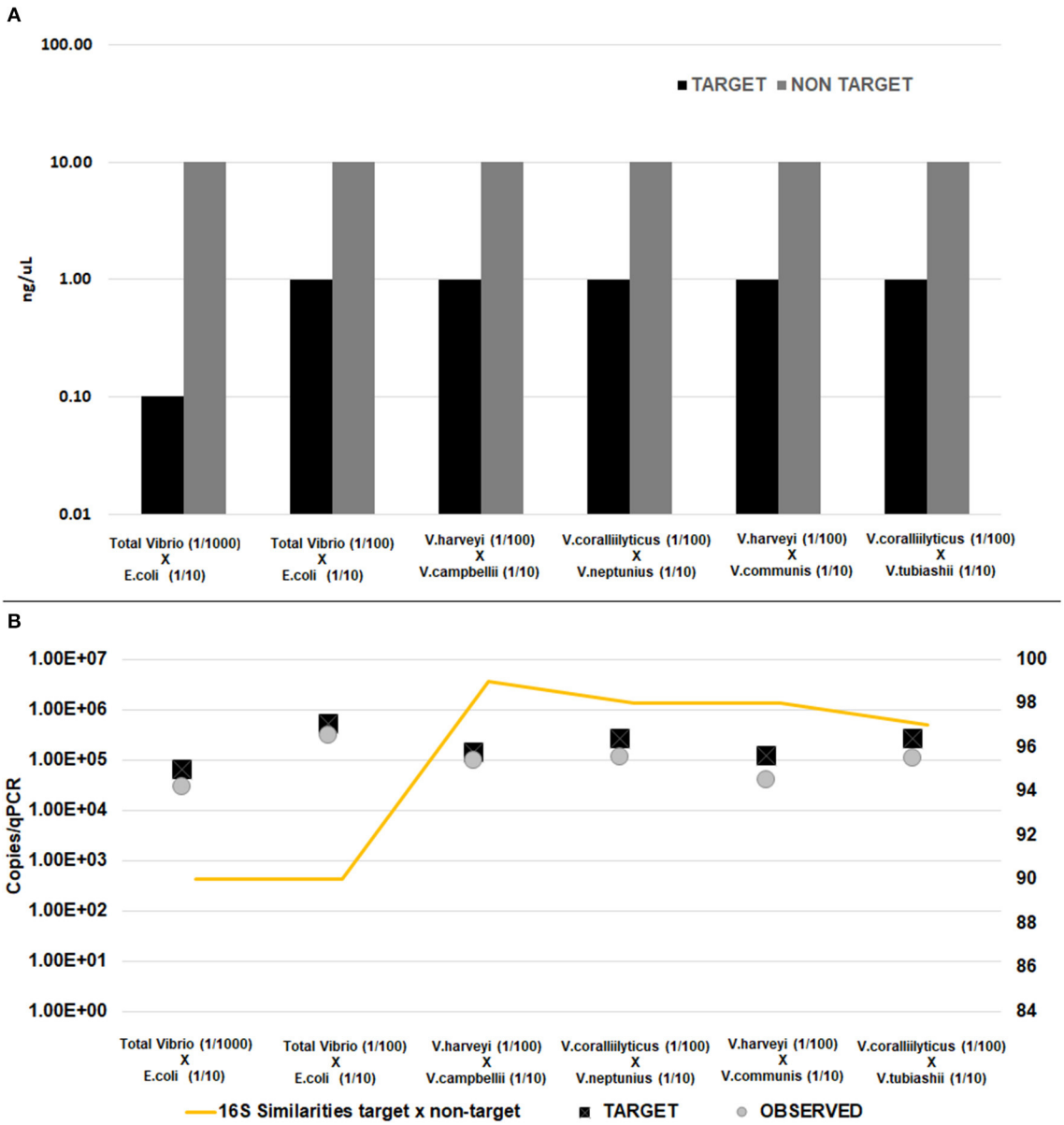


FIGURE 1 | Amplification of target DNA in a background of closed related non-target. **(A)** Target and non-target DNA concentration used in the mixed reaction. **(B)** Copies detected by qPCR. All non-target DNA had 10-fold excess compare with target DNA. Target DNA i.e. (total *Vibrio*, *V. harveyi,* and *V. coralliilyticus*) were diluted at least 1:100 while non-target (i.e., *E. coli*, *V. campbellii*, *V. communis*, *V. tubiashii,* and *V. neptunius*) were diluted 1:10. Similarities of 16S rRNA between target and non-target are represented in line.

ANOVA was not significant. Copy numbers of *V. coralliilyticus* and total *Vibrio* 16S rRNA were positively correlated (Spearman's Rho Correlation, $p = 0.00056$).

DISCUSSION

Quantification of *Vibrio* populations has been investigated in the last decades through culture-independent methods targeting DNA and whole cells (Eilers et al., 2000; Heidelberg et al., 2002; Thompson et al., 2004). For instance, Thompson and co-authors were able to quantify vibrios at 37 – 8 $\times$ 10^3 cells/mL by quantitative PCR combined with constant denaturant capillary electrophoresis (qPCR-CDCE) of environmental DNA while methods based on detection of rRNA have quantified similar levels of individual cells via fluorescence *in situ* hybridization (FISH) followed by enumeration by microscopy of flow-cytometry (FCM). Eilers et al. (2000) detected 8 $\times$ 10^3–1 $\times$ 10^4 by using FISH and Heidelberg et al. (2002) detected vibrios 5 $\times$ 10^3–1 $\times$ 10^5 cells/mL by FCM. Here, we detected *Vibrio* 16S rRNAs at 3.85 $\times$ 10^3–1.5 $\times$ 10^8 cells/mL from environmental

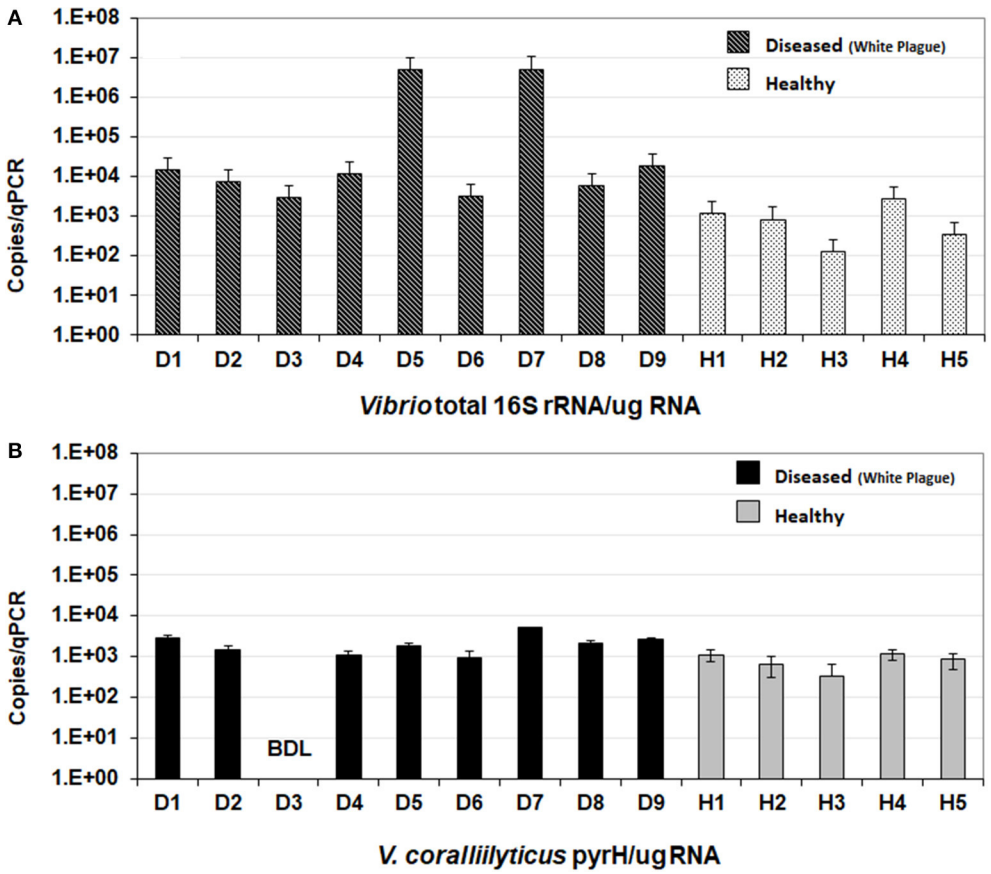


FIGURE 2 | Detection of *Vibrio* species by qPCR approach in coral samples from Abrolhos region. **(A)** Number of total *Vibrio* specie detected by 16S rRNA gene amplification. **(B)** Number of *V. coralliilyticus* detected by pyrH gene amplification. W, White Plague; H, healthy; BDL, below detection limit.

corals, showing a very rapid, specific, and sensitive tool for vibrios detection.

We suggest that both cell number and cell activity are important parameters for the investigation of coral disease. Kim et al. (2003) showed the importance of *pyrH* gene expression in *Vibrio*. Mutants unable to produce UMP kinase directly reducing bacterial growth and decreasing infectivity, revealing that *pyr*H expression has relevance in host colonization.

Luna et al. (2010) detected *V. harveyi* in tropical stony showing tissue necrosis and it was the most represented species recovered from diseased corals. Moreover, the inoculation of *V. harveyi* in healthy colonies of *P. damicornis* induced white plague disease. Surprisingly, we did not detect active *V. harveyi* in these samples. It is possible that *V. harveyi* is associated with the corals at a low level or is not actively expressing RNA due to a dormant like state or is not expressing *pyrH*. Further studies to compare diversity recovered by DNA and RNA-based methods will be necessary to shed further light on this.

V. coralliilyticus has been reported to have specific role in coral disease as a causative agent of WS in *Montipora aequituberculata*, *Pachyseris speciosa,* and *P. damicornis* corals (Sussman et al., 2008; Luna et al., 2010). In the present study *Vibrio corallilyticus* seems also been involved as one of causative agent of WPD in *Mussismilia*, which is reinforced by the statistical significance of Hp2. Active vibrios were more abundant in disease than heathy samples, with the highest vibrio activity detected in the two samples (D5 and D7) collected during summer months when disease outbreaks have been shown to be most prevalent (**Figure 2**; Francini-Filho et al., 2010). Active populations of *V. coralliilyticus* were detected in 13 of 14 coral sampled.

Although, in this study we found that diseased coral had higher vibrio counts than heathy ones and *V. corallilyticus* may have a role in WPD, it's still not clear if a consortium of vibrios instead of a single *Vibrio* species can produce WPD in *M. braziliensis* in the Abrolhos Bank. The idea that a collection or consortium of microbiota play a direct role in the causation of any given disease has recently been discussed in some studies as pathobiome concept (Vayssier-Taussat et al., 2014; Sweet and Bulling, 2017). The pathobiome breaks down the idea of "one pathogen = one disease" and highlights the role of certain members within the microbiome in causing pathogenesis. Moreover, studies of infectious agents have demonstrated that Koch's and Hill's fundamental postulates of "one microbe = one disease" has its limits (Vayssier-Taussat et al., 2014).

We note that *one* of the eight diseased corals tested did not reveal active *V. coralliilyticus* which may indicate that *V. coralliilyticus* is not necessary to establish the disease. However, we cannot rule out the possibility of multiple etiological agents for the same set of symptoms. At least three types of WPD has been described I (Dustan, 1977), II (Richardson et al., 1998a), and III (Richardson et al., 2001), differing in the rate of progression across a coral's surface and affect different species (Richardson et al., 2001; Sutherland et al., 2004). The literature reports three main bacteria specie as the causative pathogen involved in the WPD: *Sphingomonas* (Richardson et al., 1998b), *Aurantimonas coralicida* (Denner et al., 2003) in the Caribbean, and *Thalassomonas loyana* (Thompson et al., 2006) in the Red Sea, while additional agents of the similarly defined WS include several vibrios. Because no pathogen has been unequivocally verified as responsible for WPD, the debate regarding whether a definitive pathogen exists or whether different pathogens or bacterial consortia produce a similar disease phenotype in different coral species still remains (Roder et al., 2014a,b). Given the inherent difficulties of assigning a pathogen to WPD, in the Great Barrier Reef and Indo-Pacific region, WP-like phenotypes have been denominated WS (Willis et al., 2004).

CONCLUSIONS

This study developed a reliable tool for the detection of active vibrios extensively tested with pure cultures and then with environmental samples from Abrolhos as a first attempt to disclose possible associations between vibrios and disease in *Mussismilia*.

It is important to highlight that the robustness of the RT-qPCR assay is supported by accurate quantification of target with the presence of competing non-*V. coralliilyticus* bacterial DNA which had a minimal impact on the target detection. Indeed, because the target organism is embedded within a matrix of other microbial and host cells in the holobiont, it is mandatory to establish accurate quantification in a background of complex targets to attain accurate detection. The real time PCR tool will be valuable for the detection of vibrios in corals as a possible management tool to foresee disease outbreaks.

The current study developed a practical and economic system for the real-time detection and quantification of vibrios from coral tissues. We were able to detect total vibrios using the 16S rRNA gene and *V. coralliilyticus* using the *pyrH* gene. With our protocols, the lowest limit detected for total vibrios and *V. coralliilyticus* from coral tissues of environmental samples was 134 and 658 copies/μg RNA, respectively. This qPCR assay allows the monitoring of samples naturally infected with *V. coralliilyticus* and may be useful for monitoring the health of coral reefs.

AUTHOR CONTRIBUTIONS

LC and JT conceived and designed experiments. LC, KP, and RL carried out experiments and collected data, LC, JT, and AM performed data analysis. AM and GG contributed in field collection. LC wrote the paper. LC, JT, AM, RB, CT, and FT contributed with corrections and discussion. All authors contributed substantially to revisions.

ACKNOWLEDGMENTS

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the international financial support by 1 year (process number 238997/2012-6), which allowed LACT develop the study at MIT in USA. We thank Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for the grants in support the Brazilians research in Rio de Janeiro and São Paulo states.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02272/full#supplementary-material

REFERENCES

Ahmed, R., Rafiquzaman, S. M., Hossain, M. T., Lee, J.-M., and Kong, I.-S. (2015). Species-specific detection of *Vibrio alginolyticus* in shellfish and shrimp by real-time PCR using the groEL gene. *Aquac. Int.* 24, 157–170. doi: 10.1007/s10499-015-9916-5

Alavandi, S. V., Manoranjita, V., Vijayan, K. K., Kalaimani, N., and Santiago, T. C. (2006). Phenotypic and molecular typing of *Vibrio harveyi* isolates and their pathogenicity to tiger shrimp larvae. *Lett. Appl. Microbiol.* 43, 566–570. doi: 10.1111/j.1472-765X.2006.01986.x

Antonius, A. (1981). "The 'band'diseases in coral reefs," in *Proceedings of the Fourth International Coral Reef Symposium,* Vol. 2 (Manila).

Aronson, R. B., and Precht, W. F. (2001). White-band disease and the changing face of Caribbean coral reefs. *Hydrobiologia* 460, 25–38. doi: 10.1023/A:1013103928980

Bidault, A., Richard, G. G., Le Bris, C., and Paillard, C. (2015). Development of a Taqman real-time PCR assay for rapid detection and quantification of *Vibrio tapetis* in extrapallial fluids of clams. *PeerJ.* 3:e1484. doi: 10.7717/peerj.1484

Bourne, D., and Webster, N. S. (2013). "Chapter10: Coral reef bacterial communities," in *Prokaryotes*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin; Heidelberg: Springer Verlag), 163–187.

Bythell, J. C., Pantos, O., and Richardson, L. (2004). "White plague, white band, and other 'white' diseases," in *Coral Health and Disease,* eds E. Rosenberg and Y. Loya (Berlin: Springer-Verlag), 351–365.

Santos Ede, O., Alves, N. Jr., Dias, G. M., Mazotto, A. M., Vermelho, A., Vora, G. J., et al. (2011). Genomic and proteomic analyses of the coral pathogen *Vibrio coralliilyticus* reveal a diverse virulence repertoire. *ISME J.* 5, 1471. doi: 10.1038/ismej.2011.19

De'ath, G., Fabricius, K. E., Sweatman, H., and Puotinen, M. (2012). The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17995–17999. doi: 10.1073/pnas.1208909109

De'ath, G., Lough, J. M., and Fabricius, K. E. (2009). Declining coral calcification on the great barrier reef. *Science* 323, 116–119. doi: 10.1126/science.1165283

Denner, E. B., Smith, G. W., Busse, H. J., Schumann, P., Narzt, T., Polson, S. W., et al. (2003). *Aurantimonas coralicida* gen. nov., sp. nov., the causative agent

of white plague type II on Caribbean scleractinian corals. *Int. J. Syst. Evol. Microbiol.* 53(Pt 4), 1115–1122. doi: 10.1099/ijs.0.02359-0

Dustan, P. (1977). Vitality of reef coral populations off Key Largo, Florida: recruitment and mortality. *Environ. Geol.* 2, 51–58. doi: 10.1007/BF02430665

Eilers, H., Pernthaler, J., Glockner, F. O., and Amann, R. (2000). Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl. Environ. Microbiol.* 66, 3044–3051. doi: 10.1128/AEM.66.7.3044-3051.2000

Fernando, S. C., Wang, J., Sparling, K., Garcia, G. D., Francini-Filho, R. B., de Moura, R. L., et al. (2015). Microbiota of the major South Atlantic reef building coral Mussismilia. *Microb. Ecol.* 69, 267–280. doi: 10.1007/s00248-014-0474-6

Francini-Filho, R., Reis, R., Meirelles, P., Moura, R., Thompson, F., Kikuchi, R., et al. (2010). Seasonal prevalence of white plague like disease on the endemic Brazilian reef coral *Mussismilia braziliensis*/Prevalencia estacional de la enfermedad de la plaga blanca en el coral endémico de Brasil *Mussismilia braziliensis*. *Lat. Am. J. Aquat. Res.* 38:292. doi: 10.3856/vol38-issue2-fulltext-16

Freitas, M. O., Abilhoa, V., and Costa e Silva, G. H. (2011). Feeding ecology of *Lutjanus analis* (Teleostei: *Lutjanidae*) from Abrolhos bank, Eastern Brazil. *Neotrop. Ichthyol.* 9, 411–418. doi: 10.1590/S1679-62252011005000022

Garcia, G. D., Gregoracci, G. B., Santos Ede, O., Meirelles, P. M., Silva, G. G., Edwards, R., et al. (2013). Metagenomic analysis of healthy and white plague-affected *Mussismilia braziliensis* corals. *Microb. Ecol.* 65, 1076–1086. doi: 10.1007/s00248-012-0161-4

Garrido-Maestu, A., Lozano-León, A., Rodríguez-Souto, R. R., Vieites-Maneiro, R., Chapela, M.-J., and Cabado, A. G. (2016). Presence of pathogenic *Vibrio* species in fresh mussels harvested in the southern Rias of Galicia (NW Spain). *Food Control* 59, 759–765. doi: 10.1016/j.foodcont.2015.06.054

Gauger, E., Smolowitz, R., Uhlinger, K., Casey, J., and Gómez-Chiarri, M. (2006). *Vibrio harveyi* and other bacterial pathogens in cultured summer flounder, *Paralichthys dentatus*. *Aquaculture* 260, 10–20. doi: 10.1016/j.aquaculture.2006.06.012

Goarant, C., and Merien, F. (2006). Quantification of *Vibrio penaeicida*, the etiological agent of Syndrome 93 in New Caledonian shrimp, by real-time PCR using SYBR Green I chemistry. *J. Microbiol. Methods* 67, 27–35. doi: 10.1016/j.mimet.2006.02.013

Guerrero-Ferreira, R. C., and Nishiguchi, M. K. (2010). Differential gene expression in bacterial symbionts from loliginid squids demonstrates variation between mutualistic and environmental niches. *Environ. Microbiol. Rep.* 2, 514–523. doi: 10.1111/j.1758-2229.2009.00077.x

Harris, D. C. (1995). *Quantitative Chemical Analysis, 4th Edn*. New York, NY: W. H. Freeman and Company.

Heidelberg, J. F., Heidelberg, K. B., and Colwell, R. R. (2002). Seasonality of Chesapeake Bay bacterioplankton species. *Appl. Environ. Microbiol.* 68, 5488–5497. doi: 10.1128/AEM.68.11.5488-5497.2002

Hernandez-Agreda, A., Leggat, W., Bongaerts, P., and Ainsworth, T. D. (2016).The microbial signature provides insight into the mechanistic basis of coral success across reef habitats. *MBio* 7, e00560-16. doi: 10.1128/mBio.00560-16

Kim, Y. R., Lee, S. E., Kim, C. M., Kim, S. Y., Shin, E. K., Shin, D. H., et al. (2003). Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect. Immun.* 71, 5461–5471. doi: 10.1128/IAI.71.10.5461-5471.2003

Leão, Z. M. A. N., and Kikuchi, R. K. P. (2001). "The Abrolhos reefs of Brazil," in *Ecological Studies. Coastal Marine Ecosystems of Latin America,* Vol. 144, eds U. Seeliger and B. Kjerfve (Berlin: Springer), 83–96.

Lee, S. E., Kim, S. Y., Kim, C. M., Kim, M. K., Kim, Y. R., Jeong, K., et al. (2007). The pyrH gene of *Vibrio vulnificus* is an essential *in vivo* survival factor. *Infect. Immun.* 75, 2795–2801. doi: 10.1128/IAI.01499-06

Lee, Z. M., Bussema, C. III, and Schmidt, T. M. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res.* 37, D489–D493. doi: 10.1093/nar/gkn689

Luna, G. M., Bongiorni, L., Gili, C., Biavasco, F., and Danovaro, R. (2010). Vibrio harveyi as a causative agent of the White Syndrome in tropical stony corals. *Environ. Microbiol. Rep.* 2, 120–127. doi: 10.1111/j.1758-2229.2009.00114.x

McCleary, S., and Henshilwood, K. (2015). Novel quantitative TaqMan(R) MGB real-time PCR for sensitive detection of *Vibrio aestuarianus* in *Crassostrea gigas*. *Dis. Aquat. Org.* 114, 239–248. doi: 10.3354/dao02869

McDole, T., Nulton, J., Barott, K. L., Felts, B., Hand, C., Hatay, M., et al. (2012). Assessing coral reefs on a Pacific-wide scale using the microbialization score. *PLoS ONE* 7:e43233. doi: 10.1371/journal.pone.0043233

Nicolas, J. L., Basuyaux, O., Mazurie, J., and Thebault, A. (2002). Vibrio carchariae, a pathogen of the abalone *Haliotis tuberculata*. *Dis. Aquat. Org.* 50, 35–43. doi: 10.3354/dao050035

Nshimyimana, J. P., Ekklesia, E., Shanahan, P., Chua, L. H., and Thompson, J. R. (2014). Distribution and abundance of human-specific Bacteroides and relation to traditional indicators in an urban tropical catchment. *J. Appl. Microbiol.* 116, 1369–1383. doi: 10.1111/jam.12455

Osinga, R., Schutter, M., Griffioen, B., Wijffels, R. H., Verreth, J. A., Shafir, S., et al. (2011). The biology and economics of coral growth. *Mar. Biotechnol.* 13, 658–671. doi: 10.1007/s10126-011-9382-7

Pollock, F. J., Morris, P. J., Willis, B. L., and Bourne, D. G. (2010). Detection and quantification of the coral pathogen *Vibrio coralliilyticus* by real-time PCR with TaqMan fluorescent probes. *Appl. Environ. Microbiol.* 76, 5282–5286. doi: 10.1128/AEM.00330-10

Pollock, F. J., Morris, P. J., Willis, B. L., and Bourne, D. G. (2011). The urgent need for robust coral disease diagnostics. *PLoS Pathog.* 7:e1002183. doi: 10.1371/journal.ppat.1002183

Richardson, L. L., Goldberg, W. M., Carlton, R. G., and Halas, J. C. (1998a). Coral disease outbreak in the Florida Keys: plague type II. *Rev. Biol. Trop.* 46(Suppl. 5), 187–198.

Richardson, L. L., Goldberg, W. M., Kuta, K. G., Aronson, R. B., Smith, G. W., Ritchie, K. B., et al. (1998b). Florida's mystery coral-killer identified. *Nature* 392, 557. doi: 10.1038/33302

Richardson, L. L., Smith, G. W., Ritchie, K. B., and Carlton, R. G. (2001). "Integrating microbiological, microsensor, molecular, and physiologic techniques in the study of coral disease pathogenesis," in *The Ecology and Etiology of Newly Emerging Marine Diseases,* ed J. W. Porter (Dordrecht: Springer), 71–89.

Roder, C., Arif, C., Bayer, T., Aranda, M., Daniels, C., Shibl, A., et al. (2014a). Bacterial profiling of White Plague Disease in a comparative coral species framework. *ISME J.* 8, 31–39. doi: 10.1038/ismej.2013.127

Roder, C., Arif, C., Daniels, C., Weil, E., and Voolstra, C. R. (2014b). Bacterial profiling of White Plague Disease across corals and oceans indicates a conserved and distinct disease microbiome. *Mol. Ecol.* 23, 965–974. doi: 10.1111/mec.12638

Saulnier, D., De Decker, S., and Haffner, P. (2009). Real-time PCR assay for rapid detection and quantification of *Vibrio aestuarianus* in oyster and seawater: a useful tool for epidemiologic studies. *J. Microbiol. Methods* 77, 191–197. doi: 10.1016/j.mimet.2009.01.021

Sawabe, T., Kita-Tsukamoto, K., and Thompson, F. L. (2007). Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. *J. Bacteriol.* 189, 7932–7936. doi: 10.1128/JB.00693-07

Schikorski, D., Renault, T., Paillard, C., Bidault-Toffin, A., Tourbiez, D., and Saulnier, D. (2013). Development of TaqMan real-time PCR assays for monitoring *Vibrio harveyi* infection and a plasmid harbored by virulent strains in European abalone *Haliotis tuberculata* aquaculture. *Aquaculture* 392, 106–112. doi: 10.1016/j.aquaculture.2013.02.005

Stangegaard, M., Dufva, I. H., and Dufva, M. (2006). Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA. *BioTechniques* 40, 649–660. doi: 10.2144/000112153

Sussman, M., Mieog, J. C., Doyle, J., Victor, S., Willis, B. L., and Bourne, D. G. (2009). *Vibrio* zinc-metalloprotease causes photoinactivation of coral endosymbionts and coral tissue lesions. *PLoS ONE* 4:e4511. doi: 10.1371/journal.pone.0004511

Sussman, M., Willis, B. L., Victor, S., and Bourne, D. G. (2008). Coral pathogens identified for White Syndrome (WS) epizootics in the Indo-Pacific. *PLoS ONE* 3:e2393. doi: 10.1371/journal.pone.0002393

Sutherland, K. P., Porter, J. W., and Torres, C. (2004). Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. *Mar. Ecol. Prog. Ser.* 266, 273–302. doi: 10.3354/meps266273

Sweet, M. J., and Bulling, M. T. (2017). On the Importance of the Microbiome and pathobiome in coral health and disease. *Front. Mar. Sci.* 4:9. doi: 10.3389/fmars.2017.00009

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Thompson, F. L., Barash, Y., Sawabe, T., Sharon, G., Swings, J., and Rosenberg, E. (2006). *Thalassomonas loyana* sp. nov., a causative agent of the white plague-like disease of corals on the Eilat coral reef. *Int. J. Syst. Evol. Microbiol.* 56 (Pt 2), 365–368. doi: 10.1099/ijs.0.63800-0

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., et al. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl. Environ. Microbiol.* 71, 5107–5115. doi: 10.1128/AEM.71.9.5107-5115.2005

Thompson, J. R., Randa, M. A., Marcelino, L. A., Tomita-Mitchell, A., Lim, E., and Polz, M. F. (2004). Diversity and dynamics of a north atlantic coastal *Vibrio* community. *Appl. Environ. Microbiol.* 70, 4103–4110. doi: 10.1128/AEM.70.7.4103-4110.2004

Thompson, J. R., Rivera, H. E., Closek, C. J., and Medina, M. (2014). Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Front. Cell. Infect. Microbiol.* 4:176. doi: 10.3389/fcimb.2014.00176

Vayssier-Taussat, M., Albina, E., Citti, C., Cosson, J. F., Jacques, M. A., Lebrun, M. H., et al. (2014). Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. *Front. Cell. Infect. Microbiol.* 4:29. doi: 10.3389/fcimb.2014.00029

Voet, D., and Voet, J. G. (2004). *Biochemistry, 3rd Edn*. Hoboken, NJ: Wiley.

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., and Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environ. Microbiol.* 9, 2707–2719. doi: 10.1111/j.1462-2920.2007.01383.x

Willis, B. L., Page, C. A., and Dinsdale, E. A. (2004). "Coral disease on the great barrier reef," in *Coral Health and Disease,* eds Y. Loya and E. Rosenberg (Berlin; Heidelberg: Springer), 69–104.

Wilson, B., Muirhead, A., Bazanella, M., Huete-Stauffer, C., Vezzulli, L., and Bourne, D. G. (2013). An improved detection and quantification method for the coral pathogen *Vibrio coralliilyticus*. *PLoS ONE* 8:e81800. doi: 10.1371/journal.pone.0081800

Woodley, C. M., Downs, C. A., Bruckner, A. W., Porter, J. W., and Galloway, S. B. (2015). *Diseases of Coral.* Hoboken, NJ: John Wiley & Sons.

Work, T. M., and Aeby, G. S. (2006). Systematically describing gross lesions in corals. *Dis. Aquat. Org.* 70, 155–160. doi: 10.3354/dao070155

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chimetto Tonon, Thompson, Moreira, Garcia, Penn, Lim, Berlinck, Thompson and Thompson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

HYPOTHESIS AND THEORY
published: 30 June 2017
doi: 10.3389/fmicb.2017.01220

Engineering Strategies to Decode and Enhance the Genomes of Coral Symbionts

Rachel A. Levin[1,2,3], Christian R. Voolstra[4], Shobhit Agrawal[4], Peter D. Steinberg[1,2,5], David J. Suggett[3] and Madeleine J. H. van Oppen[6,7]*

[1] Centre for Marine Bio-Innovation, The University of New South Wales, Sydney, NSW, Australia, [2] School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW, Australia, [3] Climate Change Cluster, University of Technology Sydney, Ultimo, NSW, Australia, [4] Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [5] Sydney Institute of Marine Science, Mosman, NSW, Australia, [6] Australian Institute of Marine Science, Townsville, QLD, Australia, [7] School of BioSciences, The University of Melbourne, Parkville, VIC, Australia

Elevated sea surface temperatures from a severe and prolonged El Niño event (2014–2016) fueled by climate change have resulted in mass coral bleaching (loss of dinoflagellate photosymbionts, *Symbiodinium* spp., from coral tissues) and subsequent coral mortality, devastating reefs worldwide. Genetic variation within and between *Symbiodinium* species strongly influences the bleaching tolerance of corals, thus recent papers have called for genetic engineering of *Symbiodinium* to elucidate the genetic basis of bleaching-relevant *Symbiodinium* traits. However, while *Symbiodinium* has been intensively studied for over 50 years, genetic transformation of *Symbiodinium* has seen little success likely due to the large evolutionary divergence between *Symbiodinium* and other model eukaryotes rendering standard transformation systems incompatible. Here, we integrate the growing wealth of *Symbiodinium* next-generation sequencing data to design tailored genetic engineering strategies. Specifically, we develop a testable expression construct model that incorporates endogenous *Symbiodinium* promoters, terminators, and genes of interest, as well as an internal ribosomal entry site from a *Symbiodinium* virus. Furthermore, we assess the potential for CRISPR/Cas9 genome editing through new analyses of the three currently available *Symbiodinium* genomes. Finally, we discuss how genetic engineering could be applied to enhance the stress tolerance of *Symbiodinium*, and in turn, coral reefs.

Keywords: synthetic biology, genetic engineering, dinoflagellate, *Symbiodinium*, zooxanthellae, coral bleaching

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Michael Sweet, University of Derby, United Kingdom
Anthony William Larkum, University of Technology Sydney, Australia

****Correspondence:***
Rachel A. Levin
rachylevin@gmail.com

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *16 March 2017*
Accepted: *16 June 2017*
Published: *30 June 2017*

Citation:
Levin RA, Voolstra CR, Agrawal S, Steinberg PD, Suggett DJ and van Oppen MJH (2017) Engineering Strategies to Decode and Enhance the Genomes of Coral Symbionts. Front. Microbiol. 8:1220. doi: 10.3389/fmicb.2017.01220

INTRODUCTION

Photosynthetic dinoflagellates are critical primary producers in the aquatic environment, yet, their functional genomics are largely unexplored (Leggat et al., 2011; Murray et al., 2016). *Symbiodinium* is considered one of the most important dinoflagellate genera given its role as the essential photosymbiont of many tropical reef invertebrates, notably reef-building corals (Trench and Blank, 1987). Provision of photosynthetically derived metabolites from *Symbiodinium* to the coral host drives coral calcification and growth that forms the foundation of coral reef ecosystems (Muscatine and Porter, 1977; Muscatine, 1990; Kirk and Weis, 2016). Thermal and light stress

cause photosynthetic dysfunction of *Symbiodinium* and increased leakage of harmful reactive oxygen species from their cells, a process considered largely responsible for the dissociation of *Symbiodinium* from corals characterized as "coral bleaching" (Warner et al., 1999; Suggett et al., 2008; Weis, 2008; Levin et al., 2016). *Symbiodinium* has therefore become established as a major focus for research globally, and in effect, a model genus for dinoflagellates.

Dinoflagellates evolved an estimated 520 million years ago (Moldowan and Talyzina, 1998) and exhibit substantial evolutionary divergence from model eukaryotic organisms including other microalgae such as *Chlamydomonas* and diatoms (Shoguchi et al., 2013). Consequently, dinoflagellates possess unusual biological features that have hindered research progress, such as some of the largest known nuclear genomes (1.5–112 Gbp, typically exceeding the size of the human haploid genome), permanently condensed liquid-crystalline chromosomes, trans-splicing of polycistronic mRNAs, and plastid genomes that are divided up into minicircles (Shoguchi et al., 2013; Zhang et al., 2013; Lin et al., 2015; Murray et al., 2016). The *Symbiodinium* genus evolved an estimated 50 million years ago and is highly diverse, containing nine major evolutionary lineages or "clades" (A–I; Coffroth and Santos, 2005; Pochon et al., 2006; Pochon and Gates, 2010) with hundreds of genetically distinct "types/sub-clades" considered to be different species[1] (Tonk et al., 2013). Genetic factors that promote differences in stress tolerance between *Symbiodinium* variants (both inter- and intra-specific) strongly influence coral gene expression and bleaching susceptibility (Berkelmans and van Oppen, 2006; DeSalvo et al., 2010; Yuyama et al., 2012; Levin et al., 2016). However, the capacity to fully explore *Symbiodinium* genetics is currently restricted by a lack of genetic engineering capability. Genetic engineering has been central to the study of gene function and phenotypic enhancement in organisms ranging from microbes to mammals and a key platform for socioeconomic industries and biotechnologies; yet only two cases of transgene expression in *Symbiodinium* have ever been validated (ten Lohuis and Miller, 1998; Ortiz-Matamoros et al., 2015a).

In 1998, a type A1 strain was transformed at very low efficiencies using silicon carbide whiskers with plasmids encoding expression constructs with plant, plant-viral, and agrobacterial promoters (nos, CaMV 35S, and p1′2′) to drive transcription of antibiotic resistance genes (*nptII* and *hptII*) and a reporter gene (*GUS*) (ten Lohuis and Miller, 1998); however, these results have yet to be reproduced. It was not until 2015 that another case of transgene expression in *Symbiodinium* was reported (Ortiz-Matamoros et al., 2015b). Plasmids encoding expression constructs with plant and plant-viral promoters (nos and double CaMV 35S) to drive transcription of a herbicide resistance gene (*bar*) and a reporter gene (*GFP*) were introduced to type A1, B1, and F1 strains using glass beads. Whilst cells transiently exhibited improved herbicide resistance and suggestive *GFP* signal, transformations were not validated through DNA, RNA, or protein analysis (Ortiz-Matamoros et al., 2015b). Further transformation of these strains was attempted using *Agrobacterium* carrying plasmids with the same expression constructs, but the transformants were transient and unable to divide (Ortiz-Matamoros et al., 2015a). Of these studies, none attempted manipulation of ecologically relevant genes thereby limiting new insight gained into *Symbiodinium* biology.

Therefore, in an attempt to overcome the bottleneck that has become established in transforming *Symbiodinium* (and other dinoflagellates), we recommend a new approach that capitalizes on the recent surge in "omics" breakthroughs (**Figure 1**). By evaluating the rapidly increasing supply of next-generation sequencing (NGS) data, we propose a genetic engineering framework for *Symbiodinium* that may markedly advance our understanding of these important dinoflagellates. Furthermore, genetic manipulation of *Symbiodinium* in order to reduce coral bleaching has been hypothesized as a strategy to facilitate coral management as reefs continue to rapidly deteriorate under climate change (van Oppen et al., 2017). Combatting the impacts of climate change and conserving marine organisms are both key goals for sustainable development set forth by the United Nations[2]. Thus, we believe genetic engineering of *Symbiodinium* may open a novel avenue to achieve these goals by protecting corals from climate change.

TAILORING A GENETIC ENGINEERING FRAMEWORK FOR *Symbiodinium*

Fundamental components of *Symbiodinium* biology have recently been uncovered through a boom in NGS (**Figure 1**), particularly the assembly of the first *Symbiodinium* genomes and transcriptomes, direct correlation between *Symbiodinium* transcriptional and physiological states, and discovery of genes from viruses actively infecting *Symbiodinium* cells. Furthermore, NGS of *Symbiodinium* has revealed genetic elements that may allow for transformation of *Symbiodinium*. In the following sections, we detail how unique *Symbiodinium* promoters, specific *Symbiodinium* genes underpinning important phenotypes, and a viral internal ribosomal entry site recognized by *Symbiodinium* ribosomes could be integrated to build expression constructs for *Symbiodinium*.

TRANSCRIPTIONAL PROMOTERS AND TERMINATORS

Currently, dinoflagellate nuclear genome assemblies are all from the genus *Symbiodinium* (types A1, B1, and F1; Shoguchi et al., 2013; Lin et al., 2015; Aranda et al., 2016), emphasizing the importance of *Symbiodinium* to dinoflagellate research. The assemblies have revealed the immense size of *Symbiodinium* genomes with 36,850–49,109 genes, unidirectional gene orientation, prevalent gene tandem arrays, microRNAs along with putative gene targets, and unique promoter architecture (Shoguchi et al., 2013; Lin et al., 2015; Aranda et al., 2016).

[1] http://www.symbiogbr.org/

[2] https://sustainabledevelopment.un.org/?menu=1300, last accessed March 2017.

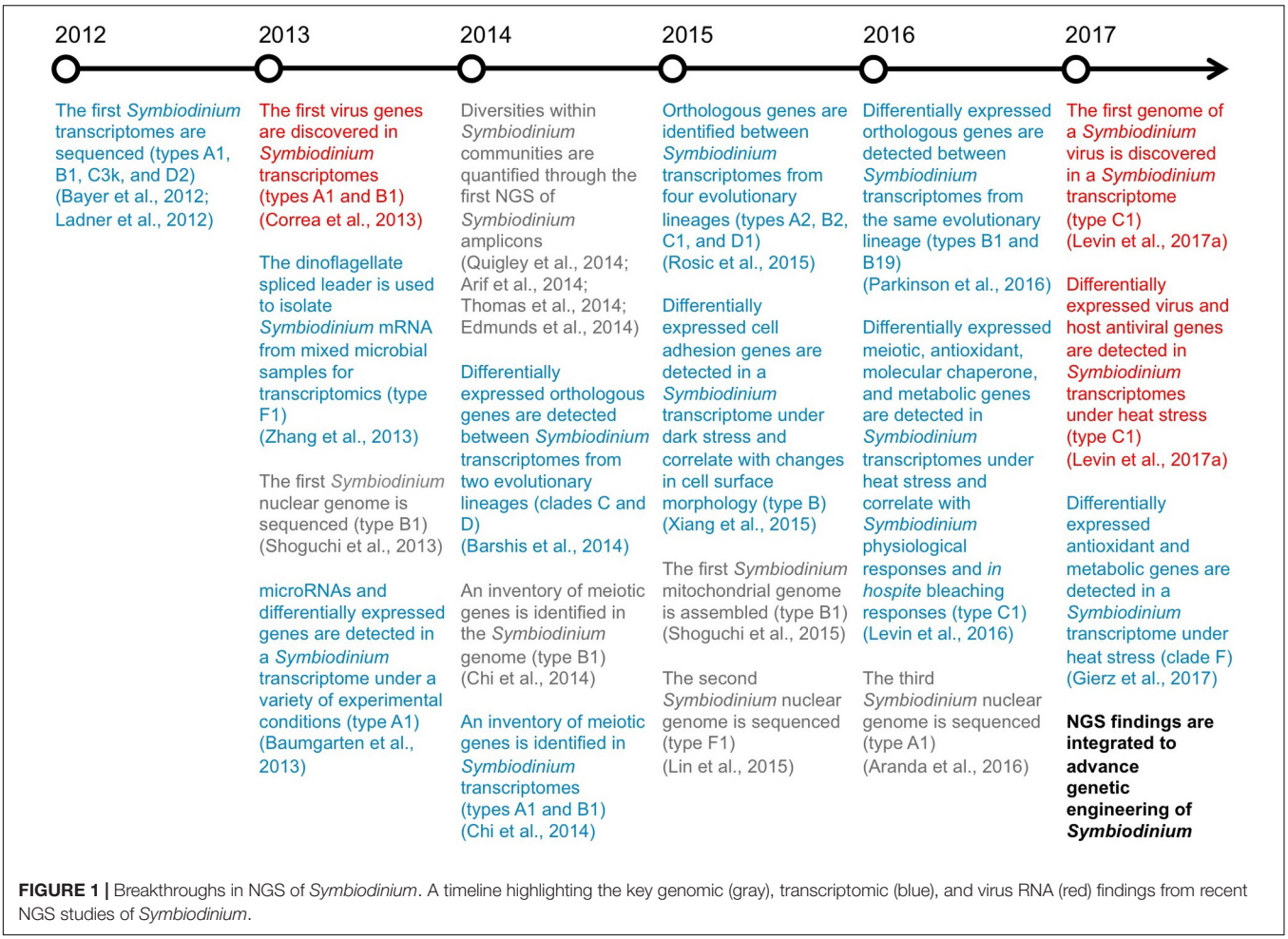


FIGURE 1 | Breakthroughs in NGS of *Symbiodinium*. A timeline highlighting the key genomic (gray), transcriptomic (blue), and virus RNA (red) findings from recent NGS studies of *Symbiodinium*.

Rather than the traditional TATA-box of eukaryotic promoters, *Symbiodinium* promoters appear to have a TTTT-box that is followed by a unique transcription start site (YYANWYY), branch point (YTNAY), and acceptor for the dinoflagellate spliced leader (AG) (Lin et al., 2015). Additionally, instead of the typical eukaryotic polyadenylation signal AAUAAA, dinoflagellate terminators use AAAAG/C (Bachvaroff and Place, 2008). Hence, utilization of endogenous *Symbiodinium* promoters and terminators (as opposed to promoters and terminators from other organisms) would likely improve expression and stability of transgenes introduced into *Symbiodinium*. By chance, the CAMV 35S (plant-viral) promoter happens to contain all of the described *Symbiodinium* promoter elements, and the CAMV 35S (plant-viral) and nos (plant) terminators both contain the dinoflagellate polyadenylation signal; this may have contributed to their ability to drive transgene expression in *Symbiodinium* previously (ten Lohuis and Miller, 1998; Ortiz-Matamoros et al., 2015a).

Recent transcriptomic studies have identified highly expressed *Symbiodinium* nuclear genes that can be genome-mapped to uncover strong, endogenous promoters and their corresponding terminators. These promoters and terminators can be isolated from purified genomic DNA (gDNA) through PCR and incorporated into custom DNA expression constructs for *Symbiodinium* (**Figure 2**). Among the most highly expressed transcripts in *Symbiodinium* transcriptomes are genes for peridinin-chlorophyll *a*-binding protein, carotenoo-chlorophyll *a*-*c*-binding protein, major basic nuclear protein 2, dinoflagellate viral nucleoprotein, and glyceraldehyde-3-phosphate dehydrogenase (Baumgarten et al., 2013; Levin et al., 2016; Parkinson et al., 2016); though all are multi-copy genes (Shoguchi et al., 2013; Lin et al., 2015; Aranda et al., 2016). Ideally, highly expressed nuclear genes chosen for promoter selection should not have high copy numbers, as their expression levels may largely be due to prevalence in the genome rather than strong promoters. Constitutively expressed nuclear genes are also desirable for selection of promoters that drive consistent transcription regardless of experimental conditions, and thus, drive reliable transgene expression.

To illustrate this approach of *Symbiodinium* promoter selection, we examined NGS data from a type A1 *Symbiodinium* strain for which the nuclear genome has been recently sequenced (Aranda et al., 2016) and the transcriptional responses to various conditions (temperatures, ionic stress, dark stress, and contrasting circadian rhythm time points) have been determined (Baumgarten et al., 2013). Locus 144 and Locus 1768 in the

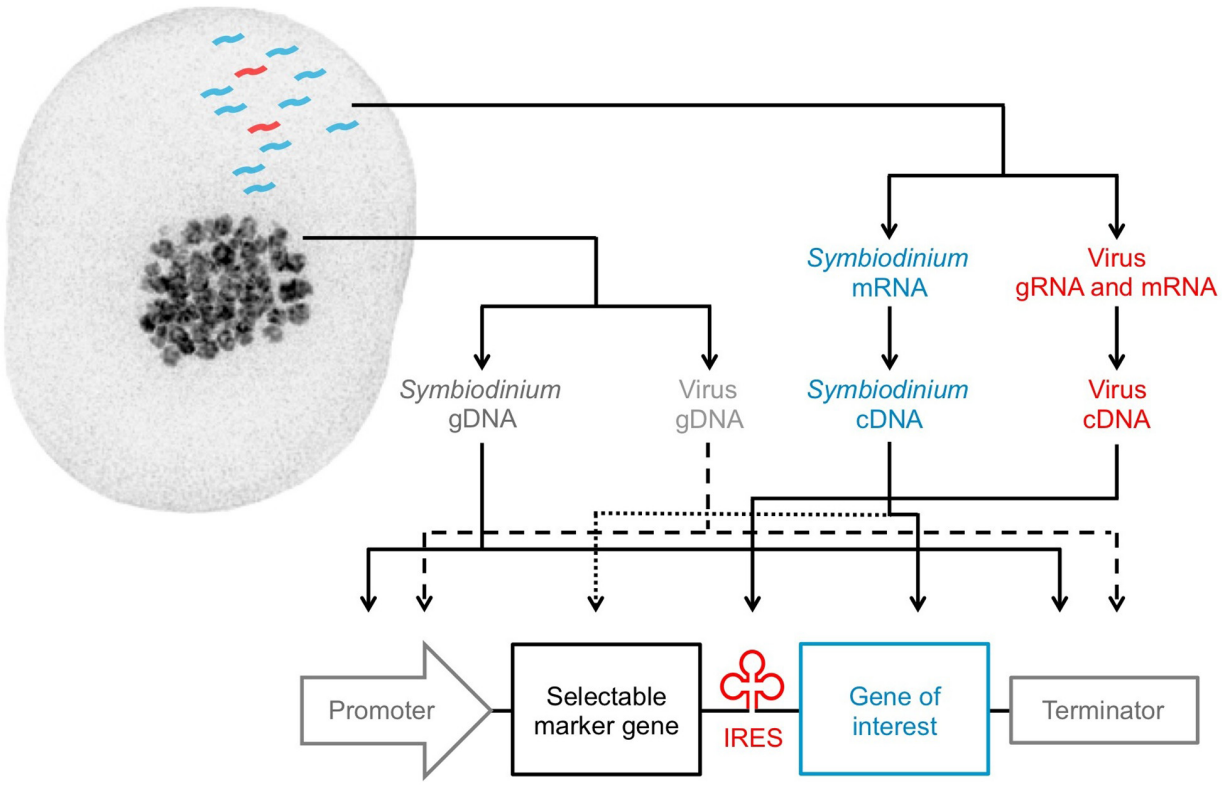


FIGURE 2 | Design of a tailored expression construct for *Symbiodinium*. Genetic elements that can be isolated from *Symbiodinium* cells: *Symbiodinium* genomic DNA (dark gray), *Symbiodinium* messenger RNA (blue), resident virus genomic DNA (light gray), resident virus genomic or messenger RNA (red). Solid lines (identified elements) and dashed or dotted lines (unidentified elements) are used to arrange the elements into a *Symbiodinium* expression construct. The pictured *Symbiodinium* cell (type C1) was stained with DAPI for imaging on a DeltaVision OMX Blaze microscope (excitation/emission: 405 nm/419–465 nm).

type A1 transcriptome, a subunit of a large neutral amino acids transporter and a putative ATP-binding cassette transporter gene, both show high expression across all conditions (average expression in the top 2% of all genes; Baumgarten et al., 2013) and map tightly to the type A1 genome scaffolds 710 and 484, respectively. No significant open reading frames are found >5 kb up- or down-stream of either gene, confirming that they are not part of tandem arrays. For each gene, all *Symbiodinium* promoter elements are within 1 kb of the start codon, and the dinoflagellate polyadenylation signal is found ∼300 bp after the stop codon. These promoter and terminator regions could therefore be isolated and utilized to drive high and consistent expression of transgenes in a *Symbiodinium* expression construct.

GENES OF INTEREST

Recent transcriptomic studies have been fundamental in the discovery of *Symbiodinium* nuclear genes that underpin phenotypic traits, such as those related to cell adhesion (e.g., *GspB*, *Svep1*, *Slap1*; Xiang et al., 2015), sexual reproduction (e.g., *Msh4*, *Msh5*, *Spo11-2*; Chi et al., 2014; Levin et al., 2016; Gierz et al., 2017), antiviral response (e.g., *Birc3*, *Ns1bp*, *Ifih1*; Levin et al., 2017a), and antioxidant activity/thermal tolerance (e.g., *Fe-sod*, *Mn-sod*, *Pxrd*, *Hsp70*; Levin et al., 2016; Gierz et al., 2017). *Symbiodinium* antioxidant genes are of particular interests because of their potential role in defining bleaching susceptibility of the coral host (Krueger et al., 2015; Levin et al., 2016). For instance, iron-type superoxide dismutase (*Fe-sod*) genes are believed to minimize thermally induced oxidative damage to photosynthetic apparatuses and leakage of harmful reactive oxygen species from type C1 *Symbiodinium* cells—determinants of coral bleaching (Weis, 2008); however, these genes are not expressed at detectable levels in all *Symbiodinium* variants (Krueger et al., 2015; Levin et al., 2016). A *Fe-sod* gene could therefore be inserted after a strong *Symbiodinium* promoter in an expression construct to drive its over-expression for evaluation of its phenotypic influence on *Symbiodinium*. Endogenous genes of interest should be isolated through PCR of complementary DNA (cDNA) reverse transcribed from purified mRNA, since gDNA introns may prevent proper expression in constructs (**Figure 2**).

Expression of exogenous genes of interest in *Symbiodinium* could also greatly advance investigations of ecological processes central to coral reef health. For instance, documenting competition between *Symbiodinium* types, transmission and acquisition of *Symbiodinium* types by the coral host, and shuffling of *Symbiodinium* types within host tissues

(Toller et al., 2001; van Oppen et al., 2001; Little et al., 2004; Berkelmans and van Oppen, 2006; Byler et al., 2013; Boulotte et al., 2016) is currently reliant upon sequencing since it is not possible to visually differentiate many types. As a result, studies have been restricted to low temporal and spatial resolution relative to real-time imaging. Instead, the ability to color-code *Symbiodinium* types through genetic transformation with various fluorescent proteins could illuminate these phenomena by enabling real-time imaging for visually differentiating types. Additionally, tagging endogenous genes of interest through fluorescent protein fusions would permit imaging of protein localization within *Symbiodinium* cells and potential protein secretion out of *Symbiodinium* cells (Xiang et al., 2015). When selecting appropriate fluorescent proteins, it will be imperative to consider the extreme autofluorescence of *Symbiodinium* (Shaner et al., 2005); for example, venus (excitation/emission: 515/528 nm), tdTomato (excitation/emission: 554/581 nm), and mCherry (excitation/emission: 587/610 nm) are promising candidates as their fluorescence properties are off-peak of the *Symbiodinium* excitation and emission spectra (Hennige et al., 2009; Jiang et al., 2012). Finally, codon optimization may be necessary for optimal exogenous gene expression in *Symbiodinium* since codon usage of *Symbiodinium* genes can be divergent from foreign genes (Levin et al., 2017a) and even between *Symbiodinium* nuclear and minicircle genes (Bayer et al., 2012).

SELECTABLE MARKER GENES

Although antibiotics have previously been used to select transformed *Symbiodinium* (ten Lohuis and Miller, 1998; Ortiz-Matamoros et al., 2015a), their use is problematic for two main reasons. Firstly, eliminating wild-type *Symbiodinium* in culture requires high concentrations of antibiotics (e.g., 3 mg/ml of G418 or hygromycin; ten Lohuis and Miller, 1998), making experimentation and long-term maintenance of transformed cell lines extremely costly. It is also important to note that natural antibiotic resistances are not uniform across all strains (Supplementary Table 1), so dosage curves are necessary before conducting transformation trials. Secondly, dinoflagellates including *Symbiodinium* require symbiotic bacteria to grow optimally (Alavi et al., 2001; Croft et al., 2005; Miller and Belas, 2006; Ritchie, 2012). Since eukaryotic antibiotics can also be toxic to prokaryotes (Gonzalez et al., 1978; Colanduoni and Villafranca, 1986; Pline et al., 2001; Vicens and Westhof, 2003), bacterial communities in *Symbiodinium* cultures are removed during antibiotic selection.

To preserve symbiotic bacteria, alternatives to antibiotic selection markers should be considered, such as genes that provide growth advantages under specific conditions by increasing pathogen resistance, increasing thermal tolerance, or allowing for utilization of non-metabolized carbohydrates (Breyer et al., 2014). The precise functions of these alternative marker genes (e.g., phosphomannose isomerase) are well defined and shown to be applicable to many photosynthetic species (Stoykova and Stoeva-Popova, 2011), though their compatibilities with dinoflagellates are unknown. Discovery of endogenous selectable markers should therefore also be pursued. Recent *Symbiodinium* transcriptomic studies have uncovered genes involved in selection-relevant phenotypes like photosynthetic ability at unique light regimes (Parkinson et al., 2016) or tolerance to increased temperature regimes (Levin et al., 2016). These *Symbiodinium* genes could first be expressed in more easily transformed microalgae like *Chlamydomonas* and diatoms to gauge the potential for their up-regulation to grant a significant selectable advantage under specific conditions.

VIRAL ELEMENTS

Viral promoters and terminators, internal ribosome entry sites (IRES), and 2A peptides are staple regulatory elements incorporated in expression constructs since they have evolved to be recognized by eukaryotic machinery for efficient and stable foreign gene expression (Benfey and Chua, 1990; Martínez-Salas, 1999; Levin et al., 2014). *Symbiodinium* transcriptomics have led to the discovery of genes, as well as an entire RNA genome, from novel eukaryotic viruses that infect *Symbiodinium* (Correa et al., 2013; Levin et al., 2017a). A putative viral IRES, which allows cap-independent translation to produce separate proteins from one mRNA transcript, was found between the two open reading frames in the RNA genome of the +ssRNA virus infecting type C1 *Symbiodinium* (GenBank accession: KX538960 and KX787934; Levin et al., 2017a). The +ssRNA virus transcripts were extremely abundant in a type C1 *Symbiodinium* transcriptome (Levin et al., 2017a), and such rampant +ssRNA virus replication indicates that *Symbiodinium* ribosomes have high affinity to this IRES.

IRES sequences enable the creation of polycistronic constructs transcriptionally controlled by a single promoter (Martínez-Salas, 1999). By permitting simultaneous expression of two independent proteins from one mRNA, a bicistronic construct can achieve long-term expression of a gene of interest because the gene of interest is transcriptionally fused to the selectable marker gene (Gurtu et al., 1996; **Figure 2**). Conversely, in monocistronic constructs, the selectable marker gene often maintains expression, while the gene of interest becomes transcriptionally repressed over time if it does not increase fitness of the cell (Allera-Moreau et al., 2007). Therefore, the IRES from the *Symbiodinium* +ssRNA virus is a valuable viral element that is recognized by *Symbiodinium* ribosomes and may improve the stability of transgene expression in *Symbiodinium.* Moving forward, NGS data of viruses in *Symbiodinium* cultures (Weynberg et al., 2017) and the coral holobiont (Weynberg et al., 2014; Correa et al., 2016) should be mined for promoter, terminator, and other regulatory elements from *Symbiodinium* viruses, given the proven benefits of viral elements to genetic engineering. Once assembled, the *Symbiodinium* expression construct (**Figure 2**) can be combined with the backbone of a standard cloning plasmid; added into an artificial, replicating minicircle (Nehlsen et al., 2006; Karas et al., 2015); or serve as a repair template for CRISPR/Cas9 genome editing (Cong et al., 2013).

CRISPR/Cas9 GENOME EDITING AND *Symbiodinium*

Within the past 5 years, CRISPR/Cas9 has revolutionized genome editing by allowing precise changes to be made to target sites in the genome (Cong et al., 2013; Baek et al., 2016; Nymark et al., 2016). In short, a single guide RNA (sgRNA) is designed to recruit the Cas9 endonuclease protein and to match a specific, desired target site in the genome that must be immediately followed by a protospacer adjacent motif (PAM) sequence (5′-NGG-3′). Once complexed with Cas9, the sgRNA guides Cas9 to the target genome site. Cas9 then interacts with the PAM sequence and creates a double-strand break in the target site. The cell can either repair the double stranded break through non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Ran et al., 2013). NHEJ genome editing arises from introduction of a random mutation/insertion/deletion when the broken ends of DNA are directly ligated, which can cause the target gene to be knocked out (i.e., non-functional). Gene knockout provides insight into the role and criticality of a gene by assessing the effect of its absence. Alternatively, HDR genome editing uses a repair template flanked by 5′ and 3′ homologous arm sequences that match the up- and down-stream regions of the double-strand break. The repair template can be designed for gene knockout, introduction of a specific mutation/insertion/deletion, or genomic integration of a transgene(s)/entire expression construct (Ran et al., 2013).

Symbiodinium exhibits an asexual haploid vegetative stage (Santos and Coffroth, 2003) with sister chromatids developing in S-phase of the cell cycle (Watrin and Legagneux, 2003), but HDR has yet to be directly observed in *Symbiodinium*. Therefore, CRISPR/Cas9 genome editing of *Symbiodinium* may be restricted to NHEJ. *Ku70*, *Ku80*, and DNA ligase IV (genes central to NHEJ; Chu et al., 2015) are all expressed in *Symbiodinium* transcriptomes (Levin et al., 2016). That said, some evidence does suggest *Symbiodinium* can enter a transient sexual diploid stage (Chi et al., 2014; Wilkinson et al., 2015; Levin et al., 2016), which has been documented in other dinoflagellates (Figueroa et al., 2015). In yeast, ploidy shifts the dominant double-stranded break repair mechanism—diploid cells favor HDR, while haploid cells favor NHEJ (Lee et al., 1999). Moreover, genes specific to meiosis, a process during which HDR occurs (Thacker and Keeney, 2016), have been found in *Symbiodinium* genomes and transcriptomes (Chi et al., 2014; Lin et al., 2015; Rosic et al., 2015; Levin et al., 2016). *Msh4*, *Msh5*, and *Spo11-2* are all highly up-regulated at elevated temperatures (Levin et al., 2016), suggesting that HDR pathways in *Symbiodinium* are activated. *Brca2*, a gene that controls HDR (Holloman, 2011), is likewise up-regulated in heat stressed *Symbiodinium* (SM population: TR74441| c0_g1; MI population: TR63986| c0_g1; Levin et al., 2016). Hence, the potential for genomic integration of transgenes through HDR may improve if *Symbiodinium* are pre-stressed. HDR in *Symbiodinium* may also be increased by suppression of *Ku70*, *Ku80*, or DNA ligase IV (Chu et al., 2015).

The permanently condensed chromosomes of *Symbiodinium* could present an obstacle for CRISPR/Cas9 genome editing by possibly limiting access of sgRNAs to certain target sites. An additional challenge for genome editing is the abundance of multi-copy genes in the large *Symbiodinium* genomes. Gene redundancy can prevent knockout of gene function since the CRISPR/Cas9 system is not 100% efficient, meaning uncleaved functional gene copies can remain. Additionally, CRISPR/Cas9 targeting of genes with high copy numbers has been found to decrease cell proliferation and survival likely due to an increased frequency of DNA damage events (Aguirre et al., 2016). Also, design of sgRNAs requires a sequenced genome, but only three *Symbiodinium* genomes—each from a separate evolutionary lineage—are currently available.

As a first step to overcome some of these limitations, we analyzed the three published *Symbiodinium* genomes (types A1, B1, and F1; Shoguchi et al., 2013; Lin et al., 2015; Aranda et al., 2016) to identify conserved single copy genes. We then predicted a target site in each conserved gene with high sgRNA efficiency and specificity across the genomes (Supplementary Materials and Methods). Conserved target sites may permit CRISPR/Cas9 genome editing of *Symbiodinium* types that have yet to be sequenced. Our analysis revealed 1792 conserved single copy orthologs, 261 of which have an optimal target site compatible with all genomes (Supplementary Dataset 1a). The 261 single copy orthologs for CRISPR/Cas9 genome editing were enriched for a wide array of functional gene groups of interest, including cellular components for photosynthesis and biological pathways for oxidation-reduction and for response to UV-B (Supplementary Figure 1 and Supplementary Tables 2–4). Knockout of these genes would critically improve our understanding of *Symbiodinium* gene function, and if HDR is present in *Symbiodinium*, these sgRNA target sites could also be used to introduce genes of interest or entire *Symbiodinium* expression constructs into the genome. Furthermore, we identified sgRNA target sites in the type A1 genome scaffolds 710 and 484 (Aranda et al., 2016) immediately downstream from the potentially strong, constitutive *Symbiodinium* promoters discussed earlier (Supplementary Dataset 1b). Assuming HDR, reporter genes such as fluorescent proteins could be introduced at these sites to measure promoter activity.

The CRISPR/Cas9 system can be carried by plasmids that contain expression constructs for the Cas9, sgRNA, and in the case of HDR, the repair template with homologous arms. Target site cleavage is improved by increased CRISPR/Cas9 construct expression (Hsu et al., 2013), so strong endogenous promoters and terminators from *Symbiodinium* discussed earlier could be employed to drive transcription of Cas9 by *Symbiodinium*. However, transcription of sgRNAs requires RNA polymerase III (*Pol III*) rather than RNA polymerase II. Therefore, promoters specifically recognized by *Pol III* (e.g., promoter of the U6 snRNA gene) are needed. Such promoters have been isolated from other eukaryotes for sgRNA transcription; but, as discussed earlier, they contain motifs (e.g., TATA-box) that *Symbiodinium* lack (Goomer and Kunkel, 1992; Clarke et al., 2013). In *Symbiodinium*, 26 U6 snRNA gene copies have been identified (see Supplementary Table 5 in Shoguchi et al., 2013), one of which is unusually located in a cluster with U1, U2, U4,

U5, 5S, and spliced leader snRNA genes (type B1 genome scaffold 8131; Shoguchi et al., 2013). Thus, genomic sequences found upstream and downstream of these *Symbiodinium* U6 snRNA genes could be isolated and trialed in sgRNA expression constructs as potential promoters and terminators recognized by *Symbiodinium Pol III*. Alternatively, the CRISPR/Cas9 system can be introduced to cells as pre-complexed sgRNA and purified Cas9 protein, which can achieve higher genome editing specificity by ∼10-fold compared to CRISPR/Cas9 plasmids and also removes the need to optimize Cas9 codon usage or to find appropriate promoters that will express Cas9 or sgRNAs (Zuris et al., 2015).

INTRACELLULAR DELIVERY OF CONSTRUCTS AND COMPLEXES

Verified delivery of expression constructs into *Symbiodinium* was previously achieved using silicon carbide whiskers, which yielded very few transformants (ten Lohuis and Miller, 1998), and with *Agrobacterium*, which produced transient transformants that were unable to divide (Ortiz-Matamoros et al., 2015a). Low efficiency foreign DNA delivery may be due to obstruction by the thick, multilayer *Symbiodinium* cell covering comprised of an external polysaccharide or glycoprotein layer atop an internal cell wall (thecal plates and the pellicle) then finally the plasma membrane (Markell et al., 1992; Wakefield et al., 2000). To overcome this barrier, methods including high-voltage electroporation, bioballistics, microinjection, and viral transduction should be trialed. Continued exploration into *Symbiodinium* viruses may facilitate development of a compatible transduction system. Additionally, the first method to produce viable *Symbiodinium* protoplasts (cells with their cell wall removed) was developed (Levin et al., 2017b). Protoplasts have been instrumental in genetic manipulation of cell-walled organisms through somatic hybridization as well as by allowing for alternate DNA delivery methods (Davey et al., 2005). Protoplast-dependent methods such as polyethylene glycol-mediated transformation (Mathur and Koncz, 1998) and liposome-mediated transformation (Caboche, 1990) may improve efficiency of construct delivery into *Symbiodinium*. Cell walls also serve as a barrier to RNA/protein complexes like pre-complexed sgRNA and Cas9 protein. Thus, genome editing of *Symbiodinium* with pre-complexed sgRNA and Cas9 protein may require the use of protoplasts (Woo et al., 2015). Polyethylene glycol-mediated transformation (Woo et al., 2015), cationic lipid transformation (Zuris et al., 2015), and electroporation (Baek et al., 2016) have all been used to effectively deliver pre-complexed sgRNA and Cas9 protein through cell membranes of other eukaryotes that lacked cell walls.

CAN WE REDUCE CORAL BLEACHING WITH GENETICALLY ENHANCED *Symbiodinium?*

Coral reefs are the most diverse marine habitat per unit area (Reaka-Kudla et al., 1996; Knowlton et al., 2010) and provide world economies with nearly US$30 billion in net benefits from goods and services annually (Cesar et al., 2003). Climate change impact models predict that most reefs will be severely damaged or lost in this century unless immediate protection efforts are made (Hoegh-Guldberg et al., 2007; Pandolfi et al., 2011; Mora et al., 2016; Hughes et al., 2017) prompting calls for the development of novel mitigation and restoration approaches (Rinkevich, 2014; van Oppen et al., 2015, 2017; Piaggio et al., 2016). Exceptional genetic variability naturally exists within the genus *Symbiodinium*, suggesting that seeding vulnerable corals with more climate-change tolerant *Symbiodinium* variants could provide a means to reduce bleaching susceptibility of corals (van Oppen et al., 2015). Although, uptake of non-native *Symbiodinium* variants by corals may not be widely achievable since many coral species only associate with specific *Symbiodinium* types (LaJeunesse et al., 2004). Furthermore, shifts from innately less stress tolerant *Symbiodinium* types to more stress tolerant *Symbiodinium* types (e.g., from type C2 to D) can have negative impacts on a number of coral fitness traits including growth and fecundity (Little et al., 2004; Jones and Berkelmans, 2011).

Environmental bioengineering is an alternative strategy to safeguard against climate change (Solé, 2015; Piaggio et al., 2016). Microalgae, such as *Symbiodinium*, are clear and promising candidates for genetic engineering with the aim of regaining and preserving ecosystem-climate homeostasis (Solé, 2015) because they can significantly influence the health of entire ecosystems (Berkelmans and van Oppen, 2006; Kirk and Weis, 2016; Murray et al., 2016). Genetic engineering to increase stress tolerance of the *Symbiodinium* variants that are naturally harbored by at-risk corals holds potential to reduce bleaching susceptibility without negatively impacting the fitness of the coral host since existing *Symbiodinium*-coral partnerships would be preserved. *Fe-sod*, *Mn-sod Prxd*, and *Hsp70* genes from *Symbiodinium* (Levin et al., 2016; Gierz et al., 2017; Goyen et al., 2017) are standout candidates whose engineered up-regulation may enhance thermal and bleaching tolerance by reducing heat-induced oxidative damage, but thorough evaluation of how this artificial up-regulation contributes to long term fitness and the *Symbiodinium*-coral symbiosis would be mandatory.

Application of genetic engineering to support environmental management practices has been gaining momentum. Notably, sterile male mosquitoes have been engineered to control mosquito-borne diseases (Gabrieli et al., 2014). Field releases of the sterile males significantly reduced wild mosquito populations, supporting their value to disease control (Harris et al., 2012). Similarly, fungus-resistance has been engineered in American chestnut trees in order to restore the natural population that was nearly eradicated from the spread of a foreign fungus. Introduction of these transgenic trees into the wild may receive federal approval in just the next few years, which would make them the first threatened plant species to be restored through genetic engineering (Jacobs et al., 2013; Powell, 2014).

Considering the great promise shown by genetic engineering-based approaches to promote environmental health (Jacobs et al., 2013; Powell, 2014) and human health (Paine et al., 2005; Harris et al., 2012; Gabrieli et al., 2014), as well as to sustain

food security (Schroeder et al., 2013), it is logical for genetic engineering to be proposed as an important component of the growing repertoire of forward-looking coral reef management approaches (van Oppen et al., 2015; Piaggio et al., 2016). Due to the urgent need to protect coral reefs from climate change, the *Symbiodinium* research community must commit to an all-hands-on-deck attitude to achieve and extensively test genetic enhancement of *Symbiodinium* and other novel reef restoration strategies in the laboratory setting. In parallel, comprehensive cost-benefit-risk evaluation of the potential ecological and socioeconomic impacts from implementation of such strategies in the natural environment must be exhaustive before field-based trials are initiated. Additionally, transparent dialogs with policy makers, coral reef managers, and the general public need to be initiated now to begin the process of education and public acceptance of genetic engineering approaches for coral reef mitigation and restoration.

As we have discussed here, recent NGS breakthroughs have revealed natural genetic elements of *Symbiodinium* and their viruses (**Figure 1**). Based on these discoveries, we have developed a tailored genetic engineering framework for *Symbiodinium* based on empirical data that may also be applicable to other dinoflagellate genera. In doing so, we have opened a new prospective avenue to decode *Symbiodinium* functional genomics that may ultimately allow for engineering increased stress tolerance of *Symbiodinium* to reduce coral bleaching.

AUTHOR CONTRIBUTIONS

RL conceived the manuscript concept, analyzed NGS data, and wrote the manuscript. CRV analyzed NGS data and critically edited the theory and writing of the manuscript. SA analyzed NGS data. PS, DS, and MvO critically edited the theory and writing of the manuscript.

FUNDING

Funding from the University of New South Wales and King Abdullah University of Science and Technology (KAUST) supported the analyses presented here.

ACKNOWLEDGMENTS

We thank the Aranda and Voolstra lab for providing the type A1 (*Symbiodinium microadriaticum*) genome prior to its publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01220/full#supplementary-material

REFERENCES

Aguirre, A. J., Meyers, R. M., Weir, B. A., Vazquez, F., Zhang, C.-Z., Ben-David, U., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR-Cas9 targeting. *Cancer Discov.* 6, 914–929. doi: 10.1158/2159-8290.CD-16-0154

Alavi, M., Miller, T., Erlandson, K., Schneider, R., and Belas, R. (2001). Bacterial community associated with Pfiesteria-like dinoflagellate cultures. *Environ. Microbiol.* 3, 380–396. doi: 10.1046/j.1462-2920.2001.00207.x

Allera-Moreau, C., Delluc-Clavières, A., Castano, C., Van den Berghe, L., Golzio, M., Moreau, M., et al. (2007). Long term expression of bicistronic vector driven by the FGF-1 IRES in mouse muscle. *BMC Biotechnol.* 7:74. doi: 10.1186/1472-6750-7-74

Aranda, M., Li, Y., Liew, Y., Baumgarten, S., Simakov, O., Wilson, M., et al. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci. Rep.* 6:39734. doi: 10.1038/srep39734

Arif, C., Daniels, C., Bayer, T., Banguera-Hinestroza, E., Barbrook, A., Howe, C. J., et al. (2014). Assessing *Symbiodinium* diversity in scleractinian corals via next-generation sequencing-based genotyping of the ITS2 rDNA region. *Mol. Ecol.* 23, 4418–4433. doi: 10.1111/mec.12869

Bachvaroff, T. R., and Place, A. R. (2008). From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS ONE* 3:e2929. doi: 10.1371/journal.pone.0002929

Baek, K., Kim, D. H., Jeong, J., Sim, S. J., Melis, A., Kim, J.-S., et al. (2016). DNA-free two-gene knockout in *Chlamydomonas reinhardtii* via CRISPR-Cas9 ribonucleoproteins. *Sci. Rep.* 6:30620. doi: 10.1038/srep30620

Barshis, D. J., Ladner, J. T., Oliver, T. A., and Palumbi, S. R. (2014). Lineage-specific transcriptional profiles of *Symbiodinium* spp. unaltered by heat stress in a coral host. *Mol. Biol. Evol.* 31, 1343–1352. doi: 10.1093/molbev/msu107

Baumgarten, S., Bayer, T., Aranda, M., Liew, Y. J., Carr, A., Micklem, G., et al. (2013). Integrating microRNA and mRNA expression profiling in *Symbiodinium microadriaticum*, a dinoflagellate symbiont of reef-building corals. *BMC Genomics* 14:704. doi: 10.1186/1471-2164-14-704

Bayer, T., Aranda, M., Sunagawa, S., Yum, L. K., DeSalvo, M. K., Lindquist, E., et al. (2012). *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS ONE* 7:e35269. doi: 10.1371/journal.pone.0035269

Benfey, P. N., and Chua, N.-H. (1990). The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* 250, 959–966. doi: 10.1126/science.250.4983.959

Berkelmans, R., and van Oppen, M. J. (2006). The role of zooxanthellae in the thermal tolerance of corals: a 'nugget of hope'for coral reefs in an era of climate change. *Proc. R. Soc. Lond. B Biol. Sci.* 273, 2305–2312. doi: 10.1098/rspb.2006.3567

Boulotte, N. M., Dalton, S. J., Carroll, A. G., Harrison, P. L., Putnam, H. M., Peplow, L. M., et al. (2016). Exploring the *Symbiodinium* rare biosphere provides evidence for symbiont switching in reef-building corals. *ISME J.* 10, 2693–2701. doi: 10.1038/ismej.2016.54

Breyer, D., Kopertekh, L., and Reheul, D. (2014). Alternatives to antibiotic resistance marker genes for in vitro selection of genetically modified plants–scientific developments, current use, operational access and biosafety considerations. *Crit. Rev. Plant Sci.* 33, 286–330. doi: 10.1080/07352689.2013.870422

Byler, K. A., Carmi-Veal, M., Fine, M., and Goulet, T. L. (2013). Multiple symbiont acquisition strategies as an adaptive mechanism in the coral *Stylophora pistillata*. *PLoS ONE* 8:e59596. doi: 10.1371/journal.pone.0059596

Caboche, M. (1990). Liposome-mediated transfer of nucleic acids in plant protoplasts. *Physiol. Plant.* 79, 173–176. doi: 10.1111/j.1399-3054.1990.tb05882.x

Cesar, H., Burke, L., and Pet-Soede, L. (2003). *The Economics of Worldwide Coral Reef Degradation*. Arnhem: Cesar Environmental Economics Consulting (CEEC).

Chi, J., Parrow, M. W., and Dunthorn, M. (2014). Cryptic sex in *Symbiodinium* (Alveolata, Dinoflagellata) is supported by an inventory of meiotic genes. *J. Eukaryot. Microbiol.* 61, 322–327. doi: 10.1111/jeu.12110

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., et al. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* 33, 543–548. doi: 10.1038/nbt.3198

Clarke, B. D., Cummins, D. M., McColl, K. A., Ward, A. C., and Doran, T. J. (2013). Characterization of zebrafish polymerase III promoters for the expression of short-hairpin RNA interference molecules. *Zebrafish* 10, 472–479. doi: 10.1089/zeb.2012.0782

Coffroth, M. A., and Santos, S. R. (2005). Genetic diversity of symbiotic dinoflagellates in the genus *Symbiodinium*. *Protist* 156, 19–34. doi: 10.1016/j.protis.2005.02.004

Colanduoni, J. A., and Villafranca, J. J. (1986). Inhibition of *Escherichia coli* glutamine synthetase by phosphinothricin. *Bioorg. Chem.* 14, 163–169. doi: 10.1016/0045-2068(86)90026-X

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Correa, A. M., Ainsworth, T. D., Rosales, S. M., Thurber, A. R., Butler, C. R., and Thurber, R. L. V. (2016). Viral outbreak in corals associated with an in situ bleaching event: atypical herpes-like viruses and a new megavirus infecting *Symbiodinium*. *Front. Microbiol.* 7:127. doi: 10.3389/fmicb.2016.00127

Correa, A. M., Welsh, R. M., and Thurber, R. L. V. (2013). Unique nucleocytoplasmic dsDNA and + ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *ISME J.* 7, 13–27. doi: 10.1038/ismej.2012.75

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., and Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438, 90–93. doi: 10.1038/nature04056

Davey, M. R., Anthony, P., Power, J. B., and Lowe, K. C. (2005). Plant protoplast technology: current status. *Acta Physiol. Plant.* 27, 117–130. doi: 10.1007/s11738-005-0044-0

DeSalvo, M. K., Sunagawa, S., Fisher, P. L., Voolstra, C. R., Iglesias-Prieto, R., and Medina, M. (2010). Coral host transcriptomic states are correlated with *Symbiodinium* genotypes. *Mol. Ecol* 19, 1174–1186. doi: 10.1111/j.1365-294X.2010.04534.x

Edmunds, P. J., Pochon, X., Levitan, D. R., Yost, D. M., Belcaid, M., Putnam, H. M., et al. (2014). Long-term changes in *Symbiodinium* communities in *Orbicella annularis* in St. John, US Virgin Islands. *Mar. Ecol. Prog. Ser.* 506, 129–144. doi: 10.3354/meps10808

Figueroa, R. I., Dapena, C., Bravo, I., and Cuadrado, A. (2015). The hidden sexuality of alexandrium minutum: an example of overlooked sex in Dinoflagellates. *PLoS ONE* 10:e0142667. doi: 10.1371/journal.pone.0142667

Gabrieli, P., Smidler, A., and Catteruccia, F. (2014). Engineering the control of mosquito-borne infectious diseases. *Genome Biol.* 15, 535. doi: 10.1186/s13059-014-0535-7

Gierz, S. L., Forêt, S., and Leggat, W. (2017). Transcriptomic analysis of thermally stressed *Symbiodinium* reveals differential expression of stress and metabolism genes. *Front. Plant Sci.* 8:271. doi: 10.3389/fpls.2017.00271

Gonzalez, A., Jimenez, A., Vazquez, D., Davies, J., and Schindler, D. (1978). Studies on the mode of action of hygromycin B, an inhibitor of translocation in eukaryotes. *Biochim. Biophys. Acta* 521, 459–469. doi: 10.1016/0005-2787(78)90287-3

Goomer, R., and Kunkel, G. (1992). The transcriptional start site for a human U6 small nuclear RNA gene is dictated by a compound promoter element consisting of the PSE and the TATA box. *Nucleic Acids Res.* 20, 4903–4912. doi: 10.1093/nar/20.18.4903

Goyen, S., Pernice, M., Szabó, M., Warner, M. E., Ralph, P. J., and Suggett, D. J. (2017). A molecular physiology basis for functional diversity of hydrogen peroxide production amongst *Symbiodinium* spp.(*Dinophyceae*). *Mar. Biol.* 164, 46. doi: 10.1007/s00227-017-3073-5

Gurtu, V., Yan, G., and Zhang, G. (1996). IRES bicistronic expression vectors for efficient creation of stable mammalian cell lines. *Biochem. Biophys. Res. Commun.* 229, 295–298. doi: 10.1006/bbrc.1996.1795

Harris, A. F., McKemey, A. R., Nimmo, D., Curtis, Z., Black, I., Morgan, S. A., et al. (2012). Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nat. Biotechnol.* 30, 828–830. doi: 10.1038/nbt.2350

Hennige, S., Suggett, D. J., Warner, M. E., McDougall, K., and Smith, D. J. (2009). Photobiology of *Symbiodinium* revisited: bio-physical and bio-optical signatures. *Coral Reefs* 28, 179–195. doi: 10.1007/s00338-008-0444-x

Hoegh-Guldberg, O., Mumby, P. J., Hooten, A. J., Steneck, R. S., Greenfield, P., Gomez, E., et al. (2007). Coral reefs under rapid climate change and ocean acidification. *Science* 318, 1737–1742. doi: 10.1126/science.1152509

Holloman, W. K. (2011). Unraveling the mechanism of BRCA2 in homologous recombination. *Nat. Struct. Mol. Biol.* 18, 748–754. doi: 10.1038/nsmb.2096

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi: 10.1038/nbt.2647

Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., et al. (2017). Global warming and recurrent mass bleaching of corals. *Nature* 543, 373–377. doi: 10.1038/nature21707

Jacobs, D. F., Dalgleish, H. J., and Nelson, C. D. (2013). A conceptual framework for restoration of threatened plants: the effective model of American chestnut (*Castanea dentata*) reintroduction. *New Phytol.* 197, 378–393. doi: 10.1111/nph.12020

Jiang, J., Zhang, H., Kang, Y., Bina, D., Lo, C. S., and Blankenship, R. E. (2012). Characterization of the peridinin–chlorophyll a-protein complex in the dinoflagellate *Symbiodinium*. *Biochim. Biophys. Acta* 1817, 983–989. doi: 10.1016/j.bbabio.2012.03.027

Jones, A. M., and Berkelmans, R. (2011). Tradeoffs to thermal acclimation: energetics and reproduction of a reef coral with heat tolerant *Symbiodinium* type-D. *J. Mar. Biol.* 2011, 12. doi: 10.1371/journal.pone.0010437

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6, 1–10. doi: 10.1038/ncomms7925

Kirk, N. L., and Weis, V. M. (2016). "Animal–*Symbiodinium* symbioses: foundations of coral reef ecosystems," in *The Mechanistic Benefits of Microbial Symbionts*, ed. C. J. Hurst (Cham: Springer), 269–294. doi: 10.1007/978-3-319-28068-4_10

Knowlton, N., Brainard, R. E., Fisher, R., Moews, M., Plaisance, L., and Caley, M. J. (2010). "Coral reef biodiversity," in *Life in the World's Oceans: Diversity Distribution and Abundance*, ed. A. D. Mcintyre (Hoboken, NJ: John Wiley & Sons), 65–74. doi: 10.1002/9781444325508.ch4

Krueger, T., Fisher, P. L., Becker, S., Pontasch, S., Dove, S., Hoegh-Guldberg, O., et al. (2015). Transcriptomic characterization of the enzymatic antioxidants FeSOD, MnSOD, APX and KatG in the dinoflagellate genus *Symbiodinium*. *BMC Evol. Biol.* 15:48. doi: 10.1186/s12862-015-0326-0

Ladner, J. T., Barshis, D. J., and Palumbi, S. R. (2012). Protein evolution in two co-occurring types of *Symbiodinium*: an exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. *BMC Evol. Biol.* 12:217. doi: 10.1186/1471-2148-12-217

LaJeunesse, T. C., Thornhill, D. J., Cox, E. F., Stanton, F. G., Fitt, W. K., and Schmidt, G. W. (2004). High diversity and host specificity observed among symbiotic dinoflagellates in reef coral communities from Hawaii. *Coral Reefs* 23, 596–603. doi: 10.1007/s00338-004-0428-4

Lee, S. E., Pâques, F., Sylvan, J., and Haber, J. E. (1999). Role of yeast SIR genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths. *Curr. Biol.* 9, 767–770. doi: 10.1016/S0960-9822(99)80339-X

Leggat, W., Yellowlees, D., and Medina, M. (2011). Recent progress in *Symbiodinium* transcriptomics. *J. Exp. Mar. Biol. Ecol.* 408, 120–125. doi: 10.1016/j.jembe.2011.07.032

Levin, R. A., Beltran, V. H., Hill, R., Kjelleberg, S., McDougald, D., Steinberg, P. D., et al. (2016). Sex, scavengers, and chaperones: transcriptome secrets of divergent symbiodinium thermal tolerances. *Mol. Biol. Evol.* 33, 2201–2215. doi: 10.1093/molbev/msw119

Levin, R. A., Felsen, C. N., Yang, J., Lin, J. Y., Whitney, M. A., Nguyen, Q. T., et al. (2014). An optimized triple modality reporter for quantitative in vivo tumor imaging and therapy evaluation. *PLoS ONE* 9:e97415. doi: 10.1371/journal.pone.0097415

Levin, R. A., Suggett, D. J., Nitschke, M. R., van Oppen, M. J., and Steinberg, P. D. (2017b). Expanding the *Symbiodinium* (Dinophyceae, Suessiales) toolkit

through protoplast technology. *J. Eukaryot. Microbiol.* doi: 10.1111/jeu.12393 [Epub ahead of print].

Levin, R. A., Voolstra, C. R., Weynberg, K. D., and van Oppen, M. J. H. (2017a). Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *ISME J.* 11, 808–812. doi: 10.1038/ismej.2016.154

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., et al. (2015). The *Symbiodinium* kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* 350, 691–694. doi: 10.1126/science.aad0408

Little, A. F., Van Oppen, M. J., and Willis, B. L. (2004). Flexibility in algal endosymbioses shapes growth in reef corals. *Science* 304, 1492–1494. doi: 10.1126/science.1095733

Markell, D., Trench, R., and Iglesias-Prieto, R. (1992). Macromolecules associated with the cell walls of symbiotic dinoflagellates. *Symbiosis* 12, 19–31.

Martínez-Salas, E. (1999). Internal ribosome entry site biology and its use in expression vectors. *Curr. Opin. Biotechnol.* 10, 458–464. doi: 10.1016/S0958-1669(99)00010-5

Mathur, J., and Koncz, C. (1998). PEG-mediated protoplast transformation with naked DNA. *Methods Mol. Biol.* 82, 267–276. doi: 10.1385/0-89603-391-0:267

Miller, T. R., and Belas, R. (2006). Motility is involved in Silicibacter sp. TM1040 interaction with dinoflagellates. *Environ. Microbiol.* 8, 1648–1659. doi: 10.1111/j.1462-2920.2006.01071.x

Moldowan, J. M., and Talyzina, N. M. (1998). Biogeochemical evidence for dinoflagellate ancestors in the Early Cambrian. *Science* 281, 1168–1170. doi: 10.1126/science.281.5380.1168

Mora, C., Graham, N. A., and Nyström, M. (2016). Ecological limitations to the resilience of coral reefs. *Coral Reefs* 35, 1271–1280. doi: 10.1007/s00338-016-1479-z

Murray, S. A., Suggett, D. J., Doblin, M. A., Kohli, G. S., Seymour, J. R., Fabris, M., et al. (2016). Unravelling the functional genetics of dinoflagellates: a review of approaches and opportunities. *Perspect. Phycol.* 3, 37–52. doi: 10.1127/pip/2016/0039

Muscatine, L. (1990). The role of symbiotic algae in carbon and energy flux in reef corals. *Ecosyst. World* 25, 75–87.

Muscatine, L., and Porter, J. W. (1977). Reef corals: mutualistic symbioses adapted to nutrient-poor environments. *Bioscience* 27, 454–460. doi: 10.2307/1297526

Nehlsen, K., Broll, S., and Bode, J. (2006). Replicating minicircles: generation of nonviral episomes for the efficient modification of dividing cells. *Gene Ther. Mol. Biol.* 10, 233–244.

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

Ortiz-Matamoros, M. F., Islas-Flores, T., Voigt, B., Menzel, D., Baluška, F., and Villanueva, M. A. (2015a). Heterologous DNA uptake in cultured *Symbiodinium* spp. Aided by *Agrobacterium tumefaciens*. *PLoS ONE* 10:e0132693. doi: 10.1371/journal.pone.0132693

Ortiz-Matamoros, M. F., Villanueva, M. A., and Islas-Flores, T. (2015b). Transient transformation of cultured photosynthetic dinoflagellates (*Symbiodinium* spp.) with plant-targeted vectors Transformación de dinoflagelados fotosintéticos del género *Symbiodinium* en cultivo con vectores diseñados para plantas. *Ciencias Marinas* 41, 21–32. doi: 10.7773/cm.v41i1.2449

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., et al. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* 23, 482–487. doi: 10.1038/nbt1082

Pandolfi, J. M., Connolly, S. R., Marshall, D. J., and Cohen, A. L. (2011). Projecting coral reef futures under global warming and ocean acidification. *Science* 333, 418–422. doi: 10.1126/science.1204794

Parkinson, J. E., Baumgarten, S., Michell, C. T., Baums, I. B., LaJeunesse, T. C., and Voolstra, C. R. (2016). Gene expression variation resolves species and individual strains among coral-associated dinoflagellates within the genus *Symbiodinium*. *Genome Biol. Evol.* 8, 665–680. doi: 10.1093/gbe/evw019

Piaggio, A. J., Segelbacher, G., Seddon, P. J., Alphey, L., Bennett, E. L., Carlson, R. H., et al. (2016). Is it time for synthetic biodiversity conservation? *Trends Ecol. Evol.* 32, 97–107. doi: 10.1016/j.tree.2016.10.016

Pline, W. A., Lacy, G. H., Stromberg, V., and Hatzios, K. K. (2001). Antibacterial activity of the herbicide glufosinate on *Pseudomonas syringae* pathovar glycinea. *Pestic. Biochem. Physiol.* 71, 48–55. doi: 10.1006/pest.2001.2556

Pochon, X., and Gates, R. D. (2010). A new *Symbiodinium* clade (Dinophyceae) from soritid foraminifera in Hawai'i. *Mol. Phylogenet. Evol.* 56, 492–497. doi: 10.1016/j.ympev.2010.03.040

Pochon, X., Montoya-Burgos, J. I., Stadelmann, B., and Pawlowski, J. (2006). Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus *Symbiodinium*. *Mol. Phylogenet. Evol.* 38, 20–30. doi: 10.1016/j.ympev.2005.04.028

Powell, W. (2014). The american chestnut's genetic rebirth. *Sci. Am.* 310, 68–73. doi: 10.1038/scientificamerican0314-68

Quigley, K. M., Davies, S. W., Kenkel, C. D., Willis, B. L., Matz, M. V., and Bay, L. K. (2014). Deep-sequencing method for quantifying background abundances of *Symbiodinium* types: exploring the rare *Symbiodinium* biosphere in reef-building corals. *PLoS ONE* 9:e94297. doi: 10.1371/journal.pone.0094297

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

Reaka-Kudla, M. L., Wilson, D. E., and Wilson, E. O. (1996). *Biodiversity II: Understanding and Protecting our Biological Resources*. Washington, DC: Joseph Henry Press.

Rinkevich, B. (2014). Rebuilding coral reefs: does active reef restoration lead to sustainable reefs? *Curr. Opin. Environ. Sustain.* 7, 28–36. doi: 10.1016/j.cosust.2013.11.018

Ritchie, K. B. (2012). "Bacterial symbionts of corals and *Symbiodinium*," in *Beneficial Microorganisms in Multicellular Life Forms*, eds E. Rosenberg and U. Gophna (Berlin: Springer), 139–150. doi: 10.1007/978-3-642-21680-0_9

Rosic, N., Ling, E. Y. S., Chan, C.-K. K., Lee, H. C., Kaniewska, P., Edwards, D., et al. (2015). Unfolding the secrets of coral–algal symbiosis. *ISME J.* 9, 844–856. doi: 10.1038/ismej.2014.182

Santos, S. R., and Coffroth, M. A. (2003). Molecular genetic evidence that dinoflagellates belonging to the genus *Symbiodinium* Freudenthal are haploid. *Biol. Bull.* 204, 10–20. doi: 10.2307/1543491

Schroeder, J. I., Delhaize, E., Frommer, W. B., Guerinot, M. L., Harrison, M. J., Herrera-Estrella, L., et al. (2013). Using membrane transporters to improve crops for sustainable food production. *Nature* 497, 60–66. doi: 10.1038/nature11909

Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–909. doi: 10.1038/nmeth819

Shoguchi, E., Shinzato, C., Hisata, K., Satoh, N., and Mungpakdee, S. (2015). The large mitochondrial genome of *Symbiodinium* minutum reveals conserved noncoding sequences between Dinoflagellates and Apicomplexans. *Genome Biol. Evol.* 7, 2237–2244. doi: 10.1093/gbe/evv137

Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., et al. (2013). Draft assembly of the *Symbiodinium* minutum nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* 23, 1399–1408. doi: 10.1016/j.cub.2013.05.062

Solé, R. (2015). Bioengineering the biosphere? *Ecol. Complex.* 22, 40–49. doi: 10.1016/j.ecocom.2015.01.005

Stoykova, P., and Stoeva-Popova, P. (2011). PMI (manA) as a nonantibiotic selectable marker gene in plant biotechnology. *Plant Cell Tissue Organ Cult. (PCTOC)* 105, 141–148. doi: 10.1007/s11240-010-9858-6

Suggett, D. J., Warner, M. E., Smith, D. J., Davey, P., Hennige, S., and Baker, N. R. (2008). Photosynthesis and production of hydrogen peroxide by *symbiodinium* (Pyrrophyta) phylotypes with different thermal tolerances. *J Phycol.* 44, 948–956. doi: 10.1111/j.1529-8817.2008.00537.x

ten Lohuis, M. R., and Miller, D. J. (1998). Genetic transformation of dinoflagellates (Amphidinium and *Symbiodinium*): expression of GUS in microalgae using heterologous promoter constructs. *Plant J.* 13, 427–435. doi: 10.1046/j.1365-313X.1998.00040.x

Thacker, D., and Keeney, S. (2016). "Homologous recombination during meiosis," in *DNA Replication, Recombination, and Repair*, eds F. Hanaoka and K. Sugasawa (Tokyo: Springer), 131–151. doi: 10.1007/978-4-431-55873-6_6

Thomas, L., Kendrick, G., Kennington, W., Richards, Z., and Stat, M. (2014). Exploring *Symbiodinium* diversity and host specificity in Acropora corals from geographical extremes of Western Australia with 454 amplicon pyrosequencing. *Mol. Ecol.* 23, 3113–3126. doi: 10.1111/mec.12801

Toller, W. W., Rowan, R., and Knowlton, N. (2001). Repopulation of zooxanthellae in the Caribbean corals *Montastraea annularis* and *M. faveolata* following

experimental and disease-associated bleaching. *Biol. Bull.* 201, 360–373. doi: 10.2307/1543614

Tonk, L., Bongaerts, P., Sampayo, E. M., and Hoegh-Guldberg, O. (2013). SymbioGBR: a web-based database of *Symbiodinium* associated with cnidarian hosts on the Great Barrier Reef. *BMC Ecol.* 13:7. doi: 10.1186/1472-6785-13-7

Trench, R. K., and Blank, R. J. (1987). *Symbiodinium* Microadriaticum Freudenthal, *S. Goreauii* Sp. Nov., *S. Kawagutii* Sp. Nov. and *S. Pilosum* Sp. Nov.: Gymnodinioid Dinoflagellate symbionts of marine invertebrates 1. *J. Phycol.* 23, 469–481. doi: 10.1111/j.1529-8817.1987.tb02534.x

van Oppen, M. J., Gates, R. D., Blackall, L. L., Cantin, N., Chakravarti, L. J., Chan, W. Y., et al. (2017). Shifting paradigms in restoration of the world's coral reefs. *Glob. Chang Biol.* doi: 10.1111/gcb.13647 [Epub ahead of print].

van Oppen, M. J. H., Oliver, J. K., Putnam, H. M., and Gates, R. D. (2015). Building coral reef resilience through assisted evolution. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2307–2313. doi: 10.1073/pnas.1422301112

van Oppen, M. J. H., Palstra, F. P., Piquet, A. M. T., and Miller, D. J. (2001). Patterns of coral–dinoflagellate associations in Acropora: significance of local availability and physiology of *Symbiodinium* strains and host–symbiont selectivity. *Proc. R. Soc. Lond. B. Biol. Sci.* 268, 1759–1767. doi: 10.1098/rspb.2001.1733

Vicens, Q., and Westhof, E. (2003). Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide. *J. Mol. Biol.* 326, 1175–1188. doi: 10.1016/S0022-2836(02)01435-3

Wakefield, T. S., Farmer, M. A., and Kempf, S. C. (2000). Revised description of the fine structure of in situ" zooxanthellae" genus *Symbiodinium*. *Biol. Bull.* 199, 76–84. doi: 10.2307/1542709

Warner, M. E., Fitt, W. K., and Schmidt, G. W. (1999). Damage to photosystem II in symbiotic dinoflagellates: a determinant of coral bleaching. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8007–8012. doi: 10.1073/pnas.96.14.8007

Watrin, E., and Legagneux, V. (2003). Introduction to chromosome dynamics in mitosis. *Biol. Cell* 95, 507–513. doi: 10.1016/j.biolcel.2003.08.003

Weis, V. M. (2008). Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *J. Exp. Biol.* 211, 3059–3066. doi: 10.1242/jeb.009597

Weynberg, K. D., Neave, M., Clode, P. L., Voolstra, C. R., Brownlee, C., Laffy, P., et al. (2017). Prevalent and persistent viral infection in cultures of the coral algal endosymbiont *Symbiodinium*. *Coral Reefs* 2017, 1–12. doi: 10.1007/s00338-017-1568-7

Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A., and van Oppen, M. J. (2014). Generating viral metagenomes from the coral holobiont. *Front. Microbiol.* 5:206. doi: 10.3389/fmicb.2014.00206

Wilkinson, S. P., Fisher, P. L., van Oppen, M. J. H., and Davy, S. K. (2015). Intra-genomic variation in symbiotic dinoflagellates: recent divergence or recombination between lineages? *BMC Evol. Biol.* 15:46. doi: 10.1186/s12862-015-0325-1

Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., et al. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi: 10.1038/nbt.3389

Xiang, T., Nelson, W., Rodriguez, J., Tolleter, D., and Grossman, A. R. (2015). *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *Plant J.* 82, 67–80. doi: 10.1111/tpj.12789

Yuyama, I., Harii, S., and Hidaka, M. (2012). Algal symbiont type affects gene expression in juveniles of the coral Acropora tenuis exposed to thermal stress. *Mar. Environ. Res.* 76, 41–47. doi: 10.1016/j.marenvres.2011.09.004

Zhang, H., Zhuang, Y., Gill, J., and Lin, S. (2013). Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium* kawagutii as a case study. *Protist* 164, 510–527. doi: 10.1016/j.protis.2013.04.002

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., et al. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* 33, 73–80. doi: 10.1038/nbt.3081

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AWL declared a shared affiliation, though no other collaboration, with one of the authors DS to the handling Editor, who ensured that the process nevertheless met the standards of a fair and objective review.

Copyright © 2017 Levin, Voolstra, Agrawal, Steinberg, Suggett and van Oppen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 February 2018
doi: 10.3389/fmicb.2018.00114

The Divergence in Bacterial Components Associated with *Bactrocera dorsalis* across Developmental Stages

*Xiaofeng Zhao[1], Xiaoyu Zhang[1], Zhenshi Chen[1], Zhen Wang[1], Yongyue Lu[1] and Daifeng Cheng[1,2]**

[1] Department of Entomology, South China Agricultural University, Guangzhou, China, [2] Grouped Microorganism Research Center, South China Agricultural University, Guangzhou, China

Eco-evolutionary dynamics of microbiotas at the macroscale level are largely driven by ecological variables. The diet and living environment of the oriental fruit fly, *Bactrocera dorsalis*, diversify during development, providing a natural system to explore convergence, divergence, and repeatability in patterns of microbiota dynamics as a function of the host diet, phylogeny, and environment. Here, we characterized the microbiotas of 47 *B. dorsalis* individuals from three distinct populations by 16S rRNA amplicon sequencing. A significant deviation was found within the larvae, pupae, and adults of each population. Pupae were characterized by an increased bacterial taxonomic and functional diversity. Principal components analysis showed that the microbiotas of larvae, pupae, and adults clearly separated into three clusters. *Acetobacteraceae*, *Lactobacillaceae,* and *Enterobacteriaceae* were the predominant families in larval and adult samples, and PICRUSt analysis indicated that phosphoglycerate mutases and transketolases were significantly enriched in larvae, while phosphoglycerate mutases, transketolases, and proteases were significantly enriched in adults, which may support the digestive function of the microbiotas in larvae and adults. The abundances of *Intrasporangiaceae, Dermabacteraceae* (mainly *Brachybacterium*) and *Brevibacteriaceae* (mainly *Brevibacterium*) were significantly higher in pupae, and the antibiotic transport system ATP-binding protein and antibiotic transport system permease protein pathways were significantly enriched there as well, indicating the defensive function of microbiotas in pupae. Overall, differences in the microbiotas of the larvae, pupae, and adults are likely to contribute to differences in nutrient assimilation and living environments.

Keywords: *Bactrocera dorsalis*, development stage, microbial community, 16S rRNA, dietary, living environment

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Sze-Looi Song, University of Malaya, Malaysia
Zhanghong Shi, Fujian Agriculture and Forestry University, China
Vince Martinson, University of Rochester, United States

***Correspondence:**
Daifeng Cheng
chengdaifeng@scau.edu.cn

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *18 October 2017*
Accepted: *18 January 2018*
Published: *01 February 2018*

Citation:
Zhao X, Zhang X, Chen Z, Wang Z, Lu Y and Cheng D (2018) The Divergence in Bacterial Components Associated with Bactrocera dorsalis across Developmental Stages. Front. Microbiol. 9:114. doi: 10.3389/fmicb.2018.00114

INTRODUCTION

Many microorganisms reside on the insect exoskeleton, in the gut and hemocoel, and within insect cells (Douglas, 2015), and relationships ranging from parasitism to mutualism are built between microorganisms and insects (Berasategui et al., 2016). These microorganisms are often identified as symbionts of insects (Douglas, 2015). Most of the best-described associations among

mutualisms are based on nutritional or defensive services provided by the symbionts to their hosts. The host's physiology and behavior are often affected in such mutualisms, and the adaptability of the hosts is increased (Ohkuma and Brune, 2011; Ye et al., 2014). The symbionts provide nutrients, such as amino acids and vitamins, or digestive enzymes that aid in the degradation of fastidious dietary polymers or in the detoxification of noxious secondary metabolites (Douglas, 2009). The microorganisms can also protect their host against pathogens, parasites, parasitoids, or predators by producing toxins or antimicrobial compounds for defense (Teixeira et al., 2008; Florez et al., 2015; Hamby and Becher, 2016). Insects can generally employ the symbiont-produced defensive compounds in two different manners: (i) for the protection of the host or its offspring against antagonistic micro- or macroorganisms or (ii) as weed killers in insect fungiculture (Kaltenpoth, 2009; Ramadhar et al., 2014). Antimicrobial compounds produced by the defensive symbiont are of particular importance to insects living in enclosed, humid environments, where opportunistic fungal or bacterial infections can develop rapidly (Douglas, 2015). Studies have recently indicated that metabolic and adaptive abilities allow different bacteria to occupy their host during different host development stages and that the relationships can be multifaceted, varying in their impact on host biology (Lindow and Brandl, 2003; Turnbaugh et al., 2007; Knief et al., 2012). Hosts thus often exploit beneficial symbioses to augment their functional capabilities and to facilitate their adaptation to novel niches (Rio et al., 2006; Ye et al., 2014; Rafael et al., 2016).

For fruit fly, the pivotal roles of microbiota have been identified in recent years, and the factors that affected the structure of microbiota were also investigated. For example, the microbiotas during ontogeny of *Ceratitis capitata* have been reported to be shaped by phylogenetic, metabolic, and taxonomic diversities (Aharon et al., 2013). Some probiotics can even improve the fitness sexual performance of the males at emergence (Hamden et al., 2013). In *Drosophila melanogaster*, symbiotic bacteria play a role in mating preference by changing the levels of cuticular hydrocarbon sex pheromones (Sharon et al., 2010). For *Bactrocera dorsalis*, many studies have identified the structure and function of the gut microbioa. By culture-dependent and the 16S rRNA sequencing methods, the diversity of the cultivable gut bacterial communities associated with *B. dorsalis* have recently been investigated (Wang et al., 2011, 2014; Gujjar et al., 2017), and the development and drug resistantance of *B. dorsalis* were affected by the gut symbionts (Cheng et al., 2017; Khaeso et al., 2017).

Studies have indicated that diet and environment greatly influence the structure of microbiotas (Egert et al., 2004; Antwis et al., 2014; Rebollar et al., 2016). Microbial communities from the surrounding environment can even serve as reservoirs and source pools of colonizers (Kueneman et al., 2014; Loudon et al., 2014). *B. dorsalis* undergoes great changes in living environment during its life, as eggs and larvae in fruit (it is of great possibility to be infected by microbes for *B. dorsalis* larvae in the rotten fruits), pupae (especially in the early stage of pupation) in the ground (enclosed, humid environments, where opportunistic fungal (especially the *Metarhizium* and *Beauveria*) can develop rapidly (Vänninen et al., 2000) and adults on the branches of fruit trees. In addition, larvae must feed on food with high sugar content, the adults must feed on food with high sugar and protein content, and pupae do not eat. And we even found the control efficiency of *Metarhizium* and *Beauveria* to *B. dorsalis* in the pupal stage is very low (data unpublished). These traits may result in differences in the microbiotas of *B. dorsalis*, and *B. dorsalis* may rely on multiple microbial species for fitness and provide a unique model to investigate and compare the population dynamics of symbionts that display varying levels of integration with host biology. Available data on the microbiota of *B. dorsalis* are unfortunately limited, which restricts understanding the microbiota's influence on host traits, such as diet and living environment.

The larvae and adults of *B. dorsalis* must feed on large amounts of high sugar content food, and the larvae and pupae are exposed to environments with many pathogenic microorganisms. We thus proposed the hypothesis that the symbionts of *B. dorsalis* will change with the development stages: in the larval and adult stages, symbionts promote the host's absorption of sugar, and some symbionts in the larval and pupal stages may also generate antibiotics to enhance resistance to the pathogens. We explored three questions in the current study. We first examined whether differences in the living environments between adults, larvae, and pupae of *B. dorsalis* are reflected in differences in their bacterial communities, for example more defensive bacteria in larvae and pupae. Second, as larvae and adults must feed on high sugar content food, we tested the hypothesis that functional gene abundances in larval and adult microbiotas would reflect their capacities for sugar and protein metabolism. Finally, we investigated whether functional gene abundances in larval and pupal microbiotas reflect the resistance to pathogens.

MATERIALS AND METHODS

Rearing and Collection of *B. dorsalis*

The lab population of *B. dorsalis* was collected from a carambola (*Averrhoa carambola*) orchard (N 23° 06′ 53.09″, E 113° 24′ 51.29″) in Guangzhou, Guangdong Province in April 2008 and was reared as previously described (Cheng et al., 2017). Briefly, the flies were reared under the following conditions: 25 ± 1°C; 16:8 h light:dark cycle; 70–80% relative humidity (RH). The flies were reared with artificial diets which were treated with high pressure sterilization. Larvae, pupae, and adults of two wild populations were also collected from the cities Huizhou (HZ population) (N 23° 25′ 56.00″, E 114°, 28′ 16.61″) and Nansha (NS population) (N 22° 42′ 25.81″, E 113° 33′ 6.30″) of Guangdong Province in June 2017. For wild populations, carambolas with larvae were collected and take into the lab. Then pupation and eclosion processes went on in the sterile sands. Seventeen samples were collected for the lab population (six larvae, six pupae, and five adults); 15 samples were collected for the HZ and NS population (five larvae, five pupae, and five adults). Each sample consists of one individual.

Bacterial Community Characterization

All samples (Whole individuals) were selected and then washed with sterile water. The washed samples were transferred to centrifuge tubes containing DNA extraction buffer (with lysozyme) and grinded with a homogenizer. Then total DNA of the samples was extracted using a DNA extraction kit (Tiangen, Beijing, China) following the manufacturer's instructions. After the DNA of the samples was prepared, qPCR was used to estimate the bacteria absolute content of the samples with the universal bacterial 16S rRNA primers. A standard curve for qPCR was generated by amplifying the 16S rDNA of the *Arthrobacter* sp. isolated from the pupa of *B. dorsalis*. Approximately 465 bp of the V3–V4 region of the bacterial 16S rDNA gene was amplified by PCR according to a standard protocol. The following primers were used: F, 5′-CCTACGGGNGGCWGCAG-3′ and R, 5′-GGACTACHVGGGTATCTAAT-3′. The primers contained the A and B adapters for 454 Life Sciences pyrosequencing and a unique 12-bp error-correcting Golay barcode, which allowed multiplexing of samples in a single run. Each sample was analyzed in a total reaction volume of 25 μL that contained 2.5 μL of Takara 10× Ex Taq buffer, 1.5 μL of Mg^{2+} (25 mM), 2 μL of dNTP (2.5 mM), 0.25 μL of Takara Ex Taq (2.5 U/μL), 0.5 μL of each primer (10 μM), 16.75 μL of ddH_2O, and 1 μL of template. The PCR amplifications were performed with a 2-min incubation at 95°C followed by 30 cycles of 94°C for 30 s, 57°C for 30 s and 72°C for 30 s, and a final 5-min extension at 72°C. The PCR products were purified using QIAGEN MinElute PCR Purification Kit (QIAGEN, Hilden, Germany) to remove unincorporated primers and nucleotides. A microspectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, DE, United States) was used to measure the concentration of the purified DNA. The purified DNA was sequenced using an Illumina sequencing kit and an Illumina MiSeq sequencer (Illumina, San Diego, CA, United States).

Paired Illumina reads were merged in QIIME (Caporaso et al., 2010). After the high-quality reads were obtained, the data were filtered to remove low-complexity sequences (such as poly-A sequences) and sequences with ambiguous nucleotides, and the operational taxonomic units (OTUs) were picked using USEARCH (Edgar, 2010). The number of OTUs was calculated with mothur software (Schloss et al., 2009) at 97% similarity. An RDP classifier (Huse et al., 2010) was used with naïve Bayes settings for species annotation; the confidence threshold was set to 0.5. Using the species annotations and the read numbers of the OTUs, we generated OTU abundance profiles for all samples. OTUs with an abundance <0.005% of the total data set were removed as an additional level of quality filtering (Bokulich et al., 2013; Navas-Molina et al., 2013).

Diversity Analyses

Alpha diversity and Shannon rarefaction curves were calculated for all samples in mothur to investigate the species richness of the samples (v.1.34.0) (Schloss et al., 2009). Bray–Curtis and unweighted UniFrac distance matrices were used to calculate the beta diversity and were visualized with principal components analysis (PCA). To determine whether bacterial communities differed among host species, we conducted a shared and unique OTU analysis on the basis of an OTU table generated by QIIME. We used the unweighted pair group method with arithmetic mean (UPGMA), a hierarchical clustering method based on the arithmetic mean, to determine clustering patterns across host species. The UPGMA was used on Bray–Curtis distances of mean OTU relative abundances at the family level. The UPGMA, Bray–Curtis calculations and resulting heatmap were completed using the vegan package (Oksanen et al., 2015) in the R statistical package. Putative microbiota functions were predicted by annotating pathways of OTUs against the KEGG database using PICRUSt (Langille et al., 2013).

Statistical Analysis

The differences between treatments were compared by a one-way analysis of variance (ANOVA), followed by Tukey's test for multiple comparisons. The differences were considered significant at the $P < 0.05$ level. The data were analyzed using SPSS software. Analysis of similarity (ANOSIM) for the bacterial community of *B. dorsalis* across developmental stages were done with PRIMER 7.0.

RESULTS

Symbionts Content Quantification and Sequencing Data of 16S rRNA

Absolute content of the symbionts in flies were identified with qPCR, the results showed the symbionts content in each individual was about 10^6 CFU and there is no difference for the absolute content of the symbionts in different development stage (lab: $F_{2,14} = 0.126$, $P = 0.883$; Huizhou: $F_{2,12} = 0.717$, $P = 0.508$; and Nansha: $F_{2,12} = 1.768$, $P = 0.212$) (**Supplementary Figure S1**). After the sequencing data were subjected to demultiplexing, quality filtering and chimera removal, 50082–57433 reads were retained for the 17 lab population samples, 50176–57172 reads were obtained for the 15 YC samples, and 50037–56599 reads were obtained for the 15 NS samples (**Supplementary Table S1**). Shannon rarefaction curves for all samples showed a plateau stage, indicating adequate sampling of 16S rRNA sequences for all the samples (**Supplementary Figure S2**).

Differences in Larval, Pupal, and Adult Bacterial Communities

Significantly more OTUs were generated in the pupal samples of the three populations (lab: $F_{2,14} = 30.387$, $P < 0.001$; Huizhou: $F_{2,12} = 5.746$, $P = 0.018$; and Nansha: $F_{2,12} = 5.116$, $P = 0.025$) (**Supplementary Figure S3**). The ACE value and the Shannon and Simpson indices indicated that pupae exhibited the greatest species richness and that the richness of adults and larvae did not significantly differ (**Supplementary Table S2**). A major trend is clearly seen during development stages: the bacterial diversities of all populations were closest, on average, during the feeding

stage (larvae and adults), while pupae in complex environments showed greater bacterial diversity ($P < 0.01$).

Larvae and adults were significantly enriched in *Proteobacteria* at the phylum level more than pupae (for larvae: a cumulative relative abundance above 60, 44.3, and 91.9% in larvae for the lab, Huizhou and Nansha populations, respectively, $P < 0.01$; and for adults: a cumulative relative abundance above 57.82, 54.32, and 92.12% for the lab, Huizhou and Nansha populations, respectively, $P < 0.01$). *Actinobacteria* had a cumulative relative abundance in pupae above 20.54, 26.01, and 19.23% for the lab, Huizhou and Nansha populations, respectively (**Figure 1**).

The abundance of the dominant OTUs in each stage was compared with that in the other two stages (**Figure 2** and **Supplementary Data Sheet S1**).

For lab population, the most abundant OTUs (average relative abundance ± SD between replicates) in larvae were *Acetobacteraceae* (*Acetobacter* sp.) (54.62 ± 8.47), *Lactobacillaceae* (*Lactobacillus brevis*) (24.49 ± 3.88),

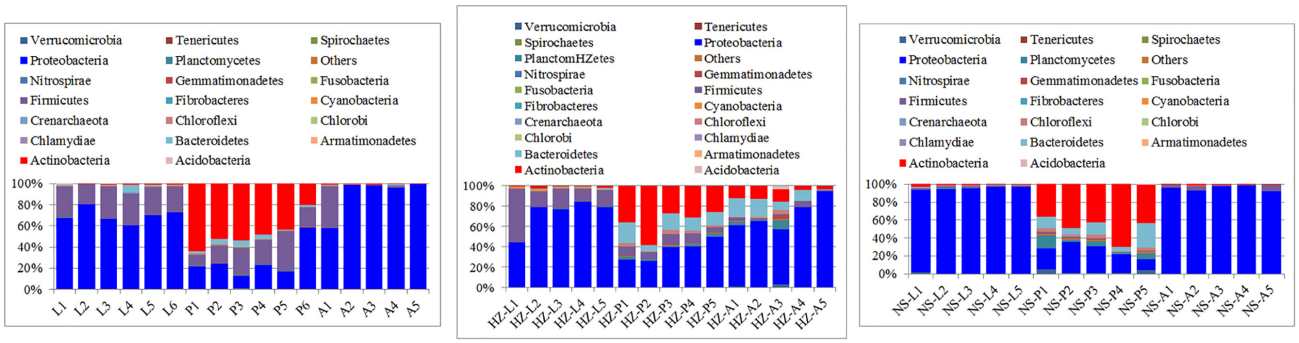


FIGURE 1 | Taxonomic compositions of microbiotas at the phylum level. Bars show proportions of taxa per individual as averaged across conspecifics and estimated from the rarefied OTU table. 'Others' group shows all phyla with relative abundance below 1% over the total number of reads. Lab population (L: larvae, P: pupae, A: adults); Huizhou population (HZ-L: larvae, HZ-P: pupae, HZ-A: adults); and Nansha population (NS-L: larvae, NS-P: pupae, NS-A: adults).

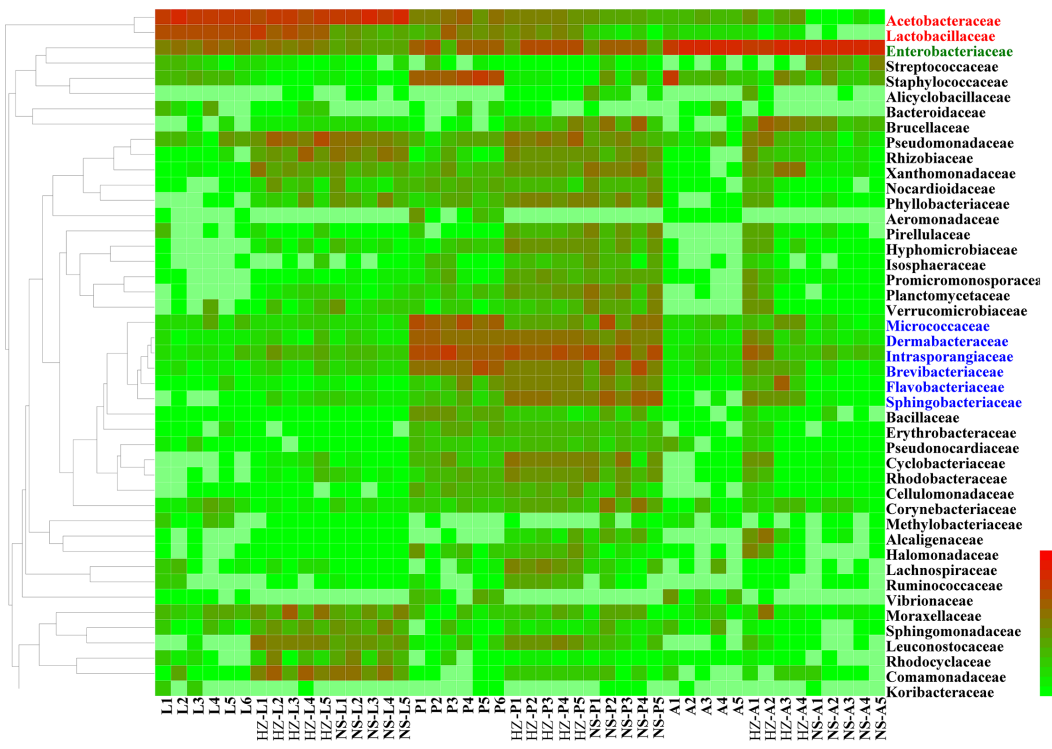


FIGURE 2 | Unweighted pair group method with arithmetic mean (UPGMA) and heatmap of bacterial families with relative abundances across different samples. Rows are bacterial families. Columns are samples. Colors indicate taxa with a higher (red) or lower (green) relative abundance in each sample. Taxonomic units in red, green and blue have significantly higher abundances in larvae, adults and pupae, respectively. Lab population (L: larvae, P: pupae, A: adults); Huizhou population (HZ-L: larvae, HZ-P: pupae, HZ-A: adults); and Nansha population (NS-L: larvae, NS-P: pupae, NS-A: adults).

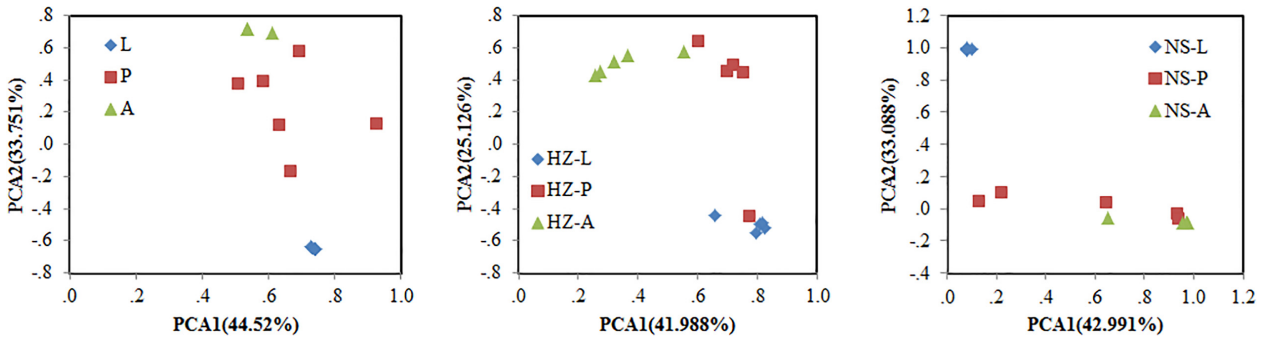


FIGURE 3 | Principal components analysis (PCA) of bacterial communities (genus level) according to development stages of the three populations (lab, Huizhou and Nansha populations). Taxonomic (OTU) clustering based on unweighted UniFrac distances. Lab population (L: larvae, P: pupae, A: adults); Huizhou population (HZ-L: larvae, HZ-P: pupae, HZ-A: adults); and Nansha population (NS-L: larvae, NS-P: pupae, NS-A: adults).

Enterobacteriaceae (4.94 ± 2.2) and *Acetobacteraceae* (*Gluconobacter* sp.) (4.77 ± 0.88); in pupae, the most abundant OTUs were *Micrococcaceae* (12.15 ± 8.54), *Intrasporangiaceae* (7.03 ± 4.97), *Brevibacteriaceae* (mainly *Brevibacterium*) (6.51 ± 6.28), and *Dermabacteraceae* (mainly *Brachybacterium*) (5.36 ± 3.01); in adults the most abundant OTU was *Enterobacteriaceae* (84.74 ± 18.11) (**Figure 2** and **Supplementary Data Sheet S1**).

For Huizhou population, the most abundant OTUs in larvae were *Acetobacteraceae* (*Acetobacter* sp.) (20.25 ± 3.11) and *Lactobacillaceae* (*Lactobacillus brevis*) (10.08 ± 3.48); the most abundant OTUs in pupae belonged to *Intrasporangiaceae* (10.79 ± 1.53), *Dermabacteraceae* (mainly *Brachybacterium*) (4.03 ± 2.22), and *Brevibacteriaceae* (mainly *Brevibacterium*) (1.63 ± 0.97); and the most abundant OTU in adults was *Enterobacteriaceae* (26.96 ± 9.52) (**Figure 2** and **Supplementary Data Sheet S1**).

For Nansha population, the most abundant OTUs in larvae were *Acetobacteraceae* (*Acetobacter* sp.) (54.04 ± 10.66); in pupae, the most abundant OTUs were *Intrasporangiaceae* (14.72 ± 1.65), *Dermabacteraceae* (mainly *Brachybacterium*) (2.36 ± 0.83), and *Brevibacteriaceae* (mainly *Brevibacterium*) (1.23 ± 0.67); and the most abundant OTU in adults was *Enterobacteriaceae* (18.89 ± 7.08) (**Figure 2** and **Supplementary Data Sheet S1**).

In conclusion, we found *Acetobacteraceae* (*Acetobacter* sp.) was the most abundant OTU in larvae of the three populations; *Intrasporangiaceae, Dermabacteraceae* (mainly *Brachybacterium*), and *Brevibacteriaceae* (mainly *Brevibacterium*) were the most abundant OTUs in pupa of the three populations; *Enterobacteriaceae* was the most abundant OTU in adult of the three populations (**Figure 2** and **Supplementary Data Sheet S1**).

Bacterial communities of larvae, pupae, and adults showed a clear pattern of specialization based on unweighted UniFrac distances with OTUs annotated at the genus level (PCA, **Figure 3**), indicating the specialization of larvae, pupae, and adults in hosting OTUs unique to each stage. PCA was also used to compare the similarity in the microbial community compositions of all samples of the three populations. Larvae, pupae, and adults each formed a distinct cluster among all samples, and these three clusters were separated from each other (**Figure 4**). The clustering pattern among samples was not influenced by the sampling population, as samples from the three populations clustered together, with the exception that results from pupal samples formed two different clusters. Moreover, ANOSIM results indicated that there were significant differences in the bacterial community of *B. dorsalis* across developmental stages (lab: $R = 0.998$, $P = 0.001$; Huizhou: $R = 0.745$, $P = 0.001$; and Nansha: $R = 0.951$, $P = 0.001$).

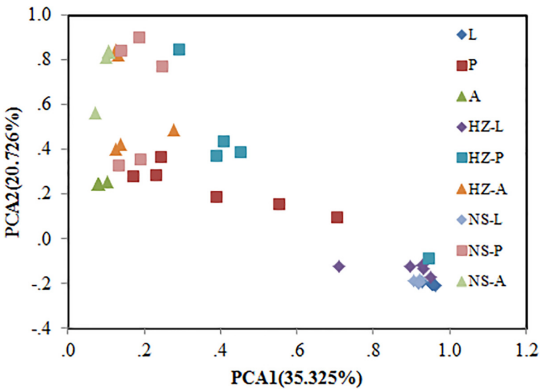


FIGURE 4 | Bray–Curtis distances-based PCA of all samples of the three populations (lab, Huizhou and Nansha populations). Lab population (L: larvae, P: pupae, A: adults); Huizhou population (HZ-L: larvae, HZ-P: pupae, HZ-A: adults); and Nansha population (NS-L: larvae, NS-P: pupae, NS-A: adults).

Functional Prediction of Larval, Pupal, and Adult Microbiotas

We addressed whether increased OTU diversity confers the host with a higher functional diversity. We predicted 4364, 4590, and 4308 level 3 KEGG Orthology (KO) groups in the predicted metagenomes of the lab, Huizhou and Nansha populations, respectively (PICRUSt analysis, **Supplementary Data Sheet S2**). The pattern of functional diversity largely followed the trend in taxonomic diversity: pupae displayed greater bacterial taxonomic and functional diversity than larvae and adults (lab: $R^2 = 0.587$,

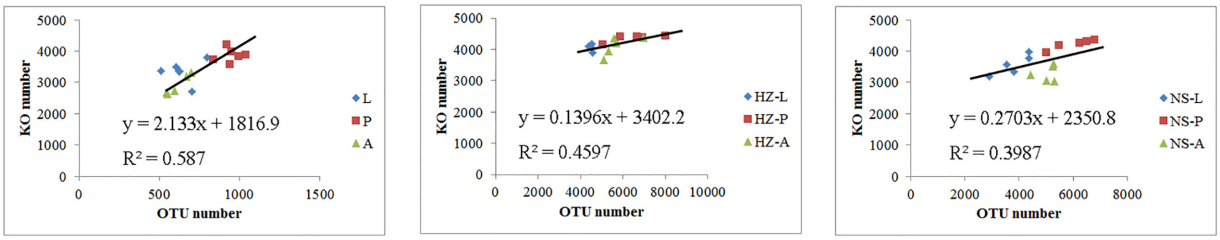


FIGURE 5 | Pearson relationship analysis of KO number and OTU number. Lab population (L: larvae, P: pupae, A: adults); Huizhou population (HZ-L: larvae, HZ-P: pupae, HZ-A: adults); and Nansha population (NS-L: larvae, NS-P: pupae, NS-A: adults).

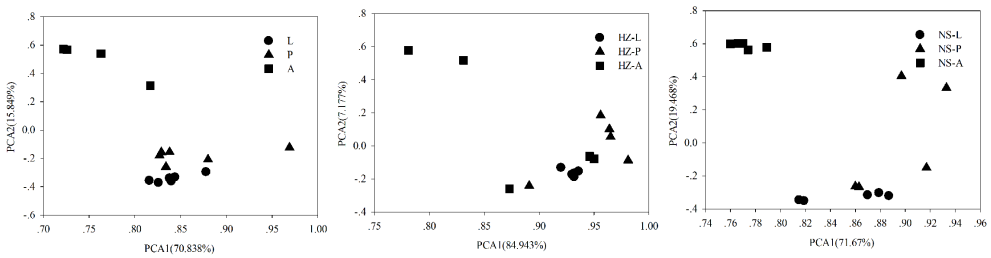


FIGURE 6 | Principal components analysis of functional diversity according to development stages of the three populations (lab, Huizhou and Nansha populations). Lab population (L: larvae, P: pupae, A: adults); Huizhou population (HZ-L: larvae, HZ-P: pupae, HZ-A: adults); and Nansha population (NS-L: larvae, NS-P: pupae, NS-A: adults).

$P < 0.01$; Huizhou: $R^2 = 0.678$, $P = 0.005$; and Nansha: $R^2 = 0.3987$, $P = 0.012$; Pearson relationship, **Figure 5**). The microbial community clusters of larvae, pupae, and adults were significantly separated at both the OTU and KO levels (**Figures 3**, **6**), indicating a strong development effect.

PICRUSt analysis predicted that phosphoglycerate mutase was significantly more abundant in larvae and adults than in pupae (lab population: $F_{2,14} = 81.873$, $P < 0.01$; Huizhou: $F_{2,12} = 25.055$, $P < 0.01$; and Nansha: $F_{2,12} = 307.792$, $P < 0.01$; **Figure 7A**). Phosphoglycerate mutase is a key enzyme in glucose metabolism and involved in converting 3-phosphoglyceric acid into 2-phosphoglyceric acid. Significantly greater number of OTUs were annotated for "antibiotic transport system ATP-binding proteins" and "antibiotic transport system permease proteins" in pupae (for antibiotic transport system ATP-binding proteins, lab population: $F_{2,14} = 37.387$, $P < 0.01$; Huizhou: $F_{2,12} = 32.737$, $P < 0.01$; and Nansha: $F_{2,12} = 37.762$, $P < 0.01$; for antibiotic transport system permease proteins, lab population: $F_{2,14} = 33.668$, $P < 0.01$; Huizhou: $F_{2,12} = 52.155$, $P < 0.01$; and Nansha: $F_{2,12} = 89.963$, $P < 0.01$; **Figures 7B,C**). And significantly greater number of OTUs were annotated for transketolase, which is involved in the pentose phosphate pathway, in adults (lab population: $F_{2,14} = 67.256$, $P < 0.01$; Huizhou: $F_{2,12} = 16.496$, $P < 0.01$; and Nansha: $F_{2,12} = 35.811$, $P < 0.01$; **Figure 7D**). Moreover, a significantly greater number of OTUs were annotated with the protease involved in the protein metabolism pathway in adults, which may indicate a function in protein metabolism (lab population: $F_{2,14} = 140.849$, $P < 0.01$; Huizhou: $F_{2,12} = 69.395$, $P < 0.01$; and Nansha: $F_{2,12} = 78.369$, $P < 0.01$; **Figure 7E**).

DISCUSSION

Although studies on fly microbiotas have been reported (Aharon et al., 2013; Aksoy et al., 2014; Augustinos et al., 2015; Michael et al., 2016; Cheng et al., 2017; Yong et al., 2017b; Zhao et al., 2017), little is known about the microbial community differences during different development stages, and the former studies mainly focused on the gut microbiotas of the flies (Aksoy et al., 2014; Augustinos et al., 2015; Michael et al., 2016; Cheng et al., 2017; Zhao et al., 2017). The number of OTUs observed in the larvae, pupae, and adults of *B. dorsalis* in this study is greater than those reported in other fly gut samples (**Figure 1**), suggesting that the function of the microbial community in *B. dorsalis* must be analyzed in detail. Actually microbiomes associated with *B. dorsalis* have been reported in earlier studies. For example, Andongma et al. (2015) have reported that the dominance of *Firmicutes* in adult stages and *Proteobacteria* in immature stages (Andongma et al., 2015); Wang et al. (2011) and Yong et al. (2017b) revealed Proteobacteria (specifically Gammaproteobacteria) to be predominant in the male adults of *B. dorsalis* (Wang et al., 2011; Yong et al., 2017b). In our study *Proteobacteria* was the dominant phylum in larvae and adults, which is consistent with previous studies (Aksoy et al., 2014; Yong et al., 2017b); however, *Actinobacteria* were mainly found in pupae, which may indicate that the pupal microbiota has a different function. Although Andongma et al. (2015) investigated symbiotic bacterial populations across life stages of *B. dorsalis*, many fewer OTUs were identified in their study; moreover, fewer *Actinobacteria* were not identified in pupa in their study (Andongma et al., 2015). The DNA extraction method may

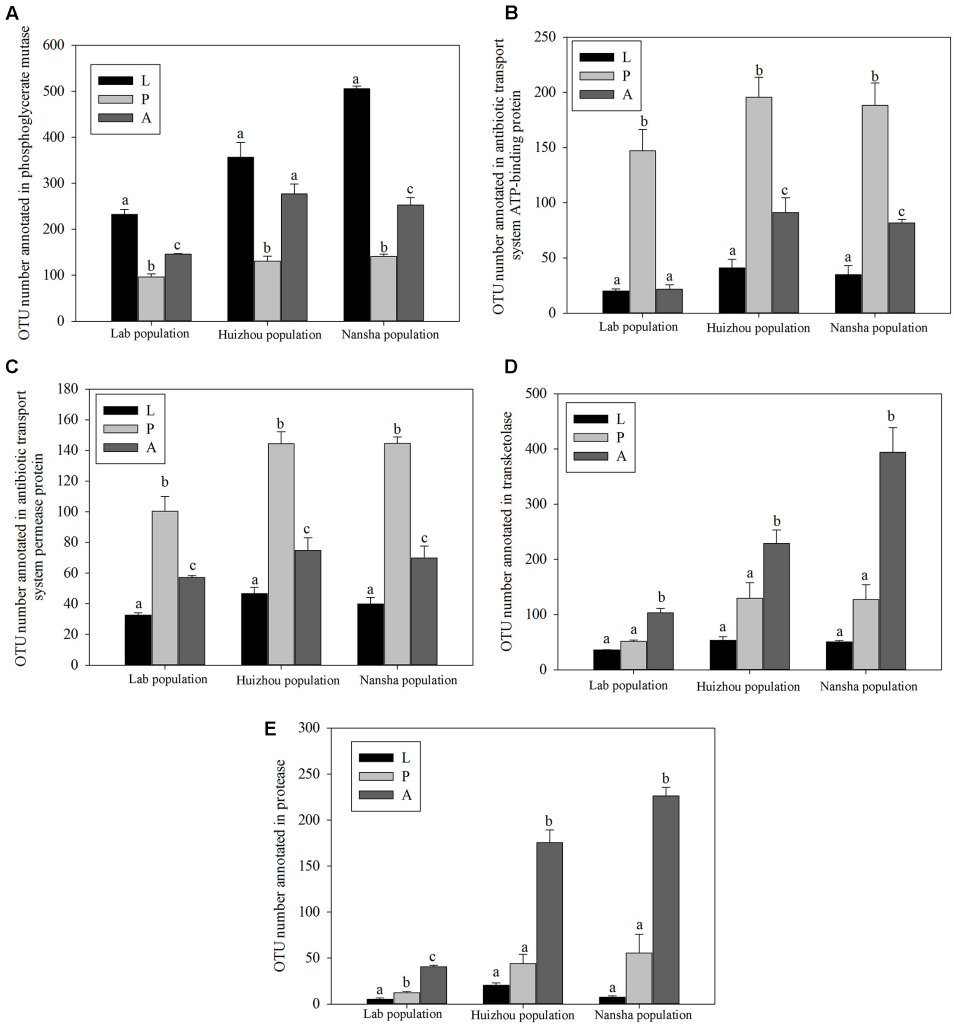


FIGURE 7 | Comparison of predicted KEGG ortholog counts. Means ± SEMs labeled with different letters are significantly different. OTU numbers annotated for phosphoglycerate mutase **(A)**, antibiotic transport system ATP-binding proteins **(B)**, antibiotic transport system permease proteins **(C)**, transketolase **(D)**, and protease **(E)**.

be the key reason for differences between their study and the current study. DNA is difficult to extract from *Actinobacteria* without lysozyme digestion because of its special cell wall (Zhang et al., 2013). Andongma et al. (2015) did not digest the sample with lysozyme, possibly causing the absence of *Actinobacteria* contributions. Moreover, they used pupae without puparium, which may explain the lack of *Actinobacteria* since bacteria of this order are often located in specific regions of the surface of insect hosts (Kaltenpoth, 2009). The greater sequencing depth in our study might also explain the greater number of identified OTUs since we obtained many more reads than Andongma reported. Actually other study has also identified *Actinobacteria* (Yong et al., 2017b).

Insects show remarkable adaptations to exploit diverse nutritional resources; these adaptations are due to the wide diversity of digestive enzymes produced by the insects themselves as well as the metabolic capabilities of symbiotic microorganisms that overcome the host's nutritional limitations (Berasategui et al., 2016). The high abundance of *Proteobacteria* observed in larvae and adults likely supports their importance in sugar metabolism. *Acetobacteraceae, Lactobacillaceae,* and *Enterobacteriaceae* were the most dominant families within this phylum that were observed in all larval and adult samples and have been reported to function in sugar metabolism (Kersters et al., 2006; Lambert et al., 2011; Yong et al., 2017a). However, we found that *Acetobacteraceae* and *Lactobacillaceae* were the most dominant families in larvae, while *Enterobacteriaceae* was the most dominant family in adults. This result may suggest that the digestion process may differ in larvae and adults, resulting in changes in the microbiota composition. Unlike the larvae, the adults must also feed on a high-protein diet, and protein in their diet can even influence the mating behavior of flies (Shelly and Kennelly, 2002; Shelly et al., 2005). An important family that is associated with most fruit flies is *Enterobacteriaceae*;

members of this group play very important roles in courtship and reproduction (Ben Ami et al., 2010). We also found that pathways involved in protein metabolism were significantly enriched in microbiotas in adults. These results indicate that microbiotas may also be involved in digesting protein in food. More evidence is needed to prove whether members of *Enterobacteriaceae* also play a role in protein digestion to influence the courtship and reproduction of *B. dorsalis*. Newell et al. (2014) also reported that some *Acetobacteraceae* species in the gut of *Drosophila* were involved in oxidative stress detoxification and encoded an efflux pump (Newell et al., 2014). Researches have suspected that *Lactobacillaceae* and *Enterobacteriaceae* contribute to digestion and protection against parasites and pathogens in insect gut (Billiet et al., 2017; Smith et al., unpublished). We thus need detailed investigation of the specific bacteria of *B. dorsalis* by pure culture methods.

Insect-associated microbes are just beginning to be exploited as promising sources of novel bioactive compounds (Dettner, 2011). Microbial symbionts providing chemical defense for the host against predators, parasites, parasitoids, and pathogens occur in several insect taxa, including beetles (Kellner, 2002), psyllids (Nakabachi et al., 2013), planthoppers (Fredenhagen et al., 1987), and solitary wasps (Kaltenpoth et al., 2005; Kaltenpoth, 2009; Kaltenpoth and Engl, 2014). The high abundance of *Actinobacteria* observed in the pupae of *B. dorsalis* support their importance in producing compounds with antimicrobial activity. *Actinobacteria* are known to be important sources suited as defensive symbionts of insects (Kaltenpoth, 2009). The number of OTUs in pupae that represent antibiotic transport system ATP-binding and antibiotic transport system permease proteins is significantly greater than that in larvae and adults, which strongly supports the defensive function of *Actinobacteria* in the pupae of *B. dorsalis*. *Actinobacteria* are particularly common and widespread in soil (Goodfellow and Williams, 1983) and are therefore regularly encountered by insects living in soil. The pupae of *B. dorsalis* consistently remained in soil until emergence. Kaltenpoth (2009) stated that three key factors probably contribute to the role of *Actinobacteria* as defensive exosymbionts in insects: (i) their ability to utilize a wide variety of carbon sources and to generally subsist on low levels of resources; (ii) the capacity of some taxa to form spores and thereby survive inhospitable conditions in the soil; and (iii) their ability to produce secondary metabolites with antibiotic properties (Goodfellow and Williams, 1983). The evolution of symbiotic interactions between *Actinobacteria* and insects might have been initiated by commensal or facultative entomopathogenic *Actinobacteria* that exploited the low amounts of compounds present on the cuticle or in the excretions of soil-dwelling insects. Once the bacteria became associated with an insect, antibiotic substances produced for the microbes' own protection might have also become beneficial for the host insect (Kaltenpoth, 2009). The specific inhabitation of *Intrasporangiaceae*, *Dermabacteraceae,* and *Brevibacteriaceae* in pupae in our study may indicate their defensive function. The defensive function of bacteria in *Brevibacteriaceae* has been previously revealed by pure culture methods, and researchers have identified the relevant bacteria and antibacterial compound (Ryser et al., 1994; Maisnierpatin and Richard, 1995). *Brachybacterium* of *Dermabacteraceae* were also identified to express strong antimicrobial activity (Liu et al., 2011). Pupae can therefore be used in future studies as a source from which *Actinomycetes* with antimicrobial activity can be isolated.

CONCLUSION

The larvae, pupae, and adults of *B. dorsalis* were observed to harbor distinct microbial flora. We performed a detailed investigation of the microbial flora of *B. dorsalis* that provides a basis for future research. Further studies to investigate the microbial composition may provide a comprehensive understanding of the differences in diet and physiological behavior among *B. dorsalis*. Moreover, host-specific microbial species, for example, those from the phylum *Actinobacteria*, can be used to develop potential compounds with antimicrobial activity that have potential value for human application.

AVAILABILITY OF DATA AND MATERIAL

Sequence data has been deposited at NCBI under Bioproject PRJNA415228.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: DC; analyzed the data: DC and YL; contributed reagents/materials/analysis tools: XfZ, XyZ, ZC, and ZW; wrote the paper: DC.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 31601693) and National Key Research and Development Project (No. 2016YFC1201200).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00114/full#supplementary-material

FIGURE S1 | Absolute content of symbionts in flies.

FIGURE S2 | Shannon rarefaction curves for all samples.

FIGURE S3 | Differences between the OTU numbers of the samples.

TABLE S1 | Summary of 16S rDNA read counts for all samples of *B. dorsalis*.

TABLE S2 | Comparison of OTU diversity of 16S rRNA gene amplicons.

DATA SHEET S1 | Abundance of the dominant OTUs.

DATA SHEET S2 | KEGG Orthology groups predicted for the samples.

REFERENCES

Aharon, Y., Pasternak, Z., Ben Yosef, M., Behar, A., Lauzon, C., Yuval, B., et al. (2013). Phylogenetic, metabolic, and taxonomic diversities shape mediterranean fruit fly microbiotas during ontogeny. *Appl. Environ. Microbiol.* 79, 303–313. doi: 10.1128/AEM.02761-12

Aksoy, E., Telleria, E. L., Echodu, R., Wu, Y., Okedi, L. M., Weiss, B. L., et al. (2014). Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. *Appl. Environ. Microbiol.* 80, 4301–4312. doi: 10.1128/AEM.00079-14

Andongma, A. A., Wan, L., Dong, Y. C., Li, P., Desneux, N., White, J. A., et al. (2015). Pyrosequencing reveals a shift in symbiotic bacteria populations across life stages of *Bactrocera dorsalis*. *Sci. Rep.* 5:9470. doi: 10.1038/srep09470

Antwis, R. E., Haworth, R. L., Engelmoer, D. J., Ogilvy, V., Fidgett, A. L., and Preziosi, R. F. (2014). *Ex situ* diet influences the bacterial community associated with the skin of red-eyed tree frogs (*Agalychnis callidryas*). *PLOS ONE* 9:e85563. doi: 10.1371/journal.pone.0085563

Augustinos, A. A., Kyritsis, G. A., Papadopoulos, N. T., Abd-Alla, A. M. M., Caceres, C., and Bourtzis, K. (2015). Exploitation of the medfly gut microbiota for the enhancement of sterile insect technique: use of *Enterobacter* sp. in larval diet-based probiotic applications. *PLOS ONE* 10:e0136459. doi: 10.1371/journal.pone.0136459

Ben Ami, E., Yuval, B., and Jurkevitch, E. (2010). Manipulation of the microbiota of mass-reared mediterranean fruit flies *Ceratitis capitata* (Diptera: Tephritidae) improves sterile male sexual performance. *ISME J.* 4, 28–37. doi: 10.1038/ismej.2009.82

Berasategui, A., Shukla, S., Salem, H., and Kaltenpoth, M. (2016). Potential applications of insect symbionts in biotechnology. *Appl. Microbiol. Biotechnol.* 100, 1567–1577. doi: 10.1007/s00253-015-7186-9

Billiet, A., Meeus, I., Van Nieuwerburgh, F., Deforce, D., Wackers, F., and Smagghe, G. (2017). Colony contact contributes to the diversity of gut bacteria in bumblebees (*Bombus terrestris*). *Insect. Sci.* 24, 270–277. doi: 10.1111/1744-7917.12284

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods.* 10, 57–59. doi: 10.1038/nmeth.2276

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods.* 7, 335–336. doi: 10.1038/nmeth.f.303

Cheng, D., Guo, Z., Riegler, M., Xi, Z., Liang, G., and Xu, Y. (2017). Gut symbiont enhances insecticide resistance in a significant pest, the oriental fruit fly *Bactrocera dorsalis* (Hendel). *Microbiome* 5:13. doi: 10.1186/s40168-017-0236-z

Dettner, K. (2011). "Potential pharmaceuticals from insects and their co-occurring microorganisms," in *Insect Biotechnology. Biologically-Inspired Systems*, Vol. 2, ed. A. Vilcinskas (Dordrecht: Springer).

Douglas, A. E. (2009). The microbial dimension in insect nutritional ecology. *Funct. Ecol.* 23, 38–47. doi: 10.1111/j.1365-2435.2008.01442.x

Douglas, A. E. (2015). Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* 60, 17–34. doi: 10.1146/annurev-ento-010814-020822

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Egert, M., Marhan, S., Wagner, B., Scheu, S., and Friedrich, M. W. (2004). Molecular profiling of 16S rRNA genes reveals diet-related differences of microbial communities in soil, gut, and casts of *Lumbricus terrestris* L. (Oligochaeta: Lumbricidae). *FEMS Microbiol. Ecol.* 48, 187–197. doi: 10.1016/j.femsec.2004.01.007

Florez, L. V., Biedermann, P. H., Engl, T., and Kaltenpoth, M. (2015). Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat. Prod. Rep.* 32, 904–936. doi: 10.1039/c5np00010f

Fredenhagen, A., Tamura, S. Y., Kenny, P. T. M., Komura, H., Naya, Y., Nakanishi, K., et al. (1987). Andrimid, a new peptide antibiotic produced by an intracellular bacterial symbiont isolated from a brown planthopper. *J. Am. Chem. Soc.* 109, 4409–4411. doi: 10.1021/ja00248a055

Goodfellow, M., and Williams, S. T. (1983). Ecology of *actinomycetes*. *Annu. Rev. Microbiol.* 37, 189–216. doi: 10.1146/annurev.mi.37.100183.001201

Gujjar, N. R., Govindan, S., Verghese, A., Subramaniam, S., and More, R. (2017). Diversity of the cultivable gut bacterial communities associated with the fruit flies *Bactrocera dorsalis* and *Bactrocera cucurbitae* (Diptera: Tephritidae). *Phytoparasitica* 45, 453–460. doi: 10.1007/s12600-017-0604-z

Hamby, K. A., and Becher, P. G. (2016). Current knowledge of interactions between *Drosophila suzukii* and microbes, and their potential utility for pest management. *J. Pest. Sci.* 89, 621–630. doi: 10.1007/s10340-016-0768-1

Hamden, H., Guerfali, M. M., Fadhl, S., Saidi, M., and Chevrier, C. (2013). Fitness improvement of mass-reared sterile males of *Ceratitis capitata* (Vienna 8 strain) (Diptera: Tephritidae) after gut enrichment with probiotics. *J. Econ. Entomol.* 106, 641–647. doi: 10.1603/EC12362

Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* 12, 1889–1898. doi: 10.1111/j.1462-2920.2010.02193.x

Kaltenpoth, M. (2009). *Actinobacteria* as mutualists: general healthcare for insects? *Trends Microbiol.* 17, 529–535. doi: 10.1016/j.tim.2009.09.006

Kaltenpoth, M., and Engl, T. (2014). Defensive microbial symbionts in Hymenoptera. *Funct. Ecol.* 28, 315–327. doi: 10.1111/1365-2435.12089

Kaltenpoth, M., Gottler, W., Herzner, G., and Strohm, E. (2005). Symbiotic bacteria protect wasp larvae from fungal infestation. *Curr. Biol.* 15, 475–479. doi: 10.1016/j.cub.2004.12.084

Kellner, R. L. (2002). Molecular identification of an endosymbiotic bacterium associated with pederin biosynthesis in *Paederus sabaeus* (Coleoptera: Staphylinidae). *Insect. Biochem. Mol. Biol.* 32, 389–395. doi: 10.1016/S0965-1748(01)00115-1

Kersters, K., Lisdiyanti, P., Komagata, K., and Swings, J. (2006). "The family *Acetobacteraceae*: the Genera *Acetobacter*, *Acidomonas*, *Asaia*, *Gluconacetobacter*, *Gluconobacter*, and *Kozakia*," in *The Prokaryotes*, eds M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (New York, NY: Springer).

Khaeso, K., Andongma, A. A., Akami, M., Souliyanonh, B., Zhu, J., Krutmuang, P., et al. (2017). Assessing the effects of gut bacteria manipulation on the development of the oriental fruit fly, *Bactrocera dorsalis* (Diptera; Tephritidae). *Symbiosis* 74, 97–105. doi: 10.1007/s13199-017-0493-4

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., et al. (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J.* 6, 1378–1390. doi: 10.1038/ismej.2011.192

Kueneman, J. G., Parfrey, L. W., Woodhams, D. C., Archer, H. M., Knight, R., and McKenzie, V. J. (2014). The amphibian skin-associated microbiome across species, space and life history stages. *Mol. Ecol.* 23, 1238–1250. doi: 10.1111/mec.12510

Lambert, M., Surhone, M. T., and Tennoe, S. F. (2011). *Lactobacillaceae*. Missoula, MT: Betascript Publishing.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Lindow, S. E., and Brandl, M. T. (2003). Microbiology of the phyllosphere. *Appl. Environ. Microbiol.* 69, 1875–1883. doi: 10.1128/AEM.69.4.1875-1883.2003

Liu, Z. X., Huang, K., Xiao, H. D., Chen, Q. H., He, J. W., and Chen, Y. G. (2011). Screening and preliminary identification of the antimicrobial strains associated with *Anthopleura xanthogrammica*. *Chin. J. Antibiot.* 36, 416–420.

Loudon, A. H., Woodhams, D. C., Parfrey, L. W., Archer, H., Knight, R., McKenzie, V., et al. (2014). Microbial community dynamics and effect of environmental microbial reservoirs on red-backed salamanders (*Plethodon cinereus*). *ISME J.* 8, 830–840. doi: 10.1038/ismej.2013.200

Maisnierpatin, S., and Richard, J. (1995). Activity and purification of linenscin OC2, an antibacterial substance produced by *Brevibacterium* linens OC2, an orange cheese coryneform bacterium. *Appl. Environ. Microbiol.* 61, 1847–1852.

Michael, E., Shay, S., Orit, S., Yulia, G., Yael, H., and Yoav, S. (2016). Impact of gut microbiota on the fly's germ line. *Nat. Commun.* 7:11280. doi: 10.1038/ncomms11280

Nakabachi, A., Ueoka, R., Oshima, K., Teta, R., Mangoni, A., Gurgui, M., et al. (2013). Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol.* 23, 1478–1484. doi: 10.1016/j.cub.2013.06.027

Navas-Molina, J. A., Peralta-Sanchez, J. M., Gonzalez, A., McMurdie, P. J., Vazquez-Baeza, Y., Xu, Z. J., et al. (2013). Advancing our understanding

of the human microbiome using QIIME. *Methods Enzymol.* 531, 371–444. doi: 10.1016/B978-0-12-407863-5.00019-8

Newell, P. D., Chaston, J. M., Wang, Y., Winans, N. J., Sannino, D. R., Wong, A. C., et al. (2014). *In vivo* function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. *Front. Microbiol.* 5:576. doi: 10.3389/fmicb.2014.00576

Ohkuma, M., and Brune, A. (2011). "Diversity, structure, and evolution of the termite gut microbial community," in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and N. Lo (Dordrecht: Springer), 413–438.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2015). *vegan: Community Ecology Package. R Package Version 2.3–1*.

Rafael, J. V., Natalia, G. J., Gloria, C.-R., Sandra, I. U. S., and Claudia Ximena, M. H. (2016). Structural differences in gut bacteria communities in developmental stages of natural populations of *Lutzomyia evansi* from Colombia's Caribbean coast. *Parasit. Vectors* 9:496. doi: 10.1186/s13071-016-1766-0

Ramadhar, T. R., Beemelmanns, C., Currie, C. R., and Clardy, J. (2014). Bacterial symbionts in agricultural systems provide a strategic source for antibiotic discovery. *J. Antibiot.* 67, 53–58. doi: 10.1038/ja.2013.77

Rebollar, E. A., Hughey, M. C., Medina, D., Harris, R. N., Ibanez, R., and Belden, L. K. (2016). Skin bacterial diversity of Panamanian frogs is associated with host susceptibility and presence of *Batrachochytrium dendrobatidis*. *ISME J.* 10, 1682–1695. doi: 10.1038/ismej.2015.234

Rio, R. V., Wu, Y. N., Filardo, G., and Aksoy, S. (2006). Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proc. Biol. Sci.* 273, 805–814. doi: 10.1098/rspb.2005.3399

Ryser, E. T., Maisnier-Patin, S., Gratadoux, J. J., and Richard, J. (1994). Isolation and identification of cheese-smear bacteria inhibitory to *Listeria* spp. *Int. J. Food. Microbiol.* 21, 237–246. doi: 10.1016/0168-1605(94)90030-2

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Sharon, G., Segal, D., Ringo, J. M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010). Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20051–20056. doi: 10.1073/pnas.1009906107

Shelly, T. E., Edu, J., and Pahio, E. (2005). Influence of diet and methyl eugenol on the mating success of males of the oriental fruit fly, *Bactrocera dorsalis* (Diptera : Tephritidae). *Fla. Entomol.* 88, 307–313. doi: 10.1653/0015-4040(2005)088[0307:IODAME]2.0.CO;2

Shelly, T. E., and Kennelly, S. (2002). Influence of male diet on male mating success and longevity and female remating in the Mediterranean fruit fly (Diptera : Tephritidae) under laboratory conditions. *Fla. Entomol.* 85, 572–579. doi: 10.1653/0015-4040(2002)085[0572:Iomdom]2.0.Co;2

Teixeira, L., Ferreira, A., and Ashburner, M. (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLOS Biol.* 6:e2. doi: 10.1371/journal.pbio.1000002

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraserliggett, C., Knight, R., and Gordon, J. I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* 449, 804–810. doi: 10.1038/nature06244

Vänninen, I., Tynijuslin, J., and Hokkanen, H. (2000). Persistence of augmented *Metarhizium anisopliae* and *Beauveria bassiana* in Finnish agricultural soils. *Biocontrol* 45, 201–222. doi: 10.1023/A:1009998919531

Wang, H., Jin, L., and Zhang, H. (2011). Comparison of the diversity of the bacterial communities in the intestinal tract of adult *Bactrocera dorsalis* from three different populations. *J. Appl. Microbiol.* 110, 1390–1401. doi: 10.1111/j.1365-2672.2011.05001.x

Wang, H. X., Jin, L., Peng, T., Zhang, H. Y., Chen, Q. L., and Hua, Y. J. (2014). Identification of cultivable bacteria in the intestinal tract of *Bactrocera dorsalis* from three different populations and determination of their attractive potential. *Pest. Manag. Sci.* 70, 80–87. doi: 10.1002/ps.3528

Ye, L., Amberg, J., Chapman, D., Gaikowski, M., and Liu, W. T. (2014). Fish gut microbiota analysis differentiates physiology and behavior of invasive Asian carp and indigenous American fish. *ISME J.* 8, 541–551. doi: 10.1038/ismej.2013.181

Yong, H. S., Song, S. L., Chua, K. O., and Lim, P. E. (2017a). High diversity of bacterial communities in developmental stages of *Bactrocera carambolae* (Insecta: Tephritidae) revealed by Illumina MiSeq sequencing of 16S rRNA gene. *Curr. Microbiol.* 74, 1076–1082. doi: 10.1007/s00284-017-1287-x

Yong, H. S., Song, S. L., Chua, K. O., and Lim, P. E. (2017b). Microbiota associated with *Bactrocera carambolae* and B. dorsalis (Insecta: Tephritidae) revealed by next-generation sequencing of 16S rRNA gene. *Meta Gene* 11, 189–196. doi: 10.1016/j.mgene.2016.10.009

Zhang, J., Xia, Z., He, J., Sun, H., and Zhang, L. (2013). Evaluation outcome of *actinobacteria* diversity in saline environment influenced by different DNA extraction methods. *Wei Sheng Wu Xue Bao* 53, 746–757.

Zhao, Y., Wang, W. Q., Zhu, F., Wang, X. Y., Wang, X. P., and Lei, C. L. (2017). The gut microbiota in larvae of the housefly *Musca domestica* and their horizontal transfer through feeding. *AMB Express* 7:147. doi: 10.1186/S13568-017-0445-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhao, Zhang, Chen, Wang, Lu and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 January 2018
doi: 10.3389/fmicb.2017.02622

Functional Characteristics of the Flying Squirrel's Cecal Microbiota under a Leaf-Based Diet, Based on Multiple Meta-Omic Profiling

Hsiao-Pei Lu[1†‡], Po-Yu Liu[1,2‡], Yu-bin Wang[1,3], Ji-Fan Hsieh[1†], Han-Chen Ho[4], Shiao-Wei Huang[1], Chung-Yen Lin[3], Chih-hao Hsieh[1,5,6,7] and Hon-Tsen Yu[1,2]*

[1] Department of Life Science, National Taiwan University, Taipei, Taiwan, [2] Genome and Systems Biology Degree Program, National Taiwan University & Academia Sinica, Taipei, Taiwan, [3] Institute of Information Science, Academia Sinica, Taipei, Taiwan, [4] Department of Anatomy, Tzu Chi University, Hualien, Taiwan, [5] Institute of Oceanography, National Taiwan University, Taipei, Taiwan, [6] Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan, [7] National Center for Theoretical Sciences, Taipei, Taiwan

Mammalian herbivores rely on microbial activities in an expanded gut chamber to convert plant biomass into absorbable nutrients. Distinct from ruminants, small herbivores typically have a simple stomach but an enlarged cecum to harbor symbiotic microbes; however, knowledge of this specialized gut structure and characteristics of its microbial contents is limited. Here, we used leaf-eating flying squirrels as a model to explore functional characteristics of the cecal microbiota adapted to a high-fiber, toxin-rich diet. Specifically, environmental conditions across gut regions were evaluated by measuring mass, pH, feed particle size, and metabolomes. Then, parallel metagenomes and metatranscriptomes were used to detect microbial functions corresponding to the cecal environment. Based on metabolomic profiles, >600 phytochemical compounds were detected, although many were present only in the foregut and probably degraded or transformed by gut microbes in the hindgut. Based on metagenomic (DNA) and metatranscriptomic (RNA) profiles, taxonomic compositions of the cecal microbiota were dominated by bacteria of the Firmicutes taxa; they contained major gene functions related to degradation and fermentation of leaf-derived compounds. Based on functional compositions, genes related to multidrug exporters were rich in microbial genomes, whereas genes involved in nutrient importers were rich in microbial transcriptomes. In addition, genes encoding chemotaxis-associated components and glycoside hydrolases specific for plant beta-glycosidic linkages were abundant in both DNA and RNA. This exploratory study provides findings which may help to form molecular-based hypotheses regarding functional contributions of symbiotic gut microbiota in small herbivores with folivorous dietary habits.

Keywords: animal-microbe interaction, ecological adaptation, gut microbiota, metabolomics, metagenomics, metatranscriptomics

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council of Argentina (CONICET), Argentina

Reviewed by:
Henning Seedorf, Temasek Life Sciences Laboratory, Singapore
Alfonso Benítez-Páez, Instituto de Agroquímica y Tecnología de Alimentos (CSIC), Spain

****Correspondence:***
Hon-Tsen Yu
ayu@ntu.edu.tw

†Present Address:
Hsiao-Pei Lu, Institute of Oceanography, National Taiwan University, Taipei, Taiwan
Ji-Fan Hsieh, Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

‡These authors have contributed equally to this work and co-first authors.

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *18 September 2017*
Accepted: *15 December 2017*
Published: *04 January 2018*

Citation:
Lu H-P, Liu P-Y, Wang Y, Hsieh J-F, Ho H-C, Huang S-W, Lin C-Y, Hsieh C and Yu H-T (2018) Functional Characteristics of the Flying Squirrel's Cecal Microbiota under a Leaf-Based Diet, Based on Multiple Meta-Omic Profiling. Front. Microbiol. 8:2622. doi: 10.3389/fmicb.2017.02622

INTRODUCTION

Mammals and their gut microbiota have co-evolved for millions of years, forming an interdependent, symbiotic relationship (Stevens and Hume, 1998; Ley et al., 2008; Leser and Molbak, 2009). Establishing a cooperative association is particularly crucial for mammalian herbivores, as they heavily rely on gut microbiota to convert plant biomass into absorbable nutrients (Wallace, 1992; Kamra, 2005; Deusch et al., 2017). To provide space for gut microbiota, mammalian herbivores typically have a complex digestive tract with an enlarged compartment (Hume, 1989). For example, ruminants (so-called "foregut fermenters"), have specialized stomach chambers to house microbes for hydrolyzing and fermenting plant fibers. In contrast, other herbivores have an enlarged chamber in the large intestine ("hindgut fermenters") to store ingesta for microbial activities (Mackie, 2002). Specifically, small mammalian herbivores usually have a well-developed cecum with a capacity ∼10 times that of their stomach (Manning et al., 1994; Campbell et al., 2000). More interestingly, their gut structure apparently has a special sorting mechanism at the ileal-cecal-colic junction, permitting fluid and fine particles (including microbes and fine plant debris) to be retained in the cecum, while concurrently allowing coarse dry matter to rapidly pass through the gut (Hume, 2002). Such digestive strategies are believed to satisfy energy demands of small mammalian herbivores with high mass-specific metabolic rates (Sakaguchi, 2003). However, in contrast to numerous studies on ruminants and ruminal microbiota (Brulc et al., 2009; Weimer, 2015; Mao et al., 2016; Deusch et al., 2017), much less is known about co-adaptation characteristics between small mammalian herbivores and their cecal microbiota.

In this study, the white-faced flying squirrel (*Petaurista alborufus lena*) inhabiting montane areas of Taiwan (Oshida et al., 2011) was selected as a target organism. We focused on a wild species, instead of domestic animals, as wild herbivores usually have a great challenge to gain sufficient energy from coarse plant material. Consequently, studies on their gut microbiota were expected to advance knowledge regarding microbial functional characteristics under specific host dietary preferences in natural habitats (Hird, 2017). Compared to other palatable plant-based dietary choices (such as seeds, fruits, and flowers), this flying squirrel species is of special consideration, because it is an arboreal obligate folivore, mainly feeding on leaf parts (including buds, petioles, young leaves, and mature leaves) of various broadleaf trees (Lee et al., 1986; Kuo and Lee, 2003). It occupies a unique feeding niche in treetops, escaping pressures of both competition and predation in forest floor ecosystems (Coley and Barone, 1996). However, since tree leaves often contain complex carbohydrates and secondary metabolites (i.e., plant defensive chemicals, such as flavonoids, alkaloids, and tannins; Mithofer and Boland, 2012), folivorous mammals must have an efficient digestive system to gather nutrients and concurrently avoid toxins from their leaf-based diets (Coley and Barone, 1996). These dual challenges may be overcome with multitudinous functions provided by gut microbiota (Kohl et al., 2014, 2016), which probably confer rapid dietary adaptation (Alberdi et al., 2016). It is noteworthy that this flying squirrel species has a large population in wide montane regions of Taiwan and East Asia (Smith and Johnston, 2016), which implies successful feeding strategies, including effective gut microbiota. Thus, the flying squirrel's cecal microbiota should be an ideal model to investigate microbial functional characteristics with regard to a high-fiber, toxin-rich diet.

Objectives were to elucidate anatomical/physiological characteristics of the flying squirrel's digestive system and functional characteristics of its cecal microbiota under a high-fiber and toxin-rich diet. In our previous study using 16S rRNA gene libraries to investigate spatial heterogeneity of the flying squirrel's gut microbiota (Lu et al., 2014), we reported that the cecum, an enlarged part of the gut, contained relatively high bacterial diversity. Here, we further investigated this specialized system, using multiple "meta-omic" approaches, including metabolomics, metagenomics and metatranscriptomics (Segata et al., 2013; Aguiar-Pulido et al., 2016). Although each meta-omic approach has been widely used alone, few studies have integrated multiple meta-omic data to understand animal-microbe interactions. In this study, functional features of the cecal microbiota were profiled with three meta-omic approaches; this enabled complementary confirmation of microbial activities (Segata et al., 2013; Aguiar-Pulido et al., 2016). Specifically, we used metabolomes to trace the metabolic fate of dietary compounds along the digestive tract, and also used parallel metagenomes and metatranscriptomes to detect discrepancies between existing genes (DNA level; potential functions) and expressed genes (RNA level; realized functions) of the cecal microbiota. These meta-omic data were also compared to open-access gut metagenomes from other mammals to reveal unique characteristics of the flying squirrel's cecal microbiota. Moreover, to provide a general understanding of this system, we measured mass, pH, and feed particle size throughout the gut, and included images of flying squirrel's skull and teeth, as well as microscopic features of cecal microorganisms (**Figure 1**). This exploratory study provides findings which may help to form molecular-based hypotheses regarding functional contributions of symbiotic gut microbiota in small herbivores with folivorous dietary habits.

MATERIALS AND METHODS

Sample Collection

Five white-faced flying squirrels (*Petaurista alborufus lena*) were captured from the mountains of Taiwan. A wild animal collecting permit (No. 0990007029) was approved by Yushan National Park Headquarters, Taiwan. Animals were dissected immediately after death. All experiments were performed in accordance with the Wildlife Conservation Act (http://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=M0120001). All five individuals were used for measurements of gut structure and feed contents, whereas three were used for the metabolomes and the remaining two were used for metagenomes/metatranscriptomes.

Gastrointestinal Anatomy and Physiology

The length and weight (with feed contents included) of the stomach, small intestine, cecum, and large intestine (i.e., four

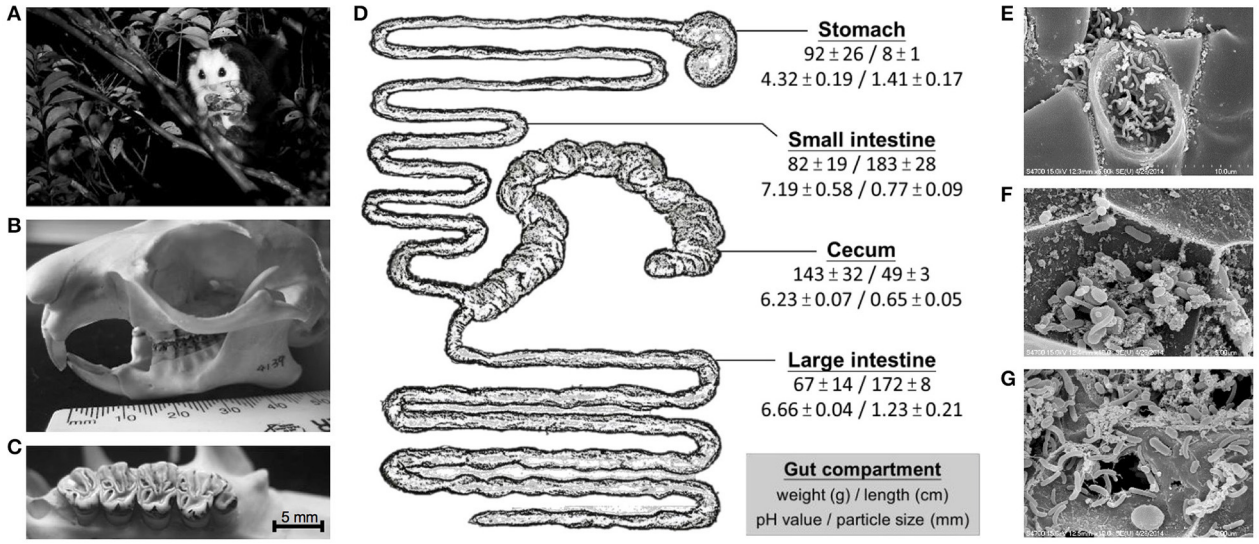


FIGURE 1 | Digestive strategies of the leaf-eating flying squirrel. **(A)** White-faced flying squirrel (*Petaurista alborufus lena*) occupies a unique feeding niche in the treetop. **(B)** The skull structure (mouth with a broad gap) enables temporarily holding a large volume of tree leaves before chewing them. **(C)** Molars with multiple narrow folds on the crown enable tree leaves to be well chewed. **(D)** Anatomical/physiological characteristics (mean ± SD) of the four main gut compartments; note the extremely enlarged cecum which contained the majority of feed contents for microbial fermentation. **(E–G)** The cecum harbored large numbers of microorganisms that act on plant debris. Photo credits: **(A)** Hsueh-Chen Chen; **(B,C)** Ji-Fan Hsieh; and **(E–G)** Han-Chen Ho.

main gut compartments) were measured and reported (Lu et al., 2012). Here, we additionally examined the pH of feed contents. Moreover, to understand how plant-based diets were processed along the digestive tract, feed contents (∼10 g from each gut compartment) were passed through the graded sieves (pore sizes: 4, 2, 1, 0.5, 0.25, and 0.125 mm) by wet-sieving (Clauss et al., 2002). Thereafter, feed particles of each size fraction (seven groups: > 4.0, 2–4, 1–2, 0.5–1, 0.25–0.5, 0.25–0.125, and <0.125 mm) were transferred onto a Petri dish, dried at 60°C for 24 h, and weighed after cooling to room temperature. Mean particle size of feed contents from each gut compartment was calculated by fitting a normal distribution (Fritz et al., 2009). Based on these measurements, we sketched the shape of the flying squirrel's digestive tract and marked physiologic characteristics of each gut compartment (**Figure 1**). In addition, for a visual image of gut microbes, feed contents from the cecum were fixed and observed with a Hitachi S-4700 field emission scanning electron microscope (FE-SEM). Specifically, specimens were pre-fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer at 4°C, post-fixed with 1% osmium tetroxide in 0.1 M cacodylate buffer at room temperature, then dehydrated through a graded series of ethanol until they were in 100% ethanol, and then they were put into 100% acetone. Specimens were critical point dried, sputter coated with gold, and examined with a FE-SEM at 15 kV.

Construction of Metabolomes

For constructing metabolomes, feed contents from three individuals were frozen (−20°C), immediately after death. For each individual, a total of 15 samples from the stomach, small intestine, cecum, and large intestine (2, 5, 4, and 4, respectively; each ∼0.5 g) were collected (Supplementary Figure 1). Each sample was homogenized with 3-fold distilled water for 5 min. The supernatant containing metabolites was retrieved by centrifugation (10,000 × g for 10 min) and mixed with 100% methanol (Sigma-Aldrich, CHROMASOLV®, for HPLC) at a ratio of 1:3, and repeated once. The supernatant was lyophilized and dissolved in 100 μl distilled water for metabolomic detection by liquid chromatography (LC)—electrospray ionization (ESI)—mass spectrometry (MS), composed of an ultra-performance LC (Ultimate 3000 RSLC, Dionex) coupled with an ESI source of quadrupled time-of-flight MS (maXis HUR-QToF system, Bruker Daltonics). Metabolites were separated by reversed-phase liquid chromatography on an HSS T3 C18 column (2.1 × 100 mm; Walters). The LC parameters were autosampler temperature at 4°C and injection volume of 10 μl with flow rate at 0.4 ml/min. Elution started from 99% mobile phase A (0.1% formic acid in pure water) and 1% mobile phase B (0.1% formic acid in ACN). Mobile phase B was held at 1% for 0.5 min, raised to 60% in 6 min, raised to 90% in 0.5 min and held for 1.5 min, and lowered to 1% in 5 min. Finally, the column was equilibrated by pumping 99% mobile phase A for 4 min. The LC–ESI–MS chromatograms were acquired by the following parameters: 190°C dry gas at 8 L/min flow rate, and 1.4 bar nebulizer gas with 4,500/3,500 capillary voltage for positive/negative ion modes. The m/z values in mass spectra were recorded for further data processing.

Metabolomic Profiles across Gut Compartments

Metabolomic data acquired from the LC–ESI–MS chromatograms were processed by TargetAnalysis (Version 1.1,

Bruker Daltonics) and XCMS (Smith et al., 2006) with optimized parameters for Bruker Q-TOF mass spectrometer (Tautenhahn et al., 2008). Metabolites were identified by MetaboSearch (Zhou et al., 2012) and matched with theoretical m/z values against the Madison Metabolomics Consortium Database (Cui et al., 2008). Identified metabolites were further annotated according to the reference library of KEGG compounds (Hattori et al., 2003), with the tolerance of LC peaks within 0.3 min and signal intensities >1,000 counts. These compound annotations were used for profiling chemical environments of each gut compartment.

The overall similarity/dissimilarity among metabolomic profiles was assessed with Jaccard distance and displayed in an ordination diagram using principal coordinates analysis (PCoA), generated by the QIIME (Version 1.9) pipeline (Caporaso et al., 2010). Detected KEGG compounds were further grouped according to their chemical structures and cellular functions (Hattori et al., 2003), with an emphasis on the quantity and signal intensity of phytochemicals (leaf-derived compounds) across gut compartments. Multiple comparisons were done by ANOVA coupling with Scheffé's test, using "stats" and "agricolae" packages (Mendiburu, 2016) in R (R Development Core Team, 2016).

Finally, metabolomic data were compiled with metagenomic and metatranscriptomic data (described below) to detect potentially crucial functional reactions of cecal microbiota under the leaf-based diet.

Construction of Metagenomes and Metatranscriptomes

For constructing metagenomes and metatranscriptomes, feed contents from two individuals were collected immediately after death and placed in RNAlater solution (Ambion, Life Technologies). For each individual, samples from the cecum (Supplementary Figure 1, sites 8–11, each ∼5 g) were pooled to represent the entire cecal microbiota. Total DNA and RNA were isolated using the AllPrep DNA/RNA Mini Kit (QIAGEN), according to manufacturer's instructions. Briefly, cecal contents were centrifuged (10,000 × g for 10 min) to remove RNAlater and re-suspended in lysis buffer (TE buffer with 5 mg/mL lysozyme) at room temperature for 5 min. After addition of RLT buffer (with beta-mercaptoethanol), lysate was homogenized by passing it 10 times through a 20-G needle with a 1-ml syringe. The DNA molecules were purified through an AllPrep DNA spin column, and RNA molecules (>200 bp, without small rRNAs and tRNAs) were isolated with an RNeasy spin column with DNase treatment (QIAGEN). Furthermore, 16S rRNAs and 23S rRNAs (typically accounting for >80% of total RNA molecules) were removed using a MICROB*Express*™ Bacterial mRNA Enrichment Kit (Ambion, Life Technologies), according to the manufacturer's instructions. Quantity and quality of DNA and RNA samples were estimated using a NanoDrop 2000 Spectrophotometer (Thermo Scientific) and an Agilent 2100 bioanalyzer (Agilent Technologies).

Coupled DNA and RNA samples were subjected to shotgun sequencing on a GS-FLX Titanium system (Roche Life Science) for characterizing metagenomes and metatranscriptomes. For RNA samples, double-stranded cDNA libraries were constructed according to the cDNA Synthesis System Kit (Roche Life Science). The DNA and cDNA libraries were converted into single-stranded DNA fragments for sequencing using a GS-FLX Titanium Rapid Library Preparation Kit (Roche Life Science). Two full runs of shotgun sequencing were conducted. Sequence data were submitted to the NCBI Sequence Read Archive (SRA; BioProject accession PRJNA267179).

Metagenomic and Metatranscriptomic Profiles of Cecal Microbiota

Raw reads of metagenome (DNA-based) and metatranscriptome (RNA-based) sequences were filtered to remove exact duplicates (possible artifacts from emulsion PCR) using CD-hits (Li and Godzik, 2006), and low-quality parts (Phred quality < 25, N content > 3%, sequences < 100 bp) were trimmed using SeqClean (http://seqclean.sourceforge.net). After quality control, a total of 569,349 metagenome reads and 483,241 metatranscriptome reads (both with average length > 300 bp) were used for bioinformatic analyses. Potential rRNA sequences were identified using BLASTn (cut-off e-value < 1e-5, with > 90% alignment) against the SILVA (containing eukaryotic and prokaryotic ribosomal RNA sequences) database (Pruesse et al., 2007). Remaining non-rRNA sequences were considered as mRNA reads for protein-coding gene annotation. Detailed sequence statistical information is shown in Supplementary Table 1.

To provide taxonomic profiles of metagenomes and metatranscriptomes, putative mRNA reads were mapped to NCBI-nr (non-redundant protein database of National Center for Biotechnology Information) database (Sayers et al., 2012) using BLASTX (cut-off e-value < 1e-5) to identify potential taxonomic origins. Taxonomic classification of sequences was retrieved and summarized according to NCBI Taxonomy. For domain—and phylum-level taxonomic profiles, all significant hits (cut-off e-value < 1e-5) were considered. For fine-level taxonomic identification, minimum identity thresholds of amino acid sequences were further restricted at 45 and 65% for family- and genus-level taxonomic groups, respectively (Konstantinidis et al., 2017).

To provide functional profiles of metagenomes and metatranscriptomes, FragGeneScan (Rho et al., 2010) based on the hidden Markov model was used to predict open reading frames (ORFs) on putative mRNA reads. The ORFs were queried to identify conserved protein families and domains with amino acid sequences against: (1) COG (Clusters of Orthologous Groups of proteins) database (Galperin et al., 2015); (2) Pfam (Protein families) database (Finn et al., 2016); and (3) KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto, 2000). Specifically, COG annotation was performed using a reverse position-specific BLAST algorithm (RPS-BLAST, cut-off e-value < 1e-5) against NCBI COG database (latest update 2017/3/28). Pfam annotation was performed using a HMMER3 (cut-off e-value < 1e-5) against Pfam database (version 31.0). In addition, to focus on carbohydrate-degrading enzymes, glycoside hydrolases (GHs) were determined according to correspondence between Pfam

and CAZy (Carbohydrate-Active enZYmes; Cantarel et al., 2009). KEGG Orthology (KO) annotation was performed using KEGG GhostKOALA (Kanehisa et al., 2016), which is designed for metagenome sequences. Moreover, to reconstruct potentially crucial functional reactions conducted by the cecal microbiota, compounds, and genes detected in the metabolome, metagenome, or metatranscriptome were all mapped to KEGG modules (Takami et al., 2012) and pathways (Hattori et al., 2003).

To evaluate individual variation, Pearson correlation coefficients (r) were used to assess degree of similarity between results derived from two flying squirrel individuals, using the "stats" package of R (R Development Core Team, 2016).

Meta-Analysis of Mammalian Gut Metagenomes

To provide an overview of functional characteristics of mammalian gut microbiota, in addition to 2 metagenome datasets from the flying squirrel's cecum, 4 metagenome datasets from cow's rumen (Brulc et al., 2009), and 39 metagenome datasets from fecal samples of zoo mammals (Muegge et al., 2011) were downloaded and analyzed. All datasets were processed with the same procedures for the ORF prediction and functional annotation, as mentioned above. Overall similarity/dissimilarity of mammalian gut metagenomes was assessed with Bray-Curtis dissimilarity and displayed in an ordination diagram using nonmetric multidimensional scaling (NMDS). An ANOSIM was done to determine whether there were significant differences in functional profiles (based on COG or Pfam) among five metagenome groups (i.e., flying squirrel's cecum, cow's rumen, zoo carnivores' feces, zoo omnivores' feces, and zoo herbivores' feces). Moreover, we performed a Kruskal-Wallis test with a threshold of adjusted p-value (FDR) <0.05 to assess significant gene families and detect different relative abundances among the five metagenome groups. Then, Dunn's test was used as a post-hoc procedure for pairwise multiple comparisons to detect the uniqueness of the flying squirrel's cecal metagenome, compared to all other gut metagenomes. Those statistical assessments were conducted using "vegan" (Oksanen et al., 2017), "RVAideMemoire" (Hervé, 2017), and "stats" packages in R (R Development Core Team, 2016).

RESULTS

Anatomical/Physiological Characteristics of the Flying Squirrel's Digestive System

The leaf-eating flying squirrel (**Figure 1A**) had a skull and molars (**Figures 1B,C**) that were specialized, facilitating mastication of tree leaves. In addition, shape and size both varied substantially across four main gut compartments, with the cecum being the largest chamber (**Figure 1D**). The pH of feed contents also differed among gut chambers (**Figure 1D**), with stomach contents having a relatively low pH (4-5), small intestine contents being slightly alkaline (pH 7-8), and contents in the cecum and large intestine being mildly acidic (both pH 6-7). Regarding feed particle sizes (**Figure 1D** and Supplementary Figure 2), feed particles in the stomach were already fine (on average, ∼1.5 mm). Nevertheless, compared to dry and coarse fibers present in the stomach, feed contents in the cecum were even finer (on average, ∼0.6 mm) with a sludge-like texture (Supplementary Figure 3). Moreover, in microscopic evaluation of cecal contents, various microbial cells were closely attached to plant debris (**Figures 1E–G**).

Metabolomic Compositions across the Flying Squirrel's Gut Compartments

Based on LC–ESI–MS metabolomic profiles, each gut compartment contained its own unique compound composition in feed contents, with clear differences between samples from the upper (stomach and small intestine) vs. lower (cecum and large intestine) portions of the digestive tract (**Figure 2**). Overall, there were ∼600 phytochemicals (mainly leaf-derived metabolites) detected among all gut samples. Several phytochemicals (∼20%) were present only in the stomach and small intestine, with relatively fewer phytochemicals detected in the cecum and large intestine (**Figure 3**), especially flavonoids, phenylpropanoids, and polyketides (**Figures 3D–F**). Moreover, considering the relative intensity of those detected phytochemicals, their signal intensities were also much lower in the hindgut than in the foregut (Supplementary Figure 4).

Metagenomes and Metatranscriptomes of the Flying Squirrel's Cecal Microbiota

Based on the annotation of protein-coding sequences, at both DNA and RNA levels, cecal microbiota were dominated by bacteria (average, 97.94% of DNA reads and 88.24% of RNA reads; Supplementary Table 2), followed by eukaryotes

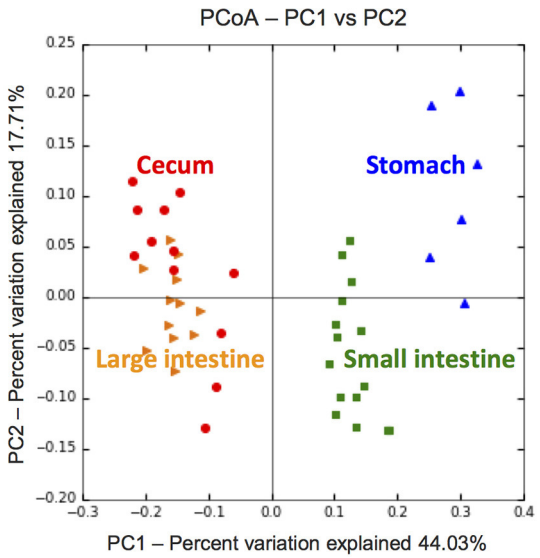


FIGURE 2 | PCoA plot of metabolomic compositions revealing distinct chemical environments across gut compartments of the flying squirrel ($N = 3$). For each individual, 2, 5, 4, and 4 metabolomic samples were collected from the stomach, small intestine, cecum, and large intestine, respectively (see Figure S1 for details).

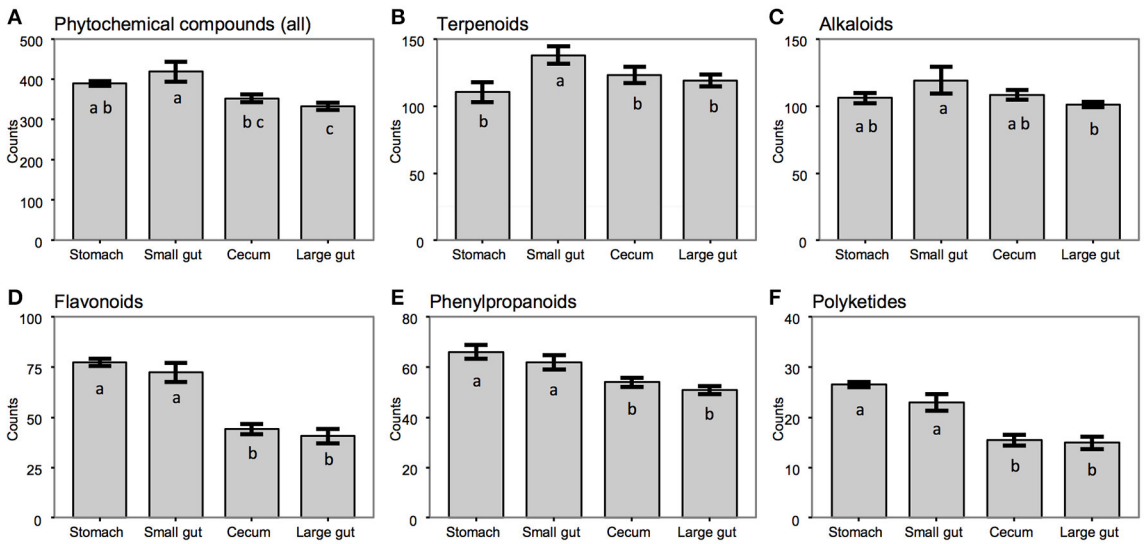


FIGURE 3 | Numbers of phytochemical compounds determined across gut compartments of the flying squirrel ($N = 3$), for **(A)** entire group of phytochemicals or **(B–F)** each subgroup. [a–c]Columns without a common letter differed ($P < 0.05$); error bars are standard error of the mean (SEM) for number of compounds detected within each gut region.

(0.77 and 11.37%), archaea (0.46 and 0.11%), and viruses (0.09 and 0.03%). Targeting bacterial composition, the phylum Firmicutes was extremely abundant in both metagenomes (∼90%) and metatranscriptomes (∼85%), with Actinobacteria, Proteobacteria, Bacteroidetes constituting only a small fraction (Supplementary Table 3). At a genus level, 577 bacterial taxa were detected, whereas only 30 taxa contained >0.5% reads in any library (**Figure 4**). Among these abundant genera, *Clostridium*, *Ruminococcus*, and *Eubacterium* (all belonging to Firmicutes) were the top three in all libraries, contributing to the majority (up to 50%) of annotated protein-coding genes in the cecal metagenome and metatranscriptome.

According to the COG annotation (**Figure 5** and Supplementary Figure 5), functional gene families had distinct abundance patterns at DNA and RNA levels. Specifically, the most abundant COG in the cecal metagenome was assigned to the ABC-type multidrug exporter component (COG1132; involved in defense mechanisms; representing ∼2% of hits; **Figure 5**). By contrast, the most abundant COG in the cecal metatranscriptome was assigned to the flagellin protein (COG1344; involved in cell motility), followed by several ABC-type sugar importer components (e.g., COG3839; involved in carbohydrate transport and metabolism). In addition, ABC-type importer components for short peptides (e.g., COG0747, involved in amino acid transport and metabolism) also had relatively high abundances at the RNA level.

Contrasting functional gene profiles in the metatranscriptome vs. metagenome were also revealed by KEGG annotation (**Figure 6** and Supplementary Figure 6), with distinct types of membrane transport components dominant at RNA—vs. DNA-levels, similar to COG results (**Figures 5**, **6**). In combination with metabolomic data, we identified KEGG pathways with high coverage of KOs and relevant compounds, including pathways involved in substrate-induced cell motility (i.e., bacterial chemotaxis and flagellar assembly; Supplementary Figures 7, 8) and pathways for biosynthesis of cellular components from plant-sourced nutrients (e.g., cellobiose, xylose, and arabinose were transported into bacterial cells and converted into other macromolecules, such as amino acids, nucleotides, and peptidoglycans; Supplementary Figures 9–16).

Among those abundantly detected COGs and KOs (**Figures 5**, **6**), it was notable that apart from ABC-type transport systems and genes associated with "housekeeping functions" (such as DNA gyrases, RNA polymerases, and translation elongation factors), one particular carbohydrate-degrading enzyme, beta-glucosidase (COG1472 or K05349), had high abundance in both metagenomes (∼0.5%) and metatranscriptomes (∼1.0%). To better characterize diversity of carbohydrate-degrading enzymes, we focused on glycoside hydrolase (GH) groups, based on the Pfam annotation. A total of 60 GH groups were detected at the DNA level, of which, 39 GH groups were also detected at the RNA level (Supplementary Table 4). According to enzymatic activities of those GH groups, diverse carbohydrate-degrading genes specific for beta-glycosidic linkages in plant polysaccharides / oligosaccharides were detected in the cecal metagenome and metatranscriptome (**Figure 7**). Among them, GH3 was the most abundant group at both DNA and RNA levels. Moreover, compared to the DNA-level background, GH9, GH48, GH43, and GH53 were enriched at the RNA level (**Figure 7**).

Basically, taxonomic and functional profiles from individuals (FS1 and FS2) were comparable (**Figures 4**–**7**), with a high similarity ($r > 0.9$ and $P < 0.01$) for either metagenomes or metatranscriptomes.

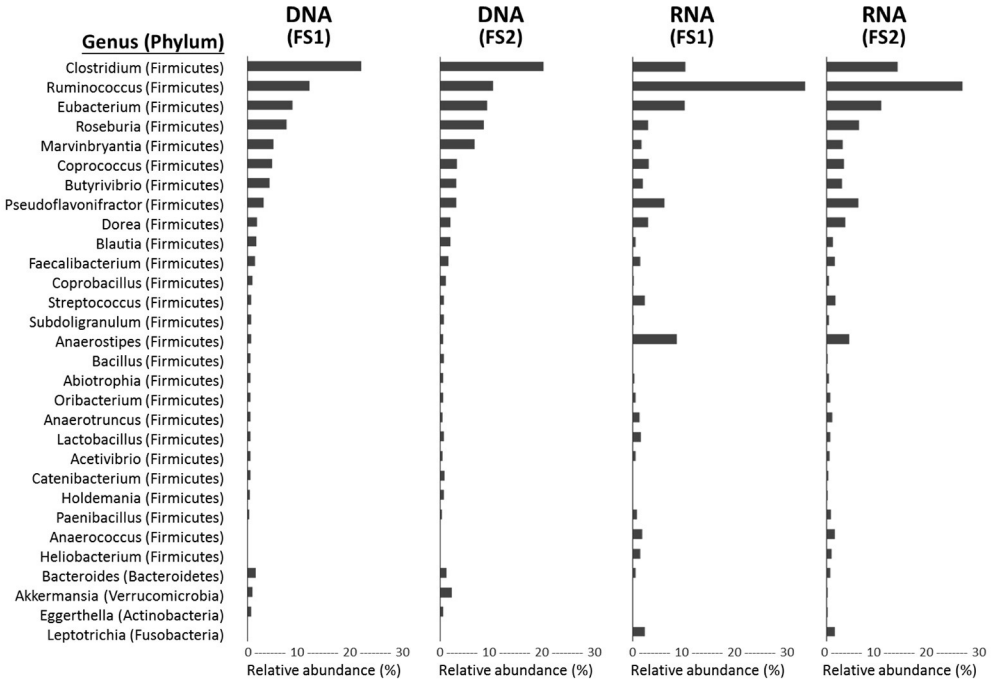


FIGURE 4 | Genus-level taxonomic compositions of metagenomes (DNA-level) and metatranscriptomes (RNA-level) of cecal microbiota from two flying squirrels (FS1 and FS2). Top 30 abundant genera that constituted >0.5% in either library are shown, with their phyla in parentheses.

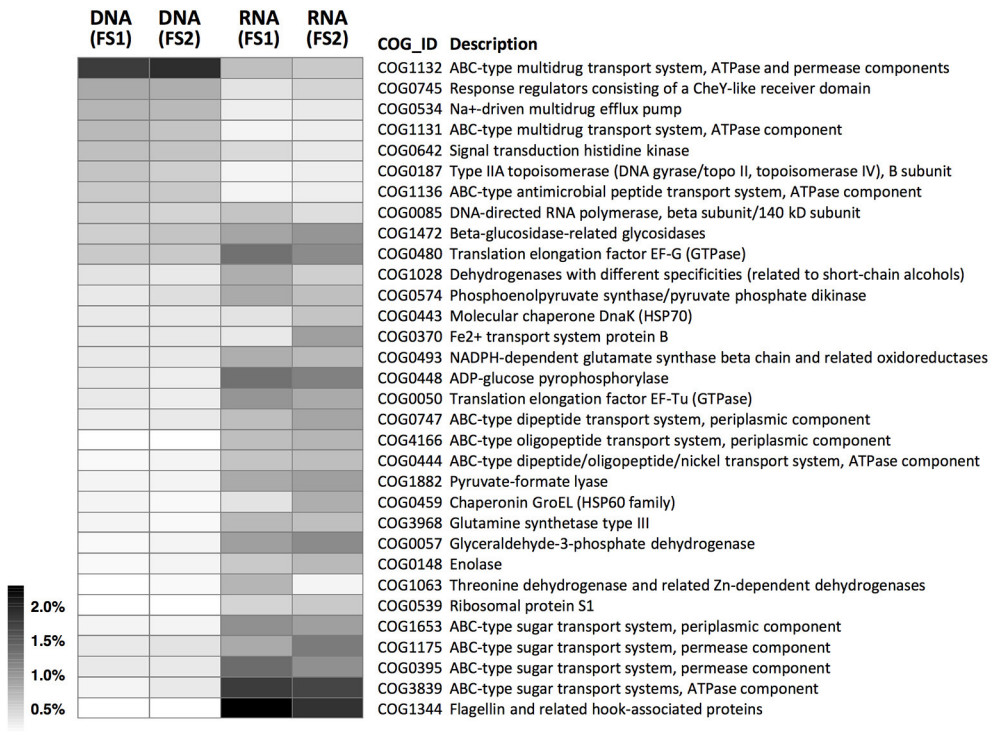


FIGURE 5 | Abundance distributions of COGs in metagenomes (DNA-level) and metatranscriptomes (RNA-level) of cecal microbiota from two flying squirrels (FS1 and FS2). Only abundant COGs that constituted >0.5% in either library are shown.

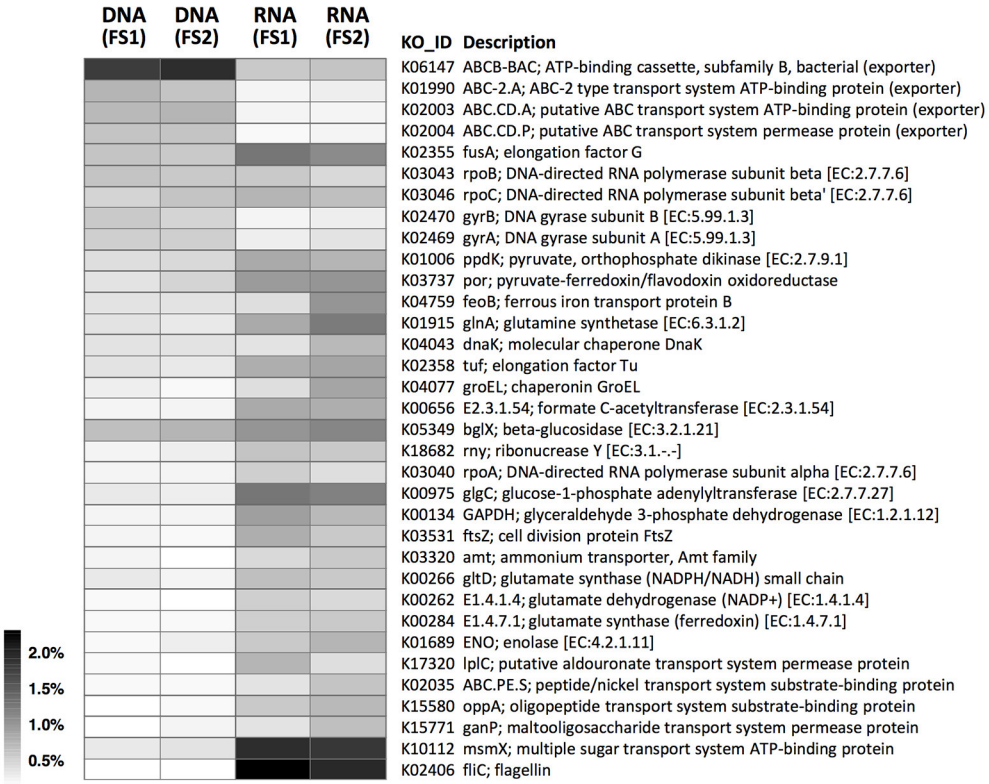


FIGURE 6 | Abundance distributions of KOs in metagenomes (DNA-level) and metatranscriptomes (RNA-level) of cecal microbiota from two flying squirrels (FS1 and FS2). Only abundant KOs that constituted >0.5% in either library are shown.

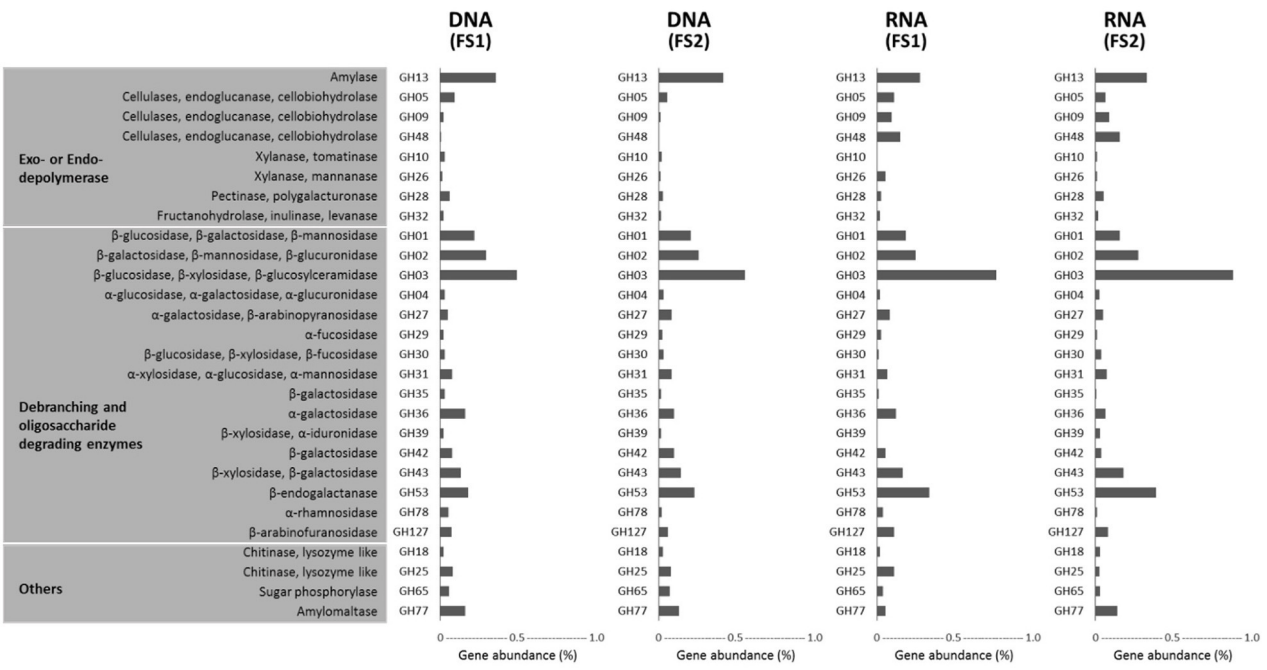


FIGURE 7 | Glycoside hydrolases (GHs) detected in metagenomes (DNA-level) and metatranscriptomes (RNA-level) of cecal microbiota from two flying squirrels (FS1 and FS2). Only abundant GHs that constituted >0.05% in either library are shown.

Meta-Analysis of Mammalian Gut Metagenomes

To provide an overview of functional characteristics of mammalian gut microbiota, in addition to 2 metagenomes from the flying squirrel's cecum, we conducted a meta-analysis on mammalian gut metagenomes with published datasets (including four datasets from the cow's rumen, and 39 datasets from fecal samples of zoo carnivores/omnivores/herbivores). Functional gene compositions based on the COG annotation (**Figure 8**) demonstrated that: (1) functional profiles of cecal metagenomes and ruminal metagenomes were distinct from those based on fecal samples; and (2) functional profiles of cecal metagenomes were also distinct from those of ruminal metagenomes. Similar functional relationships were obtained based on the Pfam annotation (Supplementary Figure 17).

To reveal unique functional characteristics of the flying squirrel's cecal microbiota, we conducted statistical tests among 5 metagenome groups (i.e., flying squirrel's cecum, cow's rumen, zoo carnivores' feces, zoo omnivores' feces, and zoo herbivores' feces). In comparisons based on relative abundances of COG functional categories, many categories had significant differences among 5 metagenome groups (Supplementary Table 5). Moreover, post-hoc pairwise comparisons focusing on specific differences between the flying squirrel's cecum and the other four groups (Supplementary Table 6) indicated that, despite several significant pairs, the flying squirrel's cecum had clear differences from all other groups in one particular category, namely defense mechanisms (V). Specifically, compared to other mammalian gut metagenomes, the flying squirrel's cecal metagenome contained a significantly higher percentage of gene function involved in defense mechanisms (Supplementary Table 5), especially for those COG gene families assigned to multidrug efflux pumps (Supplementary Table 7).

In addition, in a comparison of glycoside hydrolase (GH) groups among all 5 metagenome groups, the flying squirrel's cecal metagenome contained relatively high levels of GH3, GH43, and GH53 (Supplementary Table 8); those GH groups were specific for degradation of oligosaccharides and were actually enriched in cecal metatranscriptions (**Figure 7**).

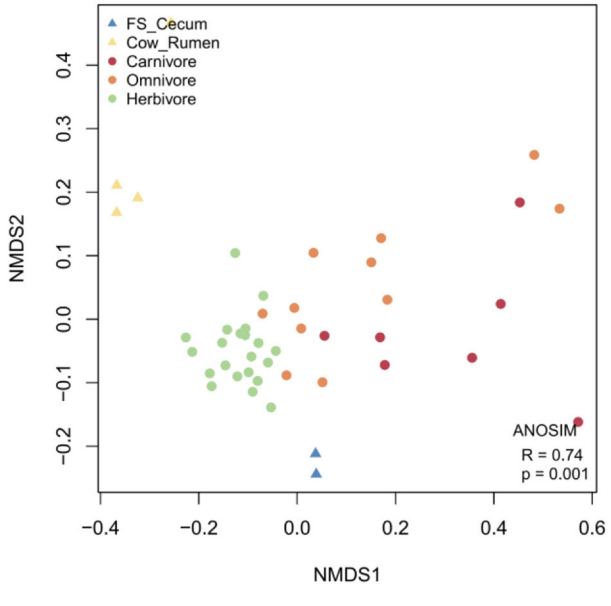


FIGURE 8 | NMDS plot based on COG compositions, revealing functional similarity / dissimilarity among mammalian gut metagenomes. In addition to two datasets from the flying squirrel (FS) cecum of this study, publicly available gut metagenomes, including four datasets from the cow's rumen and 39 datasets from the fecal samples of zoo mammals (categorized by diets: carnivore, omnivore, and herbivore), were included for comparison.

DISCUSSION

In this study, we demonstrated specialized digestive strategies of the white-faced flying squirrel (*Petaurista alborufus lena*), including powerful molars to facilitate mastication of tree leaves and an extended cecum to store ingesta for microbial activities (**Figure 1**). Notably, size distribution of feed particles in the large intestine was more similar to that in the stomach than in the cecum (**Figure 1D** and Supplementary Figure 2), indicating that some large feed particles might go directly to the large intestine (bypassing the cecum), corresponding to the description of specialized sorting at the ileal-cecal-colic junction in small herbivores (Sakaguchi, 2003). Overall, based on these anatomical/physiological characteristics, we inferred that transformation of dietary compounds into absorbable nutrients would primarily be conducted in the cecum with the aid of microbial activities.

Based on mass spectrometry to detect metabolites in feed contents, distinct gut compartments contained various types/levels of compounds (**Figure 2**). More importantly, many phytochemicals (leaf-derived compounds) were present only in the stomach and small intestine, but were not detected in the cecum and large intestine (**Figure 3**). We inferred that those phytochemicals were likely released from plant cells following exposure to a low pH and enzymes in the foregut, and later degraded or transformed by gut microbes in the hindgut (Deprez et al., 2000; Williamson et al., 2000). Some phytochemicals with sugar moieties could be used by gut microbes (Cardona et al., 2013), as a key factor structuring symbiotic microbial communities (Patra and Saxena, 2009; Laparra and Sanz, 2010; Ni et al., 2015). However, many phytochemical derivatives might be unwanted by bacterial cells (Cowan, 1999). We found that the flying squirrel's cecal metagenome contained various types of multidrug exporter genes with high abundances (**Figures 5**, **6**), which may enable gut microbes to confer resistance to leaf-derived compounds or host-defense molecules (Neyfakh, 1997; Putman et al., 2000; Piddock, 2006). Alternatively, those efflux pumps may have other roles relevant to bacterial behavior in the cecal environment, such as quorum sensing and cell motility (Martinez et al., 2009; Buckner et al., 2016). Further studies are required to determine targets and roles of those genes.

Based on metagenomes and metatranscriptomes, the flying squirrel's cecal microbial communities were exceedingly dominated by bacterial Firmicutes taxa, including diverse genera (**Figure 4**). These taxonomic profiles based on protein-coding

sequences were consistent with our previous results based on 16S rRNA gene sequences (Lu et al., 2014), suggesting that these Firmicutes taxa in general have adapted to the flying squirrel's cecum and contribute to major microbial activities in the gut environment.

We detected up to 60 types of glycoside hydrolases in the flying squirrel's cecal metagenome (Supplementary Table 4). This high diversity of GH groups was comparable to the diversity detected in other mammalian herbivores (Brulc et al., 2009; Flint et al., 2012; Wang et al., 2013; Jose et al., 2017), whereas detailed combinations and relative abundances of GH groups seemed to differ among animals with distinct gut structures and dietary preferences (Supplementary Table 8). For example, there were relatively high abundances of GH3, GH43, and GH53 in the flying squirrel's cecal metagenome, with even enriched levels in the cecal metatranscriptome (**Figure 8**). Despite different substrate specificities, these three dominant GH groups are involved in the degradation of beta-linkage oligosaccharides (Cantarel et al., 2009), releasing various monosaccharides. It is noteworthy that these GH enzymes may also have important roles in the degradation of phytochemicals (Deprez et al., 2000; Williamson et al., 2000), especially phenolic metabolites (e.g., flavonoids) that typically contain a sugar glycosidic linkage. That probably accounted for decreasing flavonoid compounds

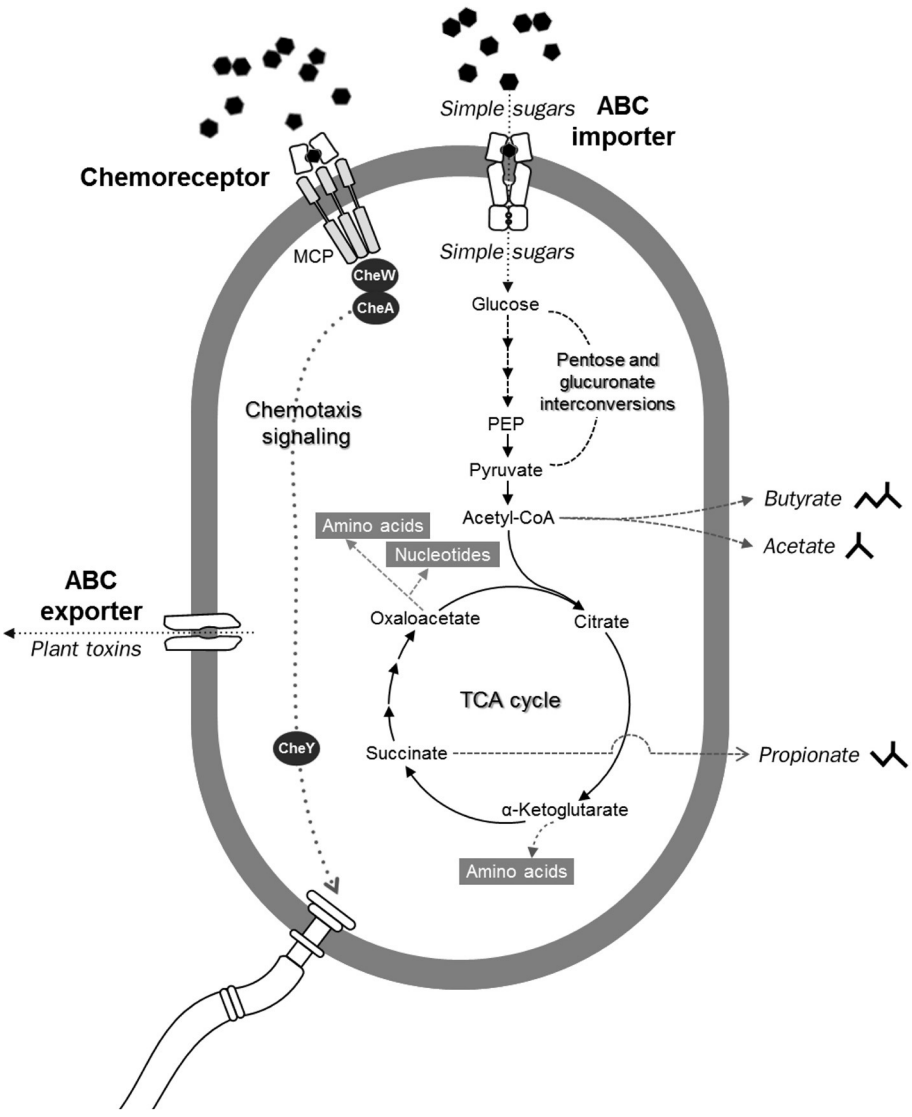


FIGURE 9 | Diagrammatic illustration of potentially crucial microbial functions in the flying squirrel's cecum, based on genes and compounds detected in the metagenome, metatranscriptome, and metabolome. From top to bottom: simple sugars were transported into microbial cells through various ABC importers, and subsequently fermented into short-chain fatty acids. Meanwhile, substrate-binding components of ABC importers were also involved in chemotaxis signaling, enabling bacteria to move toward higher concentrations of sugars by flagellar motility. Left: ABC exporters involved in defense mechanisms may enable gut microbes to export toxic phytochemical derivatives.

in the cecum (**Figure 3**). More importantly, those GH sequences contained a high nucleotide variation, and in fact, were assigned to various bacterial taxa (mostly belonging to Firmicutes, as listed in **Figure 4**), suggesting the importance of those enzymatic activities for microbes living in the flying squirrel's cecum.

In addition to carbohydrate-degrading ability (**Figure 8**), several genes involved in ABC-type sugar importers (Supplementary Figure 9) were detected in the flying squirrel's cecal microbiota, with many up-regulated at the RNA level (**Figures 5**, **6**). Based on these findings, we inferred that the cecal microbiota may have a substantial ability to quickly transport simple sugars into cells, after investing enzymes for hydrolyzing plant fibers into monosaccharides (**Figure 9**; upper part). Moreover, pathways involved in substrate-induced cell motility (Di Paola et al., 2004), including chemotaxis signaling (Supplementary Figure 7) and flagellar structure (Supplementary Figure 8), were detected in both metagenomes and metatranscriptomes, indicating directional cell movement (i.e., sensing the gradient of surrounding nutrients and moving toward stimulatory chemicals; **Figure 9**; top to bottom of left side) were likely important for cecal microbiota. Notably, chemotaxis signaling and ABC-type transport systems have overlapping components (**Figure 9**; upper part); that is, the substrate-binding proteins of the ABC-type imports (Supplementary Figure 9) also mediate environmental stimuli for chemotaxis receptors (i.e., methyl-accepting chemotaxis protein, MCP; Supplementary Figure 7; Neumann et al., 2010). Thus, when sugar-binding genes were highly expressed (**Figures 5**, **6**), it may not only allow microbes to quickly import sugars but also assist them to move toward sugar-rich microhabitats, thereby optimizing sugar acquisition in the cecum.

After being transported into the bacterial cell, sugar monomers (such as glucose, xylose, and arabinose) would be fermented into short-chain fatty acids (e.g., butyrate, acetate, and propionate) and subsequently converted into other cellular components (**Figure 9**; top to bottom of right side). In the cecal metagenome and metatranscriptome, pathways regarding pentose and glucuronate interconversions (Supplementary Figure 10), energy generating from sugar fermentation (Supplementary Figures 11, 12), and biosynthesis of various macromolecules (Supplementary Figures 13–16) were well-represented, indicating that plant-sourced sugars could be efficiently transformed into bacterial biomass in the flying squirrel's cecum, with adequate fermentation end-products to meet host requirements (Bergman, 1990; Stevens and Hume, 1998).

Our interpretation about functional characteristics of the cecal microbiota was based on the consistent patterns of two flying squirrel individuals, which may not be enough to robustly demonstrate how symbiotic microbes respond to unique environmental conditions in the cecum. Nevertheless, this study pointed out the importance of investigations into gut microbiota of small mammalian herbivores. Future research focusing on more individuals from leaf-eating flying squirrels and other small mammalian herbivores is demanded to capture the functional uniqueness of the cecal microbiota, considering different feeding adaptations and gut structures of large vs. small mammalian herbivores (Stevens and Hume, 1995; Mackie, 2002).

In the present study, we delineated anatomical/physiological characteristics of the flying squirrel's digestive system (**Figure 1**) and demonstrated functional characteristics of cecal microbiota based on multiple meta-omic data, including metabolomic profiles (**Figures 2**, **3**) and parallel metagenome-metatranscriptome profiles (**Figures 4**–**8**). Here, we summarize crucial metabolic capacities of the flying squirrel's cecal microbiota (**Figure 9**), including the ability to: (1) secrete various glycoside hydrolases to degrade plant fibers into simple sugars; (2) transport simple sugars into the cells while also moving toward sugar-rich microhabitats; and (3) ferment sugars into short-chain fatty acids to generate energy for synthesizing amino acids and nucleotides. Moreover, compared to other mammalian gut metagenomes, the cecal metagenome of the flying squirrel tends to contain more multidrug exporter genes, which may enable gut microbes to export leaf-derived compounds. This study provided a molecular basis to promote understanding functional characteristics of symbiotic gut microbiota of small mammals with folivorous dietary habits in the wild.

DATA ACCESSIBILITY

NCBI Sequence Read Archive: SRX2118787—SRX2118794.

AUTHOR CONTRIBUTIONS

H-PL, P-YL, and H-TY conceived the study design and collected samples; H-PL, P-YL, J-FH, S-WH, and H-CH conducted experiments; H-PL, P-YL, YW, and C-YL conducted bioinformatics analyses; H-PL, CH, and H-TY wrote the first draft. All authors contributed to data interpretation and preparation of the final manuscript. All authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

We acknowledge helpful comments from Manyuan Long and Jer-Horng Wu. We thank the Metabolomics Core (TCX-D800) and Geen-Dong Chang for his help in spectrometry analysis, Technology Commons in College of Life Science and Center for Systems Biology (NTU, Taiwan) for conducting liquid chromatography-mass spectrometry (LC-MS). We thank the Genome Research Center in National Yang-Ming University for sequencing. This work was supported by the Ministry of Science 1245 and Technology, Taiwan (MOST 1032311B002001 and 1062633B006004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02622/full#supplementary-material

REFERENCES

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., and Narasimhan, G. (2016). Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evol. Bioinform. Online* 12, 5–16. doi: 10.4137/EBO.S36436

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L., and Gilbert, M. T. P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation? *Trends Ecol. Evol.* 31, 689–699. doi: 10.1016/j.tree.2016.06.008

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* 70, 567–590.

Brulc, J. M., Antonopoulos, D. A., Miller, M. E. B., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., et al. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1948–1953. doi: 10.1073/pnas.0806191105

Buckner, M. M. C., Blair, J. M. A., La Ragione, R. M., Newcombe, J., Dwyer, D. J., Ivens, A., et al. (2016). Beyond antimicrobial resistance: evidence for a distinct role of the AcrD efflux pump in *Salmonella biology*. *mBio* 7:e01916-16. doi: 10.1128/mBio.01916-16

Campbell, J. L., Eisemann, J. H., Williams, C. V., and Glenn, K. M. (2000). Description of the gastrointestinal tract of five lemur species: *Propithecus tattersalli*, *Propithecus verreauxi* coquereli, *Varecia variegata*, *Hapalemur griseus*, and *Lemur catta*. *Am. J. Primatol.* 52, 133–142. doi: 10.1002/1098-2345(200011)52:3<133::AID-AJP2>3.0.CO;2-#

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Cardona, F., Andres-Lacueva, C., Tulipani, S., Tinahones, F. J., and Queipo-Ortuno, M. I. (2013). Benefits of polyphenols on gut microbiota and implications in human health. *J. Nutr. Biochem.* 24, 1415–1422. doi: 10.1016/j.jnutbio.2013.05.001

Clauss, M., Lechner-Doll, M., and Streich, W. J. (2002). Faecal particle size distribution in captive wild ruminants: an approach to the browser/grazer dichotomy from the other end. *Oecologia* 131, 343–349. doi: 10.1007/s00442-002-0894-8

Coley, P. D., and Barone, J. A. (1996). Herbivory and plant defenses in tropical forests. *Annu. Rev. Ecol. Syst.* 27, 305–335. doi: 10.1146/annurev.ecolsys.27.1.305

Cowan, M. M. (1999). Plant products as antimicrobial agents. *Clin. Microbiol. Rev.* 12, 564–582.

Cui, Q., Lewis, I. A., Hegeman, A. D., Anderson, M. E., Li, J., Schulte, C. F., et al. (2008). Metabolite identification via the madison metabolomics consortium database. *Nat. Biotechnol.* 26, 162–164. doi: 10.1038/nbt0208-162

Deprez, S., Brezillon, C., Rabot, S., Philippe, C., Mila, I., Lapierre, C., et al. (2000). Polymeric proanthocyanidins are catabolized by human colonic microflora into low-molecular-weight phenolic acids. *J. Nutr.* 130, 2733–2738.

Deusch, S., Camarinha-Silva, A., Conrad, J., Beifuss, U., Rodehutscord, M., and Seifert, J. (2017). A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments. *Front. Microbiol.* 8:1605. doi: 10.3389/fmicb.2017.01605

Di Paola, V., Marijuan, P. C., and Lahoz-Beltra, R. (2004). Learning and evolution in bacterial taxis: an operational amplifier circuit modeling the computational dynamics of the prokaryotic 'two component system' protein network. *Biosystems.* 74, 29–49. doi: 10.1016/j.biosystems.2004.01.003

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3, 289–306. doi: 10.4161/gmic.19897

Fritz, J., Hummel, J., Kienzle, E., Arnold, C., Nunn, C., and Clauss, M. (2009). Comparative chewing efficiency in mammalian herbivores. *Oikos* 118, 1623–1632. doi: 10.1111/j.1600-0706.2009.17807.x

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M. (2003). Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* 125, 11853–11865. doi: 10.1021/ja036030u

Hervé, M. (2017). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics*. R package version 0.9–68.

Hird, S. M. (2017). Evolutionary biology needs wild microbiomes. *Front. Microbiol.* 8:725. doi: 10.3389/fmicb.2017.00725

Hume, I. D. (1989). Optimal digestive strategies in mammalian herbivores. *Physiol. Zool.* 62, 1145–1163. doi: 10.1086/physzool.62.6.30156206

Hume, I. D. (2002). Digestive strategies of mammals. *Acta Zool. Sin.* 48, 1–19.

Jose, V. L., More, R. P., Appoothy, T., and Arun, A. S. (2017). In depth analysis of rumen microbial and carbohydrate-active enzymes profile in Indian crossbred cattle. *Syst. Appl. Microbiol.* 40, 160–170. doi: 10.1016/j.syapm.2017.02.003

Kamra, D. N. (2005). Rumen microbial ecosystem. *Curr. Sci.* 89, 124–135. doi: 10.1007/978-94-009-1453-7

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kohl, K. D., Stengel, A., and Dearing, M. D. (2016). Inoculation of tannin-degrading bacteria into novel hosts increases performance on tannin-rich diets. *Environ. Microbiol.* 18, 1720–1729. doi: 10.1111/1462-2920.12841

Kohl, K. D., Weiss, R. B., Cox, J., Dale, C., and Dearing, M. D. (2014). Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecol. Lett.* 17, 1238–1246. doi: 10.1111/ele.12329

Konstantinidis, K. T., Rossello-Mora, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi: 10.1038/ismej.2017.113

Kuo, C. C., and Lee, L. L. (2003). Food availability and food habits of Indian giant flying squirrels (*Petaurista philippensis*) in Taiwan. *J. Mammal.* 84, 1330–1340. doi: 10.1644/BOS-039

Laparra, J. M., and Sanz, Y. (2010). Interactions of gut microbiota with functional food components and nutraceuticals. *Pharmacol. Res.* 61, 219–225. doi: 10.1016/j.phrs.2009.11.001

Lee, P., Progulske, D., and Lin, Y. (1986). Ecological studies on two sympatric Petaurista species in Taiwan. *Bull. Inst. Zool. Acad. Sin.* 25, 113–124.

Leser, T. D., and Molbak, L. (2009). Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host. *Environ. Microbiol.* 11, 2194–2206. doi: 10.1111/j.1462-2920.2009.01941.x

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., et al. (2008). Evolution of mammals and their gut microbes. *Science* 320, 1647–1651. doi: 10.1126/science.1155725

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Lu, H. P., Lai, Y. C., Huang, S. W., Chen, H. C., Hsieh, C. H., and Yu, H. T. (2014). Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. *Sci. Rep.* 4:6185. doi: 10.1371/journal.pone.0163720

Lu, H. P., Wang, Y. B., Huang, S. W., Lin, C. Y., Wu, M., Hsieh, C. H., et al. (2012). Metagenomic analysis reveals a functional signature for biomass degradation by cecal microbiota in the leaf-eating flying squirrel (*Petaurista alborufus lena*). *BMC Genomics* 13:466. doi: 10.1186/1471-2164-13-466

Mackie, R. I. (2002). Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution. *Integr. Comp. Biol.* 42, 319–326. doi: 10.1093/icb/42.2.319

Manning, P. J., Ringler, D. H., and Newcomer, C. E. (1994). *The Biology of the Laboratory Rabbit*. San Diego, CA: Academic Press.

Mao, S. Y., Huo, W. J., and Zhu, W. Y. (2016). Microbiome-metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain in a goat model. *Environ. Microbiol.* 18, 525–541. doi: 10.1111/1462-2920.12724

Martinez, J. L., Sanchez, M. B., Martinez-Solano, L., Hernandez, A., Garmendia, L., Fajardo, A., et al. (2009). Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol. Rev.* 33, 430–449. doi: 10.1111/j.1574-6976.2008.00157.x

Mendiburu, F. D. (2016). *Agricolae: Statistical Procedures for Agricultural Research*. R package version 1.2–4.

Mithofer, A., and Boland, W. (2012). Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* 63, 431–450. doi: 10.1146/annurev-arplant-042110-103854

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., Gonzalez, A., Fontana, L., et al. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* 332, 970–974. doi: 10.1126/science.1198719

Neumann, S., Hansen, C. H., Wingreen, N. S., and Sourjik, V. (2010). Differences in signalling by directly and indirectly binding ligands in bacterial chemotaxis. *EMBO J.* 29, 3484–3495. doi: 10.1038/emboj.2010.224

Neyfakh, A. A. (1997). Natural functions of bacterial multidrug transporters. *Trends Microbiol.* 5, 309–313. doi: 10.1016/S0966-842X(97)01064-0

Ni, Y. Q., Li, J., and Panagiotou, G. (2015). A molecular-level landscape of diet-gut microbiome interactions: toward dietary interventions targeting bacterial genes. *mBio* 6:e01263-15. doi: 10.1128/mBio.01263-15

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2017). *Vegan: Community Ecology Package*. R package version 2.4–4.

Oshida, T., Lin, L.-K., Chang, S.-W., Chen, Y.-J., and Lin, J.-K. (2011). Phylogeography of two sympatric giant flying squirrel subspecies, *Petaurista alborufus lena* and *P. philippensis* grandis (Rodentia: Sciuridae), in Taiwan. *Biol. J. Linn. Soc.* 102, 404–419. doi: 10.1111/j.1095-8312.2010.01576.x

Patra, A. K., and Saxena, J. (2009). Dietary phytochemicals as rumen modifiers: a review of the effects on microbial populations. *Anton. Van Leeuwenhoek Int. J. Gen. Mol. Microbiol.* 96, 363–375. doi: 10.1007/s10482-009-9364-1

Piddock, L. J. (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nat. Rev. Microbiol.* 4, 629–636. doi: 10.1038/nrmicro1464

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864

Putman, M., Van Veen, H. W., and Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. *Microbiol. Mol. Biol. Rev.* 64, 672–693. doi: 10.1128/M.M.B.R.64.4.672-693.2000

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*, *3.3.2 Edn*. Vienna: The R Foundation for Statistical Computing Platform.

Rho, M. N., Tang, H. X., and Ye, Y. Z. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38:e191. doi: 10.1093/nar/gkq747

Sakaguchi, E. (2003). Digestive strategies of small hindgut fermenters. *Anim. Sci. J.* 74, 327–337. doi: 10.1046/j.1344-3941.2003.00124.x

Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., et al. (2012). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 40, D13–D25. doi: 10.1093/nar/gkr1184

Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., and Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* 9:666. doi: 10.1038/msb.2013.22

Smith, A. T., and Johnston, C. H. (2016). *Petaurista alborufus. The IUCN Red List of Threatened Species* 2016, e.T16718A22272572.

Smith, C. A., Want, E. J., O'maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787. doi: 10.1021/ac051437y

Stevens, C. E., and Hume, I. D. (1995). *Comparative Physiology of the Vertebrate Digestive System*. New York, NY: Cambridge University Press.

Stevens, C. E., and Hume, I. D. (1998). Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiol. Rev.* 78, 393–427.

Takami, H., Taniguchi, T., Moriya, Y., Kuwahara, T., Kanehisa, M., and Goto, S. (2012). Evaluation method for the potential functionome harbored in the genome and metagenome. *BMC Genomics* 13:699. doi: 10.1186/1471-2164-13-699

Tautenhahn, R., Bottcher, C., and Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* 9:504. doi: 10.1186/1471-2105-9-504

Wallace, R. J. (1992). Rumen microbiology, biotechnology and ruminant nutrition - the application of research findings to a complex microbial ecosystem. *FEMS Microbiol. Lett.* 100, 529–534. doi: 10.1111/j.1574-6968.1992.tb05751.x

Wang, L. L., Hatem, A., Catalyurek, U. V., Morrison, M., and Yu, Z. T. (2013). Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS ONE* 8:e78507. doi: 10.1371/journal.pone.0078507

Weimer, P. J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Front. Microbiol.* 6:296. doi: 10.3389/fmicb.2015.00296

Williamson, G., Day, A. J., Plumb, G. W., and Couteau, D. (2000). Human metabolic pathways of dietary flavonoids and cinnamates. *Biochem. Soc. Trans.* 28, 16–22. doi: 10.1042/bst0280016

Zhou, B., Wang, J., and Ressom, H. W. (2012). MetaboSearch: tool for mass-based metabolite identification using multiple databases. *PLoS ONE* 7:e40096. doi: 10.1371/journal.pone.0040096

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Lu, Liu, Wang, Hsieh, Ho, Huang, Lin, Hsieh and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 March 2017
doi: 10.3389/fmicb.2017.00480

Differential Proteomic Profiles of *Pleurotus ostreatus* in Response to Lignocellulosic Components Provide Insights into Divergent Adaptive Mechanisms

Qiuyun Xiao[1,2], Fuying Ma[1], Yan Li[2], Hongbo Yu[1], Chengyun Li[2] and Xiaoyu Zhang[1]**

[1] *Key Laboratory of Molecular Biophysics of MOE, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China,* [2] *Key Laboratory of Agro-Biodiversity and Pest Management of Education Ministry of China, Yunnan Agricultural University, Kunming, China*

Pleurotus ostreatus is a white rot fungus that grows on lignocellulosic biomass by metabolizing the main constituents. Extracellular enzymes play a key role in this process. During the hydrolysis of lignocellulose, potentially toxic molecules are released from lignin, and the molecules are derived from hemicellulose or cellulose that trigger various responses in fungus, thereby influencing mycelial growth. In order to characterize the mechanism underlying the response of *P. ostreatus* to lignin, we conducted a comparative proteomic analysis of *P. ostreatus* grown on different lignocellulose substrates. In this work, the mycelium proteome of *P. ostreatus* grown in liquid minimal medium with lignin, xylan, and carboxymethyl cellulose (CMC) was analyzed using the complementary two-dimensional gel electrophoresis (2-DE) approach; 115 proteins were identified, most of which were classified into five types according to their function. Proteins with an antioxidant function that play a role in the stress response were upregulated in response to lignin. Most proteins involving in carbohydrate and energy metabolism were less abundant in lignin. Xylan and CMC may enhanced the process of carbohydrate metabolism by regulating the level of expression of various carbohydrate metabolism-related proteins. The change of protein expression level was related to the adaptability of *P. ostreatus* to lignocellulose. These findings provide novel insights into the mechanisms underlying the response of white-rot fungus to lignocellulose.

Keywords: *Pleurotus ostreatus*, proteomics, white-rot fungus, fungal adaptability, lignocellulose

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Naresh Singhal, University of Auckland, New Zealand
Seung Gu Shin, Pohang University of Science and Technology, South Korea

****Correspondence:***
Chengyun Li
licheng_yun@163.com
Xiaoyu Zhang
zhangxiaoyu@hust.edu.cn

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *15 December 2016*
Accepted: *08 March 2017*
Published: *23 March 2017*

Citation:
Xiao Q, Ma F, Li Y, Yu H, Li C and Zhang X (2017) Differential Proteomic Profiles of Pleurotus ostreatus in Response to Lignocellulosic Components Provide Insights into Divergent Adaptive Mechanisms. Front. Microbiol. 8:480. doi: 10.3389/fmicb.2017.00480

INTRODUCTION

To adapt to changing environments, fungi have developed mechanisms to sense and respond to a multitude of environmental factors such as different carbon sources (Akai, 2012; Kües, 2015). *P. ostreatus* is a white-rot fungus that can be easily cultivated on a variety of lignocellulosic substrates, owing to its ability to degrade cellulose, lignin, and hemicellulose through the action of complex oxidative and hydrolytic enzymatic systems (Fernández-Fueyo et al., 2016). However, lignin does not act as the sole source of carbon and energy; the degradation of lignin by white-rot fungi enables access to holocellulose, which is the carbon and energy source for this species. Presumably, cellulose and hemicellulose provide

carbon and energy sources for growth, whereas lignin serves a barrier to prevent *P. ostreatus* from attacking polysaccharides. Lignin likely acts as the target for enzymes participating in degradation. manganese peroxidase (MnP) and laccase are the major oxidative enzymes secreted by *P. ostreatus* that are responsible for the oxidation of lignin and a wide range of lignin-analogous compounds (Wan and Li, 2012). In addition, various auxiliary enzymes generate hydrogen peroxide, which is required for oxidation of lignin. During the lignin degradation process, aromatic radicals are produced that catalyze subsequent degradation, generating potentially toxic molecules that trigger a defense response to protect the fungus from harmful environments (Li et al., 2015b). Primary mycelial enzymes play important roles in cellular processes involving utilization of lignocellulose; earlier studies revealed that the use of conditional transitions in biological pretreatment would affect the expression of the white rot fungi genes encoding ligninolytic enzymes at the transcriptional level (Sindhu et al., 2016).

After the lignin barrier is broken, *P. ostreatus* attacks lignocellulosic polysaccharides. The most abundant hemicellulose is xylan, which is composed of pentoses such as xylose, whereas the most abundant form of cellulose is glucose. The degradation of hemicellulose and cellulose is dependent on carbohydrate-active enzymes, whose functions do not overlap (Lombard et al., 2014); therefore, a large number of different enzymes is required for hemicellulose and cellulose degradation.

Flavin adenine dinucleotide (FAD)-dependent proteins are a current research focus, as these enzymes play important roles in lignocellulose oxidation (Levasseur et al., 2013). Flavin-mediated oxidation, which involves dioxygen as the electron acceptor, is thermodynamically favorable (Hamdane et al., 2015). Previous studies of the response of flavoproteins to lignin have focused on the role of extracellular flavoprotein during lignocellulose degradation (Hernández-Ortega et al., 2012); however, there have been few reports on the role of intracellular flavoproteins in lignocellulose degradation.

In addition, the molecular mechanisms underlying the mycelial response to hemicellulose, cellulose, and lignin remain poorly understood. Recent studies have shown that cellular responses to lignin derivatives are critical for optimization of ligninolytic conditions in fungal cells (Simon et al., 2014). Therefore, elucidation of the catalytic functions of lignin-responsive enzymes is necessary.

The degradation of lignocellulose by *P. ostreatus* plays a role in the acclimation of this fungus to the environment. Adaptation to the specific environment is mediated via profound changes in the expression of genes, which leads to changes in the composition of the fungal transcriptome, proteome, and metabolome (Gaskell et al., 2016). On the basis of their activity, proteins are traditionally classified as catalysts, signaling molecules, or building blocks in cells and microorganisms. Therefore, researchers have attempted to explore the mechanism underlying the interaction between fungi and lignocellulose by proteomics. Proteomics analysis of the filamentous fungus *Trichoderma atroviride* grown on cell walls identified 24 upregulated proteins, including fungal cell wall-degrading enzymes such as *N*-acetyl-β-D-glucosaminidase and the 42-kDa protein endochitinase (Grinyer et al., 2005). Proteomic analysis of *Botrytis cinerea* revealed that proteins such as malate dehydrogenase or peptidyl-prolyl cis–trans isomerase from the mycelium were differentially expressed among strains when using CMC as the sole carbon source; these proteins are involving in host-tissue invasion, pathogenicity, and fungal development (González-Fernández et al., 2014). These studies attempted to elucidate the effects of plant cell wall composition on microbes by mixing lignocellulose or cellulose as substrates; however, they only provide limited evidence that the main components of the plant cell wall alter the gene expression in fungal cells, and that lignin and hemicellulose might also affect the growth and protein expression of fungal cells. To date, few studies have been published regarding the intracellular proteomics of the white-rot fungal response to lignocellulose.

In this work, we performed two-dimensional protein fractionation coupled with mass spectrometry to analyze the potential biological differences among *P. ostreatus* cells grown on different lignocellulose media. *P. ostreatus* was grown in Kirk's medium to which lignin, xylan, and CMC were added; this medium is commonly used in studies of the response of white-rot fungus to lignocellulose. We compared the biomass and FAD concentration in cells during cultivation. Next, proteomic profiles of *P. ostreatus* under lignocellulose culture conditions were obtained. The 2-DE expression profiles were used to analyze the intracellular proteins differentially expressed in various substrates, and differentially expressed proteins were identified by MALDI-TOF-MS. Finally, the metabolic pathways involving in the lignocellulose response in *P. ostreatus* were examined according to the differentially expressed proteins in the various substrates.

MATERIALS AND METHODS

Microorganism and Cultivation

P. ostreatus isolate BP2 obtained from the Culture Collection Center, Huazhong Agriculture University (Hubei, China) was used in this study. The strain was maintained on potato dextrose agar (PDA) slants at 4°C and activated for 1 week on new PDA slant before use, then transferred into potato dextrose broth (PDB) medium for 7 days at 28°C as inoculum.

In order to exclude influence of other organics, the strain was inoculated into a 250 ml flask with 100 mL modified Kirk's liquid medium which just contain basal salt component as basic medium (Taniguchi et al., 2005). The Kirk's liquid medium contained: 9×10^{-3} mol/L KH_2PO_4, 3×10^{-3} mol/L $MgSO_4.7H_2O$, 2×10^{-5} mol/L ammonium tartrate, 3×10^{-4} mol /L $CaCl_2.2H_2O$, 5×10^{-2} mol/L glucose, and 10 ml/L trace element contained: 7.8×10^{-3} mol/L amino acetic acid, 1.2×10^{-3} mol/L $MgSO_4.H_2O$, 2.9×10^{-3} mol/L $MnSO_4.H_2O$, 1.7×10^{-2} mol/L NaCl, 3.59×10^{-4} mol/L $FeSO_4.7H_2O$, 7.75×10^{-4} mol/L $CoCl_2$, 9.0×10^{-4} mol/L $CaCl_2$, 3.48×10^{-4} mol/L $ZnSO_4.7H_2O$, 4×10^{-5} mol/L $CuSO_4.5H_2O$, 2.1×10^{-5} mol/L $AlK(SO_4)_2.12H_2O$, 1.6×10^{-4} mol/L H_3BO_3, 4.1×10^{-5} mol/L $NaMoO_4.2H_2O$.or 100 ml Kirk's liquid medium supplemented with 0.5 g lignin(Sigma), xylan(Sigma), or cellulose(Sigma). All

experiments were accompanied by controls that lacked the lignocellulose amendment. The mycelia were collected after 7 days incubation in dark at 28°C with continuous stirring at 120 r/min, and the cultures were centrifuged for collection washed with sterilized MilliQ water for several times to separate from medium, then kept at −80°C for use.

Growth Measurement

The mycelial dry weight was used to characterize *P. ostreatus* growth condition. Base on the method reported before (Taniwaki et al., 2006), the mycelium cultured in lignin as mentioned before was weighed after cultured for 0, 3, 5, 7, 9, 11 days. Three individual cultures of the mycelium were weighed at every time point.

Analysis of FAD Concentration during *P. ostreatus* Growth

Mycelium proteins were obtained using a dynamic high pressure homogenizing (GEA Niro Soavi S.p.A), and proteins were quantified by BCA method. Intracellular FAD concentration was measured using an FAD Colorimetric/Fluorometric Assay Kit (BioVision). Experimental methods refer to product description.

Laccase Activity Assays

Laccase activity was determined spectrophotometrically as previous study described by with 14 μmol of ABTS as the substrate (Srinivasan et al., 1995). All the assays were done at pH 3.0, the optimum pH for laccase of *P. ostreatus* with ABTS as the substrate.

2-De Analysis of Mycelia Protein

Frozen mycelia were used to extract total myceliaproteins by the TCA-acetone precipitation method (Rabilloud et al., 2010). Mycelia (dry weight of 1 g) was ground to a fine powder under liquid nitrogen and was collected into 50 ml microcentrifuge tubes. Three individual cultures of the mycelium were harvested and extracted separately. Twenty milliliters cold acetone (−20°C, 10% w/v trichloroacetic acid (TCA), 0.1% w/v dithiothreito (DTT, Bio-Rad), 1 mmol/L phenylmethanesulfonyl fluoride (PMSF, Sigma) was added into the tube. After the samples were resuspended totally, the tube was incubated at −20°C for more than 12 h, and then the samples were centrifuged for 20 min at 14,000 r/min. The resulting pellet was washed with 15 mL cold acetone (0.1% w/v DTT, 1 mmol/L PMSF), then centrifuged at 14,000 r/min for 20 min. This washing procedure was repeated twice and final pellet was resuspended. The pellet was vacuum-dried and solubilized with lysis buffer containing 7 mmol/L urea, 2%CHAPS (Sigma), 10 mmol/L DTT and 0.5% biolytes (Bio-Rad). After fully dissolved, the samples were stored at −80°C for 2-DE analysis. Protein concentration was determined using Bradford's method with bovine serum albumin as standards (Fernández and Novo, 2013). Ready strip IPG strips (18 cm, 4–7 linear pH gradient, Bio-Rad) were rehydrated for 12 h with 800 μg of protein sample as most mycelial proteins were in this range according to previous studies (Jami et al., 2010). Then the IPG were carried out for the first electrophoretic dimension in a Protean IEF-Cell (Bio-Rad). The isoelectric focusing was performed with a limiting current of 50 μA/strip following the program setting: (i) 250 v, rapid, 0.5 h. (ii) 1,000 v, rapid, 0.5 h. (iii) 9,000 v, liner, 4.5 h. (iv) 9,000 v, rapid, 75,000 vh(v) 500 v, rapid, 1 h. The IPG strips were treated twice for at least 30 in with SDS equilibration buffer (6 mmol/L urea, 1.5 mmol/L Tris-Cl with pH 8.8, 30% v/v glycerol, 2% (w/v) SDS, 0.001% bromophenol blue). Ten milligrams per milliliters DTT was add to the equilibration buffer in the first step, and 25 mg/mL iodoacetamide was added in the second step. The second dimensional SDS-polyacrylamide electrophoresis (SDS-PAGE) was performed on v/v 12.5% acrylamide gel (v/v 2% SDS) by using a Protean II xi Cell system (Bio-Rad). Coomassie PAGE Blue (Bio-Rad) was used to stain the gels. The finished gels were scanned with GE Gel Scan system (GE) and analyzed with PDQuest software (7.0.1 version, Bio-Rad). In order to verify the significant change of protein/spot, three replicate 2-DE gels were visually compared by using PDQuest software. The spots/proteins appeared in all three biological replicate could be considered the infallible spots/proteins, Finally, only differences with a ratio lignocellulose/control (R) $0.5 > R$ or $R > 2$ (CV $<$ 25%), and with a *t*-test ($p < 0.05$), were considered as significant. The theoretical pIs were calculated using the ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/).

ESI-MS/MS of 2-De Spots

Then, we performed MALDI-TOF/TOF to identify significantly changed spots in one or two cultures compared with that in the control. Spots from 2-DE gels were excised and digested with trypsin for 20 h. The resulting peptide mixtures were desalted using ZipTips C18 (Millipore), and eluted onto a 96-well MALDI target plate. Then, 2 mL samples on the plate were mixed with 1 mL supersaturated CHCA solution with 0.1% TFA and 50%ACN. Mass spectrometric analysis were measured on 5800 MALDI-TOF/TOF (AB SCIEX). Briefly, mass data acquisitions were piloted by 4000 Series Explorer Software v3.0 using batched-processing and automatic switching between MS and MS/MS modes The PMF data were collected and blasted in JGI database using MASCOT software (http://matrixscience.com).

RESULTS

Lignocellulose Components Influence the Growth of Mycelium

P. ostreatus grown in Kirk's medium supplemented with lignin, xylan, and CMC was used to study the relative intensity of proteins affected by lignocellulose, and Krik's medium without lignocellulose was used as control. The mycelial dry weights of colonies grown on lignocellulose significantly differed from those of the control (**Figure 1**). The growth of fungal mycelium was suppressed on lignin relative to other cultures. Xylan and CMC served as slow-acting carbon resources; accordingly, the biomass of mycelium in xylan and CMC accumulated slowly at first, and then began to surpass that of the control 7 days after inoculation. Compared with the control, lignocellulose supplementation suppressed mycelial growth for the first 7 days of culture; subsequently, mycelia underwent adaptation to xylan

and CMC, resulting in rapid growth of *P. ostreatus* in these medium.

Lignocellulose Components Influence the FAD Levels of Mycelia

FAD is a redox cofactor that plays an important role in metabolism (**Figure 2**). The primary sources of reduced FAD levels during eukaryotic metabolism are the citric acid cycle and beta oxidation reaction pathways. FAD accumulates with time, especially during growth on lignin. After inoculation for 7 days, the FAD concentration was higher in fungi grown in lignin than in other cultures.

Lignin Influence Laccase Activity

Since laccase is the most important extracellular enzyme responsible for lignin modification, we examined its activity in the lignin group (**Figure 3**). After inoculation, the laccase activity in this group was lower than that in the control for the first 5 days; however, after culturing for 7 days, laccase activity in the lignin group was higher than that in the control.

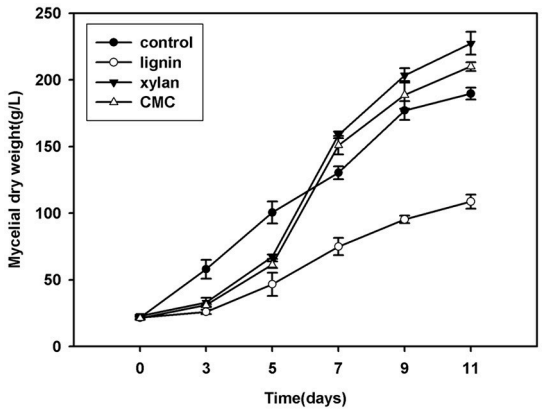


FIGURE 1 | Growth curve of mycelium (mycelium dry weight) in Krik's, lignin, cellulose, and xylan for 11 days.

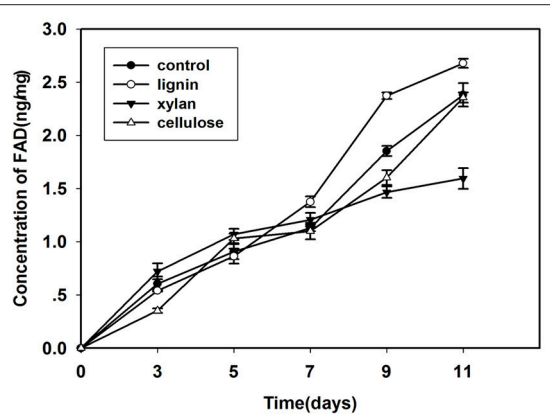


FIGURE 2 | Concentration of FAD in mycelium that cultured in control, lignin, xylan, and CMC for 11 days.

Differences between the Mycelial Proteomes during Growth in Lignocellulose and in the Control Medium

Three biological replicates for each mycelial protein of *P. ostreatus*, grown in Kirk's medium and in Kirk's medium supplemented with lignin, were separated by 2-DE. Total 531 ± 23, 496 ± 19, 567 ± 38, and 601 ± 27 protein spots were detected in the control, lignin, xylan, and CMC conditions, respectively (**Figure 4**). Proteins that were differentially expressed under various culture conditions were divided into categories according to their molecular functions and involvement in biological processes, based on the JGI database and GO (http://geneontology.org/) classification system (**Table 1**, **Figures 5**, **6**). For proteins lacking exact functional annotations in this database, we used family and domain databases (Inter Pro and Pfam) to reveal annotations of their conserved domains. Identified proteins included those involving in (i) redox processes and (ii) stress response. The stress-response group included anti-oxidation proteins and proteins involving in the response to toxic stress that are considered to play a role in the protection of cells from damage. The intensity of four spots (6, 7, 19, 20) for proteins involving in the stress response and three spots (58, 85, 112) for proteins involving in redox processes show a significant increase in all fungi grown in the three substrates relative to the control. The identified proteins also included proteins involving in (iii) carbohydrate metabolism and energy metabolism; these proteins are involving in the conversion of carbohydrates into energy to support cell processes. **Figure 5** show that the intensity of 15 spots representing proteins involving in this process was significantly decreased for fungi grown on lignin, whereas five spots representative of proteins related to carbohydrate metabolism exhibited an increase for those grown on xylan and CMC. The identified proteins also included proteins involving in (iv) protein and amino acid synthesis, (v) nucleotide metabolism, and (vi) others. Proteins in the "others" group were related to other types of metabolism or considered to have unknown functions.

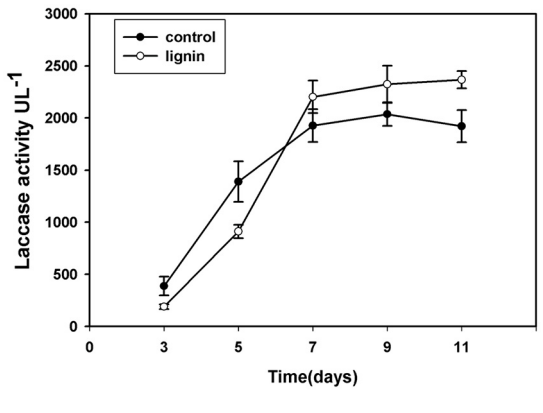


FIGURE 3 | Changes of laccase activity in Kirk's liquid medium supplemented with lignin and without lignin (control) after inoculation for 11 days.

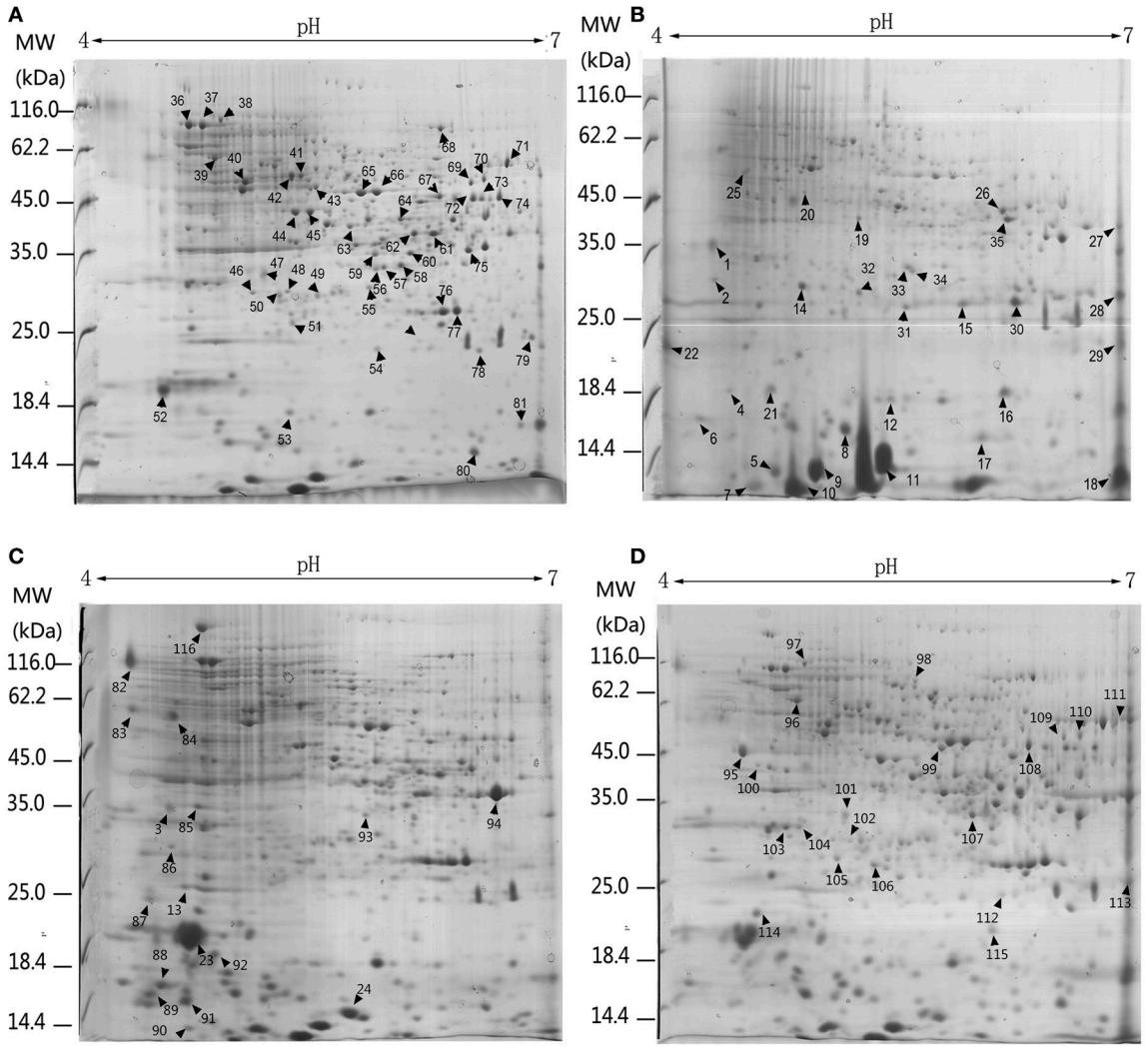


FIGURE 4 | 2-DE analysis of differential expressed proteins in *P. ostreatus* grown in liquid substrates supplemented with different components of lignocellulose. Arrows and numbers refer to differential expressed proteins **(A)** Kirk's liquid medium (control). **(B)** Kirk's liquid medium supplemented with lignin; **(C)** xylan; **(D)** CMC. The pH of the isoelectric focusing gel ranging from 4 to 7 is shown on the top of each gel. Protein ladder (molecular weight, MW) is shown on the left of the gels.

Lignin-Responsive Proteins

Base on the result of proteomics, the intensity of 36 spots was found to be significantly increased (fold $>$ 2) and that of 71 spots significantly decreased (fold $<$ 0.5; **Table 1**, **Figure 6**). Eight spots only increased or detected for fungus grown on lignin, whereas the intensity of spot 9 (oxidation-resistance protein), spot 11 (10-kDa heat shock protein), spot 28 (superoxide dismutase [Cu-Zn]), spot 29 (14-3-3 protein), and spot 30 (glutathione-S-transferase), representing proteins involving in the stress response in lignin, was 3.5-, 2.5-, 2.6-, 2.4-, and 2.5-fold higher than that of the control, respectively. Among proteins related to the redox process, the intensity of spot 10 (cytochrome c oxidase copper chaperone) increased by 2.5-fold in lignin compared to that in the control, whereas spot 22 (putative oxidoreductase) was only detected for fungus grown on lignin. Notably, spot 30 and spot 62 both corresponded to glutathione-s-transferase; however, spot 62 was not detected for the lignin group, probably because subunits of the same protein would separate during the focusing process. The intensity of 26 proteins related to carbohydrate metabolism was significantly decreased for the lignin group. Most of these proteins participate in six types of carbohydrate metabolism. Interestingly, the intensity of the carbohydrate metabolism-related protein adenylate kinase (spot 15) was 6.3-fold higher than that in the control.

Polysaccharide-Responsive Proteins

Xylan and cellulose, which are the main polysaccharides present in lignocellulose, are the primary carbon sources for fungi. In this study, CMC was used as a substitute for cellulose to study

TABLE 1 | List of proteins identified by ESI-MS/MS from *P. ostreatus* growing in lignocellulose.

Spot no.	Biological process	JGI ID	Description	MW(kDa)/pI	Mascot score	Fold change(treat/control)		
						Lignin	Xylan	CMC
3	Carbohydrate metabolism and energy metabolism	jgi|1043453	Mannose-6-phosphate isomerase	32.9/3.91	98	a	b	a
99		jgi|1036636	Phosphatidylserine decarboxylase proenzyme 2	54.4/6.19	85	0.4	0.64	2.18
109		jgi|1070334	Pyruvate kinase 1	61.6/7.47	151	0.25	0.75	1.53
100		jgi|1011623	Phosphoglycerate kinase	41.8/3.67	169	0.95	a	1.95
76		jgi|1109049	Triosephosphate isomerase	27.4/6.85	109	0.48	2.03	2.67
106		jgi|1070334	Pyruvate kinase 1	45.6/7.69	84	0.47	1.53	3.16
95		jgi|1032066	Phosphoglycerate kinase	45.0/3.44	115	a	0.49	4.25
86		jgi|1019376	Xylulose kinase	28.9/4.01	87	a	b	a
93		jgi|1102061	NADPH-dependent D-xylose reductase	34.7/6.69	99	1.22	6.41	3.1
96		jgi|1114405	6-phosphogluconate dehydrogenase, decarboxylating	62.3/3.98	75	0.3	0.37	1.54
55		jgi|1068318	Ribose-5-phosphate isomerase	31.4/6.08	90	0.34	1.59	a
69		jgi|1109004	6-phosphogluconate dehydrogenase	53.7/6.23	147	0.14	0.17	0.69
81		jgi|1108563	Glucose-6-phosphate dehydrogenase	16.4/7.89	106	0.42	a	1.14
75		jgi|1090672	Glyceraldehyde-3-phosphate dehydrogenase	34.6/6.63	135	0.44	0.72	1.51
77		jgi|1037028	Phosphogluconate dehydrogenase	26.6/7.02	126	0.11	2.62	3.55
54		jgi|1101876	Ribulose-phosphate 3-epimerase	25.9/5.5	219	0.13	1.17	1.26
38	Carbohydrate metabolism and energy metabolism	jgi|1018327	Glycosyl transferase family 4	81.8/6.11	76	0.514	0.88	1.13
71		jgi|1096444	Glucose-6-phosphate 1-dehydrogenase	58.5/6.55	140	a	0.14	3.85
70		jgi|1064981	Glucose-1-phosphate uridylyltransferase	58.7/6.26	154	0.43	0.44	1.23
72		jgi|48499	Pyruvate dehydrogenase	44.9/7.7	163	0.41	0.57	0.67
73		jgi|48714	Isocitrate dehydrogenase	47.5/5.93	209	0.13	1.11	0.72
56		jgi|1094663	Malate dehydrogenase	34.2/6.12	139	0.3	0.35	1.23
60		jgi|1075656	Malate dehydrogenase	34.1/6.09	91	0.34	2.06	1.87
23		jgi|1097340	Pyruvate carboxylase 2	18.0/4.34	87	a	7.57	2.16
78		jgi|1113799	Rhamnogalacturonan acetylesterase	25.6/6.02	181	a	a	a
94		jgi|1082594	Carbon catabolite-derepressing protein kinase	35.0/7.45	89	a	4.82	0.96
36		jgi|1053961	ATPase	100.3/5.27	239	0.23	2.58	1.15
40		jgi|1044485	Hydrogen-transporting ATPase	57.0/5.34	204	a	0.77	1.25
115		jgi|1089099	Glutamine synthetase	22.8/6.98	148	a	1.3	2.13
15		jgi|1087999	Adenylate kinase	27.7/6.77	104	6.29	2.33	2.26
13	Nucleotide metabolism	jgi|1105829	GDP-mannose transporter	23.7/4.26	92	a	b	a
88		jgi|1099408	Suppressor of kinetochore protein 1	17.4/3.63	76	a	b	a
84		jgi|47938	Alpha-1,3/1,6-mannosyltransferase ALG2	49.7/4.09	64	0.45	2.35	a

(Continued)

TABLE 1 | Continued

Spot no.	Biological process	JGI ID	Description	MW(kDa)/pI	Mascot score	Fold change(treat/control)		
						Lignin	Xylan	CMC
79	Nucleotide metabolism	jgi|1037108	Scavenger mRNA decapping enzyme	25.4/6.11	204	0.19	a	a
46		jgi|1054232	Ribonuclease T2	41.5/5.92	198	0.17	0.27	0.62
47		jgi|1083505	40S ribosomal protein	32.1/5.21	72	0.38	0.15	0.94
41		jgi|1056351	cysteine-type endopeptidase	55.0/5.81	88	a	a	1.19
44		jgi|1088444	GTP binding	42.0/5.11	109	a	1.31	1.23
57		jgi|1037683	Ribosomal	33.6/5.99	95	0.3	0.92	1.23
68		jgi|1054296	GTPase	93.4/6.27	136	0.28	0.22	1.4
103		jgi|1090777	Glucosamine 6-phosphate N-acetyltransferase	30.9/3.91	99	0.45	0.45	1.55
59		jgi|1095212	Endo/exonuclease	34.7/5.25	92	0.33	0.76	1.76
113		jgi|1106249	Tethering factor for nuclear proteasome sts1	24.0/8.28	88	2.23	1.28	2.16
101		jgi|1075990	cAMP-dependent protein kinase regulatory subunit	34.1/6.25	78	a	1.63	2.19
108		jgi|1101333	40S ribosomal protein S29	45.2/7.36	144	0.46	0.61	2.63
24		jgi|1066340	60S ribosomal protein L43	14.7/4.98	106	2.15	7.78	3.34
27		jgi|1047882	Dimethyl adenosine transferase	36.1/9.0	79	2.32	0.464	4.64
90		jgi|185993	60S ribosomal protein L23-B	14.3/4.26	83	b	b	b
89	Other metabolism	jgi|1114368	Calmodulin	31.3/3.57	100	a	b	a
63		jgi|52279	Formamidase	42.9/5.22	137	0.3	0.67	0.93
39		jgi|1098883	Amidase	55.7/5.09	188	0.27	a	1.12
64		jgi|1093313	Delta-aminolevulinic acid dehydratase	35.8/5.79	151	a	1.03	1.52
65		jgi|1054502	Enolase	47.1/5.55	157	0.15	1.17	1.74
66		jgi|1107810	CoA- transferase	41.8/5.93	218	0.28	1.21	1.99
97		jgi|155235	Mitochondrial import inner membrane translocase	115.6/4.7	100	a	0.44	2.26
61		jgi|1067896	Putative hydrolase	40.5/6.32	157	0.47	1.74	2.65
25		jgi|1065645	Putative Sugar transporter	57.5/3.97	83	2.31	21.19	4.62
114		jgi|1094266	Diphosphoinositol polyphosphate phosphohydrolase	20.7/4.77	82	1.76	4.91	4.86
35		jgi|1107810	Putative CoA-transferase	41.9/5.93	256	4.29	4.78	8.18
102		jgi|1101425	Sterol 3-beta-glucosyltransferase	32.6/6.24	86	b	b	b
83	Protein and amino acid synthesis	jgi|1089644	Leucine carboxyl methyltransferase	50.0/3.47	76	a	b	a
45		jgi|1076233	3-isopropylmalate dehydrogenase	40.7/5.56	182	a	0.39	0.67
48		jgi|176309	Glucosamine-6-phosphate isomerase	32.8/5.81	136	0.47	a	0.68
74		jgi|1094925	Adenosylhomocysteinase	47.3/5.93	141	a	1.3	0.85
111		jgi|1010364	SWI5-dependent HO expression protein 3	60.2/7.74	95	a	a	0.88
37		jgi|1091537	Alanine–tRNA ligase	106.3/5.52	116	a	2.54	1.21
49		jgi|1099302	Methyltransferase	31.0/5.7	206	0.37	a	1.23
67		jgi|1060865	Aspartate aminotransferase	45.4/6.54	176	0.21	0.37	1.28

(Continued)

TABLE 1 | Continued

Spot no.	Biological process	JGI ID	Description	MW(kDa)/pI	Mascot score	Fold change(treat/control)		
						Lignin	Xylan	CMC
107		jgi|1112899	Phenylalanine ammonia-lyase	33.4/7.08	96	a	0.75	1.59
104		jgi|1027245	E3 ubiquitin-protein ligase TOM1	31.6/4.64	102	1.21	1.18	2.17
110		jgi|1039251	Serine/threonine-protein kinase MEC1	61.4/7.65	79	2	0.53	2.57
98		jgi|48252	Glycylpeptide N-tetradecanoyltransferase	85.2/6.66	80	a	a	b
22	Redox processes	jgi|1066477	Putative oxidoreductase	22.8/3.22	72	b	a	a
10		jgi|1027050	Cytochrome c oxidase copper chaperone	13.9/4.93	74	2.55	0.51	0.26
112		jgi|1098138	Aldehyde dehydrogenase	21.9/6.37	94	0.37	0.32	1.49
34		jgi|1045076	Putative oxidoreductase	32.8/5.56	147	2.24	2.24	1.79
58		jgi|1081301	Putative Glucose/ribitol dehydrogenase	33.1/6.17	184	0.42	1.42	1.89
105		jgi|1029362	Isocitrate dehydrogenase [NAD] subunit 1	28.3/6.19	90	1.68	1.63	2.1
31		jgi|1086766	Putative oxidoreductase	26.3/5.69	146	4.24	2.97	2.24
2		jgi|19749	Cytochrome c oxidase assembly protein	31.1/3.55	108	2.98	2.19	2.32
8		jgi|1039979	3-isopropylmalate dehydrogenase	15.5/6.24	82	3.26	6.52	2.6
85		jgi|1098138	Potassium-activated aldehyde dehydrogenase	35.4/4.65	63	a	3.38	3.24
21		jgi|1015418	3-hydroxyanthranilate 3,4-dioxygenase	18.6/4.14	80	2.67	4.22	3.76
14		jgi|1030848	Putative aryl-alcohol dehydrogenase	31.4/4.95	95	9.25	11.13	4.63
33		jgi|1076970	Putative oxidoreductase	32.3/5.5	133	b	b	b
5		jgi|1015750	Cytochrome c oxidase assembly protein 3	14.6/4.65	88	a	a	a
16		jgi|1087944	Putative nitronate monooxygenase	19.2/7.34	69	b	b	b
29	Stress response	jgi|1064479	14-3-3 protein homolog	23.0/8.26	109	2.37	a	a
30		jgi|1077356	Glutathione-S-Trfase	26.9/5.77	131	2.53	0.52	0.78
9		jgi|1077250	Oxidation resistance protein 1	14.4/5.76	105	3.52	0.69	0.84
11		jgi|185767	10 kDa heat shock protein, mitochondrial	14.4/5.76	64	4.23	1.26	0.85
28		jgi|1113505	Superoxide dismutase [Cu-Zn]	28.1/9.0	100	2.6	0.52	1.38
87		jgi|1022101	Thiamine thiazole synthase	22.8/3.52	79	0.42	2.19	1.66
62		jgi|1058013	Glutathione-S-Trfase_C-like	42.1/6.0	98	a	1.38	2.56
19		jgi|1108100	Inheritance of peroxisomes protein 1	40.0/5.42	71	5.78	4.56	4.79
6		jgi|1032742	Oxidant-induced cell-cycle arrest protein 5	16.6/3.58	97	b	b	b
7		jgi|1025559	Monothiol glutaredoxin-5	13.5/4.34	100	b	b	b
20		jgi|1106868	Alternative oxidase, mitochondrial	45.4/4.61	90	b	b	b
91	Unkown function	jgi|1048439	Uncharacterized protein C6B12.14c	15.4/4.17	111	a	b	0
50		jgi|1080037	PLC-like phosphodiesterase	32.7/5.0	89	0.14	0.2	0.41
26		jgi|1022166	Purine phosphorylase	43.7/7.33	120	2.32	2.78	0.7
51		jgi|1097135	Calcium ion binding protein	25.4/5.32	194	a	0.7	1.23
43		jgi|1048986	Predicate protein	46.5/5.34	182	a	a	1.24

(Continued)

TABLE 1 | Continued

Spot no.	Biological process	JGI ID	Description	MW(kDa)/pI	Mascot score	Fold change(treat/control)		
						Lignin	Xylan	CMC
80		jgi|1016239	Putative protein	15.2/5.95	122	a	1.14	1.29
53		jgi|1081199	Putative protein	16.8/5.71	94	a	0.93	1.33
18		jgi|1046446	Hypothetical protein	14.7/9.0	84	5.44	1.41	1.47
1		jgi|1075945	Hypothetical protein	34.9/3.58	69	4.47	2.77	1.52
17		jgi|1012191	Hypothetical protein	15.4/7.66	99	3.26	4.08	3.98
32		jgi|1107762	Hypothetical protein	30.0/5.38	192	9.22	1.84	4.61
92		jgi|176718	Putative uncharacterized protein YKL131W	18.8/4.76	141	1.18	5.54	6.36
52		jgi|1033638	Putative protein	21.1/4.68	91	0.48	26.11	11.14
82		jgi|1048248	Bifunctional lycopene cyclase/phytoene synthase	114.5/3.34	96	a	b	b
4		jgi|166436	Hypothetical protein	18.7/3.64	98	b	b	b
12		jgi|174389	Hypothetical protein	17.1/5.76	104	b	b	b

a, particularly low; b, particularly high.

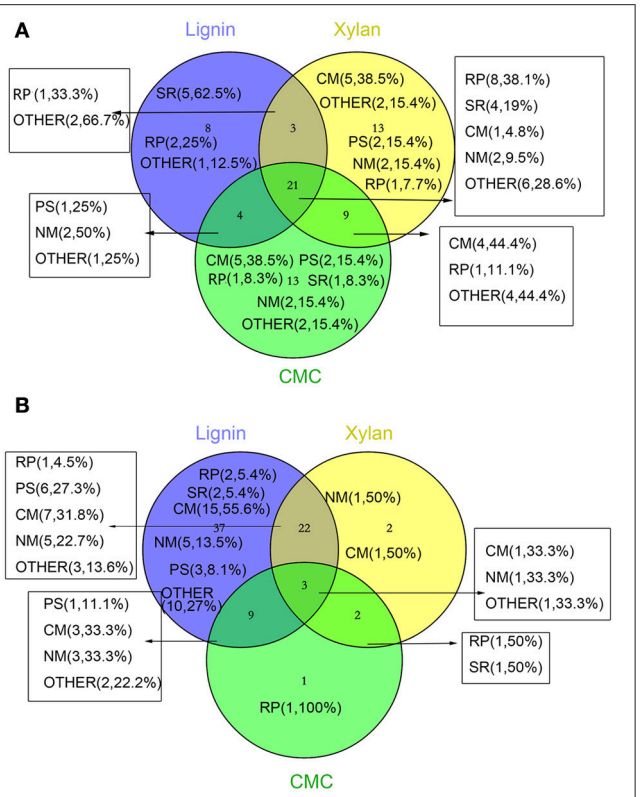


FIGURE 5 | Venn diagram representing the distribution of number and function of validated and significantly changed proteins according to proteome. The numbers in parentheses indicate the amount and percentage (the percentage of the proteins in increased or decreased proteins in different treatments) of protein in this section. **(A)** increased proteins **(B)** decreased proteins. Abbreviations refer to different metabolic processes: RP, redox process; NM, nucleotide metabolism; PS, protein and amino acid synthesis; SR, stress response; CM, Carbohydrate metabolism and energy metabolism; OTHER, other metabolism and unkown function.

the effect of cellulose on *P. ostreatus*. Differentially expressed proteins displayed similar expression patterns in xylan and CMC; for both substrates, most proteins showing an increase in abundance were associated with carbohydrate metabolism. Ten carbohydrate metabolism-related proteins showed higher abundance in the two substrates than in the control. **Table 1** show that the intensity of spot 95 (phosphoglycerate kinase) and spot 106 (pyruvate kinase), which represented proteins involving in the glycolysis/gluconeogenesis pathway, was 4.3- and 3.2-fold higher in the CMC group than in the control. Spot 71 (glucose-6-phosphate 1-dehydrogenase) and spot 77 (phosphogluconate dehydrogenase), which represented proteins involving in the pentose phosphate pathway, had 3.9- and 3.6-fold higher abundance in the CMC group than in the control. However, these spots showed lower abundance in the xylan group. In addition, the intensity of spot 93 (NADPH-dependent D-xylose reductase) was 6.4-fold higher in the xylan group and 3.1-fold higher in the CMC group than in the control. Spot 86, which was identified as a xylulose kinase, was only detected in the xylan group. D-xylose reductase and xylulose kinase are

both involving in the pentose and glucuronate interconversion pathway. In other species, these two proteins are involving in xylan degradation and energy release. The intensity of spot 23 (pyruvate carboxylase) was 7.6-fold higher in the xylan group than in the control group, but only 2.2-fold higher in the CMC group than in the control group.

DISCUSSION

Lignocellulose is the main substrate used for cultivation of edible fungi. Hemicellulose and cellulose are carbon sources for fungal growth; however, another main component of lignocellulose, lignin, affects the degradation of fiber by fungi. The presence of lignin limits the access of cellulotytic enzymes to cellulose, that may influence the efficiency of enzymatic hydrolysis of cellulose and hemicellulose (Kumar et al., 2012). This effect is not observed in white-rot fungus, in which lignin is degraded by the extracellular oxidative system. However, the growth of this fungus is affected by a series of lignin derivatives; previous studies have shown that various lignin-related para-phenolic benzoic acids, para-phenolic cinnamic acids, and para-phenolic phenylpropionic acids elicit increased inhibition of growth in white-rot fungus (Buswell and Eriksson, 1994). In addition, higher concentrations of aromatic aldehydes were shown to be more toxic than the corresponding carboxylic acid (Dekker et al., 2002). These findings are consistent with those of the previous work showing that the growth of *P. ostreatus* is inhibited by lignin (Barakat et al., 2012). In the present study, although the fungus was still able to grow on lignin, the relative growth rate increased 7 days after inoculation. The rapid growth of mycelia in the control group was presumably related to the rapid consumption of nutrients. An alternative explanation for this observation is that the fungus began to adapt to the lignin-based medium. To date, little is known about the effects of lignin on mycelial growth and the stress response in fungi.

Lignin degradation is an extracellular oxidative process, and the production of H_2O_2 is temporally related to lignin degradation (Achyuthan et al., 2010). *P. ostreatus* has a range of extracellular enzymes that generate H_2O_2 for utilization by ligninolytic enzymes (Akpinar and Urek, 2014). Superoxide dismutase, ascorbate peroxidases, and glutathione reductase are key enzymes involving in reducing H_2O_2 in the ascorbate-glutathione cycle in cells (Yousuf et al., 2012; Choudhury et al., 2013; Yang et al., 2013). These proteins, which are induced in response to numerous environmental stresses, mediate the detoxification of reactive oxygen species. The enzymes related to the oxidative stress response were more abundant in the lignin condition, indicating a better response to H_2O_2 in the mycelium of *P. ostreatus* when compared to that in other culture conditions. These proteins, which are expressed in response to increased concentrations of extracellular H_2O_2, scavenge excess intracellular reactive oxygen species to protect cells from oxidative damage.

Inhibition of the transformation of carbon sources is another effect of oxidative stress on *P. ostreatus* (Filomeni et al., 2015). In the present study, most proteins involving in carbohydrate and energy metabolism were less abundant in the lignin group. This suggests that the inhibition of energy metabolism in response to lignin restricts mycelial growth. In the present study, as the adaptability of fungi to lignin increased, this restriction was gradually lifted, allowing slow accumulation of mycelial biomass to occur.

Recent research has suggested that laccase may play an important role in the fungal defense against oxidative stress, which acts as an element of the stress response (Giardina et al., 2010). It has been observed that oxidative stress induces the expression of ligninolytic enzymes in some basidiomycetes (Viswanath et al., 2014). In our study, the activity of laccase increased with time in the lignin group, and the increase in laccase expression appeared to increase the resistance of *P. ostreatus* to oxidative stress. The increase in laccase activity was therefore considered to enhance the adaptability of *P. ostreatus* to lignin in a gradual manner.

Interestingly, we found that the intensity of a 14-3-3 protein was significantly increased in the lignin group. The 14-3-3 proteins, which are rarely reported in fungi, are known to be upregulated in plants in response to pathogenic fungi. Previous studies have suggested that 14-3-3 proteins may control a negative feedback loop to prevent harmful overactivation of defense responses in plants (Lozano-Durán and Robatzek, 2015). Our results suggest a prominent role for 14-3-3 proteins in the fungal response to stress; however, it is not clear how lignin regulates the expression of this protein. The question of whether the expression of this protein relates to lignin needs further study.

The present results elucidate the relationship of the expression of antioxidative intracellular proteins and laccase with the defense response to exogenous H_2O_2—induced oxidative stress in fungi grown on lignin (Strong and Claus, 2011). Although the expression of these proteins promoted the adaptability of *P. ostreatus* to lignin, it is possible that alternative stress response mechanisms may additionally be associated with adaptation to growth in such environments.

Cellulose and hemicellulose in lignocellulose are the main nutrient sources for *P. ostreatus*. In fungi, the cAMP–PKA and TOR pathways respond to carbon and nitrogen signals to regulate a myriad of functions, including protein synthesis, ribosome biogenesis, autophagy, polarized cellular growth, cell-cycle progression, and filamentation (Liu et al., 1993). TOR signaling activates the expression of genes required for ribosome biogenesis, including those encoding ribosomal proteins, ribosomal RNA (rRNA), and tRNA (Dobrenel et al., 2016). Our findings additionally showed that cAMP-dependent protein kinase and three ribosomal proteins involving in sugar sensing were significantly upregulated in fungi grown on xylan and CMC. Furthermore, xylan and CMC regulate the adaptation of the fungus to the environment via their signaling pathways. Therefore, after inoculation for 7 days, the mycelial growth rate was observed to increase rapidly.

In a previous study, sensing of glucose as the preferred carbohydrate source was extensively studied in the yeast model organism (Braunsdorf et al., 2016). In the presence of glucose, genes required for growth on alternative carbon sources are repressed (Bahn et al., 2007). For *P. ostreatus*, the natural growth

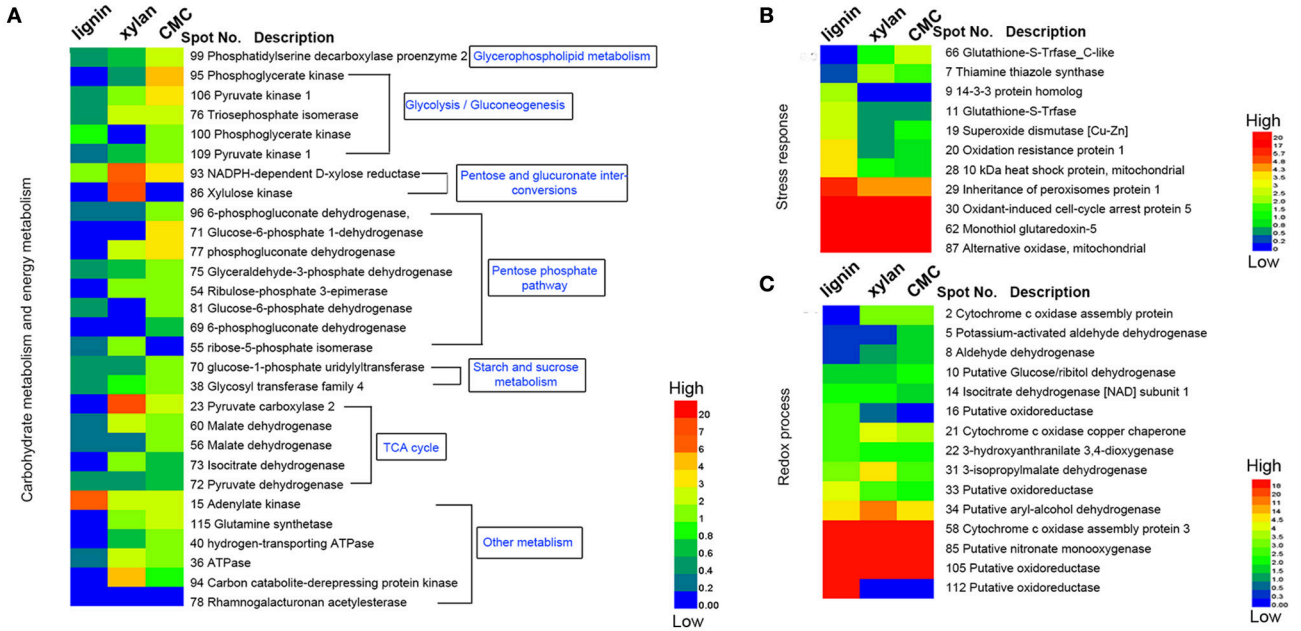


FIGURE 6 | Heatmap of the fold changes of differential proteins related to carbohydrate metabolism, stress response and redox process. (A) Fold change of proteins related to carbohydrate metabolism and energy metabolism. **(B)** Fold change of proteins response to stress. **(C)** Fold change of proteins related to redox process. The data are presented in matrix format in which rows represent the individual proteins and the columns represent each culture. Each cell in the matrix represents the fold change of a protein at an individual substrate compared with control. The red and green colors in cells reflect low and high change fold, respectively.

environment lacks glucose; accordingly, this fungus has evolved an effective method for regulation of natural polysaccharides. Various filamentous fungi, including *Neurospora crassa,* are capable of growth on pentose (Li et al., 2014). The genomes of pentose-utilizing fungi are a useful resource for mining novel gene elements, such as D-xylose transporters for metabolic engineering in *S. cerevisiae*. The xylose metabolism pathway consists of three enzymes, namely xylose reductase, xylitol dehydrogenase, and xylulokinase, which have been studied in relation to the metabolic engineering of *S. cerevisiae* for xylose fermentation (Farwick et al., 2014). This has been a subject of great interest over the past decade, as xylose is easier to obtain in nature (Li et al., 2015a). Despite these endeavors to improve xylose fermentation, the yields and productivity for ethanol obtained from xylose, using engineered *S. cerevisiae*, are much lower than those for ethanol obtained by glucose fermentation (Kurosawa et al., 2013). The high intensity of D-xylose reductase and xylulose kinase in *P. ostreatus* grown on xylan may be related to increased xylose metabolism under xylan regulation. However, this is not the only carbon metabolism pathway that is enhanced under xylan regulation; the expression of malate dehydrogenase, pyruvate carboxylase, ATPase, and adenylate kinase, which are involving in TCA metabolism, is also increased on xylan. The enhancement of xylose metabolism and other carbohydrate metabolism pathways greatly promotes the utilization of polysaccharides by *P. ostreatus*.

The hydrolysis product of CMC is glucose; therefore, the response mechanism of *P. ostreatus* for CMC is similar to that for glucose. Previous studies proved that GTPase activity may be indicative of the activation of signaling pathways in the presence of glucose as a carbon source, and almost half of the identified signaling-related proteins are G-protein coupled receptors or small GTPases (Post and Brown, 1996; Gancedo, 2008). GTPases are present at high levels in CMC, suggesting that it activates this signaling pathway. Addition of glucose to cells growing on non-fermentable carbon sources, or to stationary-phase cells, triggers a wide variety of regulatory processes directed toward the exclusive and optimal utilization of the preferred carbon source (Gancedo, 2008). Pyruvate kinase, phosphoglycerate kinase, triosephosphate isomerase, and phosphoglycerate kinase are upregulated in fungi growing on CMC, suggesting that glycolysis is activated by glucose. When glucose influx and utilization through glycolysis are stimulated, gluconeogenesis is inhibited, and there is a drastic increase in growth rate, which is preceded by a characteristic upshift in ribosomal RNA and protein synthesis.

Sugars such as xylan and cellulose are the primary fuel for most fungi (de Souza et al., 2014). The amount of available sugar may fluctuate widely, necessitating a mechanism for sensing available amounts and responding appropriately. In most organisms, this response involves changes in gene expression. Studies of the yeast glucose repression system have provided novel insights into the signaling pathway that responds to sugar. When yeast cells growing on high levels of sugar obtain most of their energy via fermentation, large amounts of sugar are metabolized through glycolysis (Johnston, 1999; Kim et al., 2013). Our findings suggest that addition of CMC and xylan to the medium significantly enhances the ability of *P. ostreatus* to transform sugars via different metabolic

pathways, and improves the adaptability of *P. ostreatus* to the environment.

Alcohol oxidation is critical for lignocellulose degradation. In our study, aryl-alcohol dehydrogenase enzymes showed higher abundance in all of the lignocellulose substrates. Moreover, aryl-alcohol dehydrogenase coupled with NADPH as a co-factor constitutes a redox system involving in aryl-alcohol/aryl-aldehyde production in the fungus that ensures steady availability of H_2O_2 for ligninolytic activities (Yang et al., 2012). Recent studies have shown that aryl-alcohol oxidases and dehydrogenase are induced by lignin derivatives and are involving in their metabolism *in vitro* (Feldman et al., 2015). Our results suggest that aryl-alcohol dehydrogenase is induced by lignin as well as lignocellulosic polysaccharides, and regulated by lignocellulose.

Flavin-containing oxidases catalyze a wide variety of different oxidation reactions; in the last decade, many flavoprotein oxidases with varied substrate specificities and reactivities have been discovered (Dijkman et al., 2013). Glucose oxidase, the best-known flavoprotein, is involving in lignocellulose degradation (Hernández-Ortega et al., 2012). To date, few studies have focused on the correlation between flavoprotein and lignocellulose degradation in cells. The only flavoproteins known to be involving in this process are the flavin-containing monooxygenases, which are widely distributed within living organisms and involving in various biological processes such as the detoxification of drugs, biodegradation of environmental aromatic compounds, and biosynthesis of antibiotics (Nakamura et al., 2012). In our study, the level of FAD increased with time; the level of FAD in fungus grown on lignocellulose was higher than that in fungus grown in the control medium, and highest in fungus grown on lignin. This indicates that the expression of FAD is regulated by lignocellulose, and that flavoprotein in cells plays an important role in the response to lignocellulose. Although it was not possible to determine which proteins are specifically regulated by lignocellulose, our findings provide novel insights into the roles of intracellular flavoproteins in the response to lignocellulose.

Some studies have shown that *P. ostreatus* selectively degrades hemicellulose when cultured with solid biomass (Ander and Eriksson, 1977; Chandra et al., 2007). This implies that *P. ostreatus* favors the use of hemicellulose as a carbon source. In our study, xylan had a certain effect on the accumulation of mycelial biomass, and we believe that xylan plays a key role in the regulation of genes related to the metabolism of xylulose. We speculate that the selective degradation of hemicellulose when *P. ostreatus* is cultured in solid biomass occurs because xylan is a carbon source that is beneficial for the growth of *P. ostreatus* (Dwivedi et al., 2011), and xylan activates the expression of genes in the xylose-related metabolic pathway, which allows *P. ostreatus* to use hemicellulose as a carbon source. There are some reports that lignin in natural lignocellulose limits the growth of fungi, that because of the structural limitation of mycelial invasion and the use of other polysaccharides (Sattler and Funnell-Harris, 2013). Our results suggest that this restriction may also be due to the inhibition of mycelial growth by lignin and the effect of lignin on carbon metabolism in *P. ostreatus* hyphae. Our results provide further understanding of the solid-state culture of *P. ostreatus*.

Elucidation of lignocellulose–fungal interactions is important for understanding fungal ecology and for the maintenance of the delicate balance of fungal symbionts in our ecosystem. Understanding the mechanism of the fungal response to lignocellulose will facilitate its application in metabolic engineering of biotechnology to optimize the bioconversion of biomass resources in the future.

AUTHOR CONTRIBUTIONS

XZ and CL designed the experiments. QX, FM, and HY wrote the manuscript. QX conducted most of the experimental work and performed analysis of data. YL assisted with experiments. All authors discussed the results and reviewed the manuscript.

FUNDING

This research was supported by the National Basic Research Program of China (2014CB138303), the High-tech Research and Development Program of China (2012AA101805) and the National Natural Science Foundation of China (J1103514).

REFERENCES

Achyuthan, K. E., Achyuthan, A. M., Adams, P. D., Dirk, S. M., Harper, J. C., Simmons, B. A., et al. (2010). Supramolecular self-assembled chaos: polyphenolic lignin's barrier to cost-effective lignocellulosic biofuels. *Molecules* 15, 8641–8688. doi: 10.3390/molecules15118641

Akai, S. (2012). "Histology of defense in plants," in *Plant Pathology, An Advanced Treatise,* eds J. Horsfall and E. B. Cowling (New York, NY, Springer), 391–434.

Akpinar, M., and Urek, R. O. (2014). Extracellular ligninolytic enzymes production by *Pleurotus eryngii* on agroindustrial wastes. *Prep. Biochem. Biotechnol.* 44, 772–781. doi: 10.1080/10826068.2013.867870

Ander, P., and Eriksson, K. E. (1977). Selective degradation of wood components by white-rot fungi. *Physiol. Plant.* 41, 239–248. doi: 10.1111/j.1399-3054.1977.tb04877.x

Bahn, Y.-S., Xue, C., Idnurm, A., Rutherford, J. C., Heitman, J., and Cardenas, M. E. (2007). Sensing the environment: lessons from fungi. *Nat. Rev. Microbiol.* 5, 57–69. doi: 10.1038/nrmicro1578

Barakat, A., Monlau, F., Steyer, J.-P., and Carrere, H. (2012). Effect of lignin-derived and furan compounds found in lignocellulosic hydrolysates on biomethane production. *Bioresour. Technol.* 104, 90–99. doi: 10.1016/j.biortech.2011.10.060

Braunsdorf, C., Mailänder-Sánchez, D., and Schaller, M. (2016). Fungal sensing of host environment. *Cell. Microbiol.* 18, 1188–1200. doi: 10.1111/cmi.12610

Buswell, J., and Eriksson, K.-E. (1994). Effect of lignin-related phenols and their methylated derivatives on the growth of eight white-rot fungi. *World J. Microbiol. Biotechnol.* 10, 169–174. doi: 10.1007/BF00360880

Chandra, R. P., Bura, R., Mabee, W. E., Berlin, D. A., Pan, X., and Saddler, J. N. (2007). Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? *Adv. Biochem. Eng. Biotechnol.* 108, 67–93. doi: 10.1007/10_2007_064

Choudhury, S., Panda, P., Sahoo, L., and Panda, S. K. (2013). Reactive oxygen species signaling in plants under abiotic stress. *Plant Signal. Behav.* 8, 1559–2324. doi: 10.4161/psb.23681

de Souza, A. P., Grandis, A., Leite, D. C., and Buckeridge, M. S. (2014). Sugarcane as a bioenergy source: history, performance, and perspectives for second-generation bioethanol. *Bioenergy Res.* 7, 24–35. doi: 10.1007/s12155-013-9366-8

Dekker, R. F., Barbosa, A. M., and Sargent, K. (2002). The effect of lignin-related compounds on the growth and production of laccases by the ascomycete, *Botryosphaeria* sp. *Enzyme Microb. Technol.* 30, 374–380. doi: 10.1016/S0141-0229(01)00503-8

Dijkman, W. P., de Gonzalo, G., Mattevi, A., and Fraaije, M. W. (2013). Flavoprotein oxidases: classification and applications. *Appl. Microbiol. Biotechnol.* 97, 5177–5188. doi: 10.1007/s00253-013-4925-7

Dobrenel, T., Caldana, C., Hanson, J., Robaglia, C., Vincentz, M., Veit, B., et al. (2016). TOR Signaling and Nutrient Sensing. *Annu. Rev. Plant Biol.* 67, 261–285. doi: 10.1146/annurev-arplant-043014-114648

Dwivedi, P., Vivekanand, V., Pareek, N., Sharma, A., and Singh, R. P. (2011). Co-cultivation of mutant Penicillium oxalicum SAU E-3.510 and *Pleurotus ostreatus* for simultaneous biosynthesis of xylanase and laccase under solid-state fermentation. *New Biotechnol.* 28, 616–626. doi: 10.1016/j.nbt.2011.05.006

Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., and Boles, E. (2014). Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5159–5164. doi: 10.1073/pnas.1323464111

Feldman, D., Kowbel, D. J., Glass, N. L., Yarden, O., and Hadar, Y. (2015). Detoxification of 5-hydroxymethylfurfural by the *Pleurotus ostreatus* lignolytic enzymes aryl alcohol oxidase and dehydrogenase. *Biotechnol. Biofuels* 8, 1. doi: 10.1186/s13068-015-0244-9

Fernández, R. G., and Novo, J. V. J. (2013). "Proteomic protocols for the study of filamentous fungi," in *Laboratory Protocols in Fungal Biology*, eds V. K. Gupta, M. G. Tuohy, M. Ayyachamy, K. M. Turner, and A. O'Donovan (New York, NY: Springer), 299–308.

Fernández-Fueyo, E., Ruiz-Dueñas, F. J., López-Lucendo, M. F., Pérez-Boada, M., Rencoret, J., Gutiérrez, A., et al. (2016). A secretomic view of woody and nonwoody lignocellulose degradation by *Pleurotus ostreatus*. *Biotechnol. Biofuels* 9, 1. doi: 10.1186/s13068-016-0462-9

Filomeni, G., De Zio, D., and Cecconi, F. (2015). Oxidative stress and autophagy: the clash between damage and metabolic needs. *Cell Death Differ.* 22, 377–388. doi: 10.1038/cdd.2014.150

Gancedo, J. M. (2008). The early steps of glucose signalling in yeast. *FEMS Microbiol. Rev.* 32, 673–704. doi: 10.1111/j.1574-6976.2008.00117.x

Gaskell, J., Blanchette, R. A., Stewart, P. E., BonDurant, S. S., Adams, M., Sabat, G., et al. (2016). Transcriptome and secretome analyses of the wood decay fungus *Wolfiporia* cocos support alternative mechanisms of lignocellulose conversion. *Appl. Environ. Microb.* 82, 3979–3987. doi: 10.1128/AEM.00639-16

Giardina, P., Faraco, V., Pezzella, C., Piscitelli, A., Vanhulle, S., and Sannia, G. (2010). Laccases: a never-ending story. *Cell. Mol. Life Sci.* 67, 369–385. doi: 10.1007/s00018-009-0169-1

González-Fernández, R., Aloria, K., Valero-Galván, J., Redondo, I., Arizmendi, J. M., and Jorrín-Novo, J. V. (2014). Proteomic analysis of mycelium and secretome of different *Botrytis cinerea* wild-type strains. *J. Proteomics* 97, 195–221. doi: 10.1016/j.jprot.2013.06.022

Grinyer, J., Hunt, S., McKay, M., Herbert, B. R., and Nevalainen, H. (2005). Proteomic response of the biological control fungus *Trichoderma atroviride* to growth on the cell walls of *Rhizoctonia solani*. *Curr. Genet.* 47, 381–388. doi: 10.1007/s00294-005-0575-3

Hamdane, D., Bou-Nader, C., Cornu, D., Hui-Bon-Hoa, G., and Fontecave, M. (2015). Flavin–protein complexes: aromatic stacking assisted by a hydrogen bond. *Biochemistry* 54, 4354–4364. doi: 10.1021/acs.biochem.5b00501

Hernández-Ortega, A., Ferreira, P., and Martínez, A. T. (2012). Fungal aryl-alcohol oxidase: a peroxide-producing flavoenzyme involved in lignin degradation. *Appl. Microbiol. Biotechnol.* 93, 1395–1410. doi: 10.1007/s00253-011-3836-8

Jami, M. S., García-Estrada, C., Barreiro, C., Cuadrado, A. A., Salehi-Najafabadi, Z., and Martín, J. F. (2010). The *Penicillium chrysogenum* extracellular proteome. Conversion from a food-rotting strain to a versatile cell factory for white biotechnology. *Mol. Cell. Proteomics* 9, 2729–2744. doi: 10.1074/mcp.M110.001412

Johnston, M. (1999). Feasting, fasting and fermenting: glucose sensing in yeast and other cells. *Trends Genet.* 15, 29–33. doi: 10.1016/S0168-9525(98)01637-0

Kim, J.-H., Roy, A., Jouandot, D., and Cho, K. H. (2013). The glucose signaling network in yeast. *Biochim. Biophys. Acta* 1830, 5204–5210. doi: 10.1016/j.bbagen.2013.07.025

Kües, U. (2015). Fungal enzymes for environmental management. *Curr. Opin. Biotechnol.* 33, 268–278. doi: 10.1016/j.copbio.2015.03.006

Kumar, L., Arantes, V., Chandra, R., and Saddler, J. (2012). The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. *Bioresour. Technol.* 103, 201–208. doi: 10.1016/j.biortech.2011.09.091

Kurosawa, K., Wewetzer, S. J., and Sinskey, A. J. (2013). Engineering xylose metabolism in triacylglycerol-producing *Rhodococcus opacus* for lignocellulosic fuel production. *Biotechnol. Biofuels* 6, 1. doi: 10.1186/1754-6834-6-134

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 1. doi: 10.1186/1754-6834-6-41

Li, C., Zhao, X., Wang, A., Huber, G. W., and Zhang, T. (2015b). Catalytic transformation of lignin for the production of chemicals and fuels. *Chem. Rev.* 115, 11559–11624. doi: 10.1021/acs.chemrev.5b00155

Li, J., Lin, L., Li, H., Tian, C., and Ma, Y. (2014). Transcriptional comparison of the filamentous fungus *Neurospora crassa* growing on three major monosaccharides D-glucose, D-xylose and L-arabinose. *Biotechnol. Biofuels* 7, 1. doi: 10.1186/1754-6834-7-31

Li, X., Yu, V. Y., Lin, Y., Chomvong, K., Estrela, R., Park, A., et al. (2015a). Expanding xylose metabolism in yeast for plant cell wall conversion to biofuels. *Elife* 4:e05896. doi: 10.7554/eLife.05896

Liu, H., Styles, C. A., and Fink, G. R. (1993). Elements of the yeast pheromone response pathway required for filamentous growth of diploids. *Science* 262, 1741–1744.

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic. Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Lozano-Durán, R., and Robatzek, S. (2015). 14-3-3 proteins in plant-pathogen interactions. *Mol. Plant Microbe Interact.* 28, 511–518. doi: 10.1094/MPMI-10-14-0322-CR

Nakamura, T., Ichinose, H., and Wariishi, H. (2012). Flavin-containing monooxygenases from *Phanerochaete chrysosporium* responsible for fungal metabolism of phenolic compounds. *Biodegradation* 23, 343–350. doi: 10.1007/s10532-011-9521-x

Post, G., and Brown, J. H. (1996). G protein-coupled receptors and signaling pathways regulating growth responses. *FASEB J.* 10, 741–749.

Rabilloud, T., Chevallet, M., Luche, S., and Lelong, C. (2010). Two-dimensional gel electrophoresis in proteomics: past, present and future. *J. Proteomics* 73, 2064–2077. doi: 10.1016/j.jprot.2010.05.016

Sattler, S., and Funnell-Harris, D. (2013). Modifying lignin to improve bioenergy feedstocks: strengthening the barrier against pathogens? *Front. Plant Sci.* 4:70. doi: 10.3389/fpls.2013.00070

Simon, O., Klaiber, I., Huber, A., and Pfannstiel, J. (2014). Comprehensive proteome analysis of the response of *Pseudomonas putida* KT2440 to the flavor compound vanillin. *J. Proteomics* 109, 212–227. doi: 10.1016/j.jprot.2014.07.006

Sindhu, R., Binod, P., and Pandey, A. (2016). Biological pretreatment of lignocellulosic biomass–An overview. *Bioresour. Technol.* 199, 76–82. doi: 10.1016/j.biortech.2015.08.030

Srinivasan, C., Dsouza, T. M., Boominathan, K., and Reddy, C. A. (1995). Demonstration of laccase in the white rot basidiomycete *Phanerochaete chrysosporium* BKM-F1767. *Appl. Environ. Microb.* 61, 4274–4277.

Strong, P., and Claus, H. (2011). Laccase: a review of its past and its future in bioremediation. *Crit. Rev. Environ. Sci. Technol.* 41, 373–434. doi: 10.1080/10643380902945706

Taniguchi, M., Suzuki, H., Watanabe, D., Sakai, K., Hoshino, K., and Tanaka, T. (2005). Evaluation of pretreatment with *Pleurotus ostreatus* for enzymatic hydrolysis of rice straw. *J. Biosci. Bioeng.* 100, 637–643. doi: 10.1263/jbb.100.637

Taniwaki, M. H., Pitt, J. I., Hocking, A. D., and Fleet, G. H. (2006). "Comparison of hyphal length, ergosterol, mycelium dry weight, and colony diameter for quantifying growth of fungi from foods," in *Advances in Food Mycology*, eds A. D. Hocking, J. I. Pitt, R. A. Samson, and U. Thrane (New York, NY: Springer), 49–67.

Viswanath, B., Rajesh, B., Janardhan, A., Kumar, A. P., and Narasimha, G. (2014). Fungal laccases and their applications in bioremediation. *Enzyme Res.* 2014:21. doi: 10.1155/2014/163242

Wan, C., and Li, Y. (2012). Fungal pretreatment of lignocellulosic biomass. *Biotechnol. Adv.* 30, 1447–1457. doi: 10.1016/j.biotechadv.2012.03.003

Yang, D.-D., François, J. M., and de Billerbeck, G. M. (2012). Cloning, expression and characterization of an aryl-alcohol dehydrogenase from the white-rot fungus *Phanerochaete chrysosporium* strain BKM-F-1767. *BMC Microbiol.* 12:126. doi: 10.1186/1471-2180-12-126

Yang, J., Dong, S., Jiang, Q., Si, Q., Liu, X., and Yang, J. (2013). Characterization and expression of cytoplasmic copper/zinc superoxide dismutase ([Cu-Zn] SOD) gene under temperature and hydrogen peroxide (H_2O_2) in rotifer Brachionus calyciflorus. *Gene* 518, 388–396. doi: 10.1016/j.gene.2012.12.101

Yousuf, P. Y., Hakeem, K. U. R., Chandna, R., and Ahmad, P. (2012). "Role of glutathione reductase in plant abiotic stress,"in *Abiotic Stress Responses in Plants*, eds P. Ahmad and M. N. V. Prasad (New York, NY: Springer), 149–158.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Xiao, Ma, Li, Yu, Li and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 June 2018
doi: 10.3389/fmicb.2018.01374

Investigation of Citrinin and Pigment Biosynthesis Mechanisms in *Monascus purpureus* by Transcriptomic Analysis

*Bin Liang[1], Xin-Jun Du[1], Ping Li[1], Chan-Chan Sun[1] and Shuo Wang[1,2]**

[1] Key Laboratory of Food Nutrition and Safety, Tianjin University of Science and Technology, Ministry of Education, Tianjin, China, [2] Beijing Advanced Innovation Center for Food Nutrition and Human Health, Beijing Technology and Business University, Beijing, China

Monascus purpureus YY-1 is widely used in food colorant production in China. Our previous study clearly illustrated the whole-genome data for YY-1 and provided useful insight into evolutionary research and industrial applications. However, the presence of citrinin, which has nephrotoxic, hepatotoxic, and carcinogenic activities, has attracted attention to the safety of *Monascus* products. In an effort to reduce the harmful effects of citrinin in *Monascus*-related products, a random mutant of *M. purpureus* YY-1 with low citrinin production (designated as "winter") was obtained in this study. To analyze the biosynthesis and regulation mechanisms of pigment and citrinin, a transcriptomic analysis of the *M. purpureus* YY-1 and winter strains was performed. Comparative transcriptomic analysis reveals *pksCT*, the essential gene for citrinin synthesis, showed a low expression level in *M. purpureus* YY-1 and winter, which suggested there might be isoenzymes in *M. purpureus* YY-1 that were responsible for the citrinin synthesis during evolution. In addition, changes in transcription factor expression may also influence the network regulating the citrinin synthesis pathway in *M. purpureus*. Moreover, the yields of pigments produced by the winter mutant were significantly increased. Repressing the central carbon metabolism and improving the acetyl-CoA pool can contribute to a high pigment yield, and enhanced NADPH regeneration can also lead to the metabolic flux of pigment production in *M. purpureus*. Investigations into the biosynthesis and regulation of citrinin and pigment production in *M. purpureus* will enhance our knowledge of the mechanisms behind the biosynthesis of fungal secondary metabolites.

Keywords: pigment, citrinin, transcriptomic analysis, *Monascus purpureus*, solid-state fermentation

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Jiangxin Wang, Shenzhen University, China
Alejandro Sanchez-Flores, Universidad Nacional Autónoma de México, Mexico

****Correspondence:***
Shuo Wang
s.wang@tust.edu.cn

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *09 January 2018*
Accepted: *06 June 2018*
Published: *28 June 2018*

Citation:
Liang B, Du X-J, Li P, Sun C-C and Wang S (2018) Investigation of Citrinin and Pigment Biosynthesis Mechanisms in Monascus purpureus by Transcriptomic Analysis. Front. Microbiol. 9:1374. doi: 10.3389/fmicb.2018.01374

INTRODUCTION

The filamentous fungus *Monascus* has a long history of being used to produce fermented foods in eastern Asia (Shi and Pan, 2011). The health benefits of *Monascus*-fermented products have been described in the Compendium of Materia Medica. *Monascus* species can produce many worthwhile secondary metabolites, such as *Monascus* pigments and monacolin K (Endo, 1979; Feng et al., 2012).

Monascus pigments, which are polyketide components, range in structure from tetraketides to octaketides. Representative classes include the anthraquinones, naphthoquinones,

hydroxyanthraquinones, and azaphilone structures, each of which exhibits an array of color hues (Mapari et al., 2010). *Monascus* pigments have been used as natural food colorants for over 1000 years worldwide, especially in China. Recently, an increasing number of investigations have shown that *Monascus* pigments exhibit biological activities, such as anti-inflammatory, anticancer and antihyperlipidemic activities (Hong et al., 2008; Lin et al., 2011). In addition, monacolin K is considered an effective agent for reducing blood cholesterol levels (Su et al., 2003). However, *Monascus spp*. also produce citrinin, a toxic product that has nephrotoxic and hepatotoxic effects in animals and humans (Blanc et al., 1995). Many countries, especially Japan, European countries and the United States, have enacted legislation to limit the content of citrinin in *Monascus*-fermented products (Endo, 1979; Blanc et al., 1995). Thus, citrinin contamination has become a hindrance to the export of *Monascus*-fermented products from China.

The generation of non-citrinin-producing strains for use in the commercial production of *Monascus*-related products is a primary task. Although the optimization of culture conditions is a traditional strategy for decreasing the content of citrinin, it is hard to block citrinin biosynthesis completely in *Monascus*. Moreover, random mutagenesis and screening is widely used in the development of low- or non-citrinin-producing mutants, but some mutants are not genetically stable (Shao et al., 2014). Therefore, it is necessary and urgent to clearly understand the biosynthetic pathways involved in *Monascus*-fermented products.

In recent years, several genes related to the biosynthesis of citrinin and pigments have been cloned and characterized. Shimizu et al. first cloned a polyketide synthase (PKS) gene from *Monascus purpureus*. The *pksCT* disruptant lost its ability to produce citrinin, suggesting that *pksCT* plays a critical role in citrinin biosynthesis. Unfortunately, the *pksCT* mutant was not genetically stable, and citrinin production was restored after successive cultivation (Shimizu et al., 2005). Similarly, deletion of *pksCT* in *Monascus aurantiacus* resulted in dramatically decreased citrinin production. Intriguingly, this mutant exhibited a stronger ability to produce red and yellow pigments (Fu et al., 2007). In addition, fungal secondary metabolite pathways are also tightly controlled at the transcriptional level. Shimizu et al. have identified a major transcriptional activator (CtnA) of citrinin biosynthesis in *M. purpureus*. The deletion of *ctnA* significantly decreased the production of the *pksCT* transcript, leading to reduced citrinin production (Shimizu et al., 2007).

Compared with the citrinin biosynthesis pathway, the mechanism regulating the biosynthesis of *Monascus* pigments is more complicated. The first pigment biosynthetic gene cluster was obtained by T-DNA random mutagenesis in *M. purpureus*. The transcriptional regulator gene *mppR1* and PKS gene *MpPKS5* are two major components of pigment biosynthesis (Balakrishnan et al., 2013). Chen et al. systematically investigated the pigment gene cluster in *Monascus ruber* M7. The wild-type strain M7 could produce red, orange, and yellow pigments. However, the *MpigE* disruption strain could only produce four kinds of yellow pigments and could barely produce red pigments. Intriguingly, the mutant could restore pigment production by supplementation with M7PKS-1, an intermediate in the production of *Monascus* pigments (Liu et al., 2016). Furthermore, the overexpression of *MpigE* had positive effects on pigment formation and led to a decrease in citrinin production (Liu et al., 2014). These researchers also identified and characterized a pigment regulatory gene (*pigR*). The *pigR* deletion strain could no longer produce pigments, but its capacity to produce citrinin was greatly enhanced (Xie et al., 2013). These results suggested a close relationship between the pigment and citrinin biosynthesis pathways.

Monascus purpureus YY-1 is widely used in food colorant production in China. Our previous study clearly illustrated the whole-genome data for YY-1 and provided useful insight into evolutionary research and industrial applications (Yang et al., 2015). In this paper, a random mutant with low citrinin production was obtained through protoplast transformation. We performed comparative transcriptomic analysis between the wild-type strain YY-1 and the mutant, which revealed the mechanisms underlying pigment and citrinin biosynthesis. Further investigations into the mechanisms regulating pigment and citrinin biosynthesis are necessary. The scientific elucidation of the complex relationship between pigment biosynthesis and citrinin biosynthesis in *Monascus* will broaden our knowledge of the mechanisms involved in the biosynthesis of fungal secondary metabolites and will provide important insights into the genetic engineering of industrial strains to increase the production of specific metabolites.

MATERIALS AND METHODS

Strain and Culture Conditions

Monascus purpureus YY-1 was obtained from Fujian Province in China (Yang et al., 2015). *Escherichia coli* DH5α was employed for DNA manipulation.

Monascus purpureus strains were grown on potato dextrose agar (PDA) at 30°C for 10 days. Spore suspensions were prepared as previously described (Yang et al., 2015). Rice (45 g) was soaked in distilled water overnight and then transferred to a 250 mL Erlenmeyer flask. Subsequently, the rice was autoclaved at 121°C for 20 min. After cooling, the steamed rice was inoculated with a 10% spore suspension and cultivated at 30°C. After 10 days of cultivation, the red yeast rice was dried at 60°C for 12 h and then ground.

Construction of Random Mutants of *M. purpureus* YY-1 With Deficient Citrinin Secretion

Random mutants were generated by transforming the *M. purpureus* YY-1 strain with a hygromycin B resistance cassette from the plasmid pHPH. *M. purpureus* spores were obtained from the PDA plates and were inoculated in 100 mL spore medium for 40 h at 30°C and 170 rpm. The mycelium was harvested by filtering, washed with lysis buffer twice and

then digested for 4 h under shaking at 100 rpm and 30°C. The protoplasts were separated by filtering through three layers of Miracloth and were collected by centrifugation at 1000 g at 10°C for 20 min. Then, the protoplasts were washed twice with 10 mL STC buffer (0.6 M sorbitol, 10 mM Tris-HCl, 10 mM $CaCl_2$, pH 6.5). The pellet was resuspended in 400 μL STC buffer, and 10 μg DNA was added to the protoplast suspension, followed by incubation on ice for 20 min. Subsequently, 100 μL PEG800 was added to the suspension, and the suspension was incubated for 5 min at room temperature. The suspension was diluted by adding 4 mL 1 M sorbitol, and centrifuging at 3000 rpm for 10 min. The pellet was resuspended in 2 mL 1 M sorbitol. Two hundred microliters of the suspension was plated onto regeneration medium. The cultures were incubated at 30°C for 12 h and then overlaid with 10 mL regeneration medium containing 100 μg/mL hygromycin B. After 7 days of cultivation, the transformants were picked onto selective medium.

Quantitative Analysis of Citrinin and Pigments

The red yeast rice powder was extracted with 75% ethanol by ultrasonic extraction for 20 min. After keeping the solution still for 30 min, the concentration of citrinin was determined by high-performance liquid chromatography (HPLC) using an RF-10Axl fluorescence detector (λ_{ex} = 331 nm, λ_{em} = 500 nm) and a ZORBAXE Eclipse XDB C18 column (5 μL, 250 × 4.6 mm). Elution was performed at 38°C with 45% acetonitrile (v/v, pH 2.5) at a flow rate of 1.0 mL/min.

Extraction of pigments for analysis followed a similar protocol to that for citrinin extraction. The red yeast rice powder was extracted with 70% ethanol at 60°C for 1 h. After filtration, the filtrate was measured at 410, 465, and 505 nm.

RNA Extraction

Mycelia from the *M. purpureus* YY-1 and winter strains for RNA isolation were harvested after 8 days of cultivation. Then, the mycelia were immediately homogenized in liquid nitrogen and stored at −80°C until used for RNA extraction.

Total RNA was extracted using a modification of a method described previously (Tian et al., 2009). The frozen mycelia were ground to a powder in a prechilled mortar with a prechilled pestle and then processed with the RNeasy® Plant Mini Kit (QIAGEN Translational Medicine Company Limited, Germany) following the manufacturer's protocol. The quantity and concentration of the final RNA samples were measured using agarose gel electrophoresis and the Eppendorf BioPhotometer® basic and stored at −80°C.

Transcriptome Construction and Analysis

Total RNA samples from each strain were purified with the Qiagen RNeasy Mini kit plus on-column treatment with DNase I to eliminate genomic DNA contamination. Reverse transcription was performed with the PrimeScript™ RT Reagent Kit (Takara Biotechnology Company Limited, Japan), according to the manufacturer's protocol. The cDNA library was sequenced at Novogene Corporation (Tianjin, China) by means of the Illumina HiSeq 2500 platform. All data in the present study were generated by sequencing independent biological triplicates.

Prior to analyzing the data, the quality of the raw sequencing reads was checked by the FastQC tool (v0.10.0).[1] Low-quality and dirty reads with adapter sequences and reads with more than 20% of bases with a QA of <25 or with "N" bases were removed using the NGS QC Toolkit (v2.3) (Patel and Jain, 2012). All clean reads were mapped against predicted transcripts from the *M. purpureus* YY-1 genomic sequence (NCBI: samn08978766) using TopHat (v2.1.1) (Trapnell et al., 2012) with at most two mismatches. HTSeq (v0.7.2) (Anders et al., 2015) was employed to calculate the raw counts of reads mapped to unique exons, and the normalized transcript abundance was described by RPKM (reads per kilobase per million mapped reads). Differential gene expression analysis was carried out by DESeq (v1.30.0) (Anders and Huber, 2010), with the raw read count and exon length as inputs. A *P-adj* value <0.05 indicated significantly different expression levels between the YY-1 and winter strains (**Supplementary Table S1**). The raw RNA-seq data are accessible in the Gene Expression Omnibus under accession number GSE107628.

Quantitative Real-Time PCR

The CFX96 real-time PCR detection system (Bio-Rad, Hercules, United States) was used for quantitative PCR analysis. Reagents were obtained from TOYOBO (One-step qPCR Kit, OSAKA, Japan). The PCR reaction mixture (20 μl) included 75 ng template RNA, 0.4 μM primers and 10 μL RNA-direct SYBR® Green Realtime PCR Master Mix, according to the manufacturer's instruction. Three replicates were analyzed. The relative transcript level of each gene was calculated by the $2^{-\Delta\Delta C_t}$ method, with the expression in the WT as the control and the expression of GAPDH as the internal standard. The primers used in this study are listed in **Supplementary Table S7**.

Statistical Significance Tests and Data Plotting

Unless otherwise noted, all statistical significance tests were done with a one-tailed homoscedastic (equal variance) *t*-test. The *R* packages used for data plotting included Pheatmap (version 0.7.7) and ggplot2 (version 0.9.3.1).

RESULTS AND DISCUSSION

Screening and Characterization of the Mutant With Deficient Citrinin Secretion but Hyperproduction of Pigment

To obtain a non-citrinin-producing mutant, the hygromycin phosphotransferase-encoding gene was introduced into *M. purpureus* YY-1 using protoplast transformation with pHPH, as described previously (Das et al., 1989). A series

[1] http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

of hygromycin-resistant insertional mutants of *M. purpureus* was generated and screened for the loss of the ability to produce citrinin. Among these transformants, one transformant showed low citrinin production in solid-state fermentation, and we designated this mutant as "winter." As shown in **Figure 1A**, citrinin production was significantly decreased in the *M. purpureus* winter strain. After 14 days of cultivation, the citrinin production by the mutant decreased by 35.6-fold, to 1.38 ± 0.22 μg/g, compared to that of the wild-type strain, which produced 49.17 ± 2.64 μg/g of citrinin. Surprisingly, the mutant produced significantly more pigment than the YY-1 strain during the late fermentation period. As shown in **Figures 1B–D**, from the eighth day to the fourteenth day, the pigment production increased significantly. After 14 days of cultivation, the yields of yellow, orange, and red pigment produced by the winter strain reached 8539 ± 426, 8470 ± 332, and 7667 ± 321 U/g, respectively. Compared with the YY-1 wild-type strain, the pigment yields of the winter strain increased by 37% (yellow), 54% (orange), and 49% (red). Thus, the generation of the winter mutant, which exhibited a high yield of pigments and a low yield of citrinin, would greatly reduce the risk of *Monascus*-related products and would provide a foundation for research on the molecular mechanism of citrinin and pigment biosynthesis.

Transcriptomic Analysis of the *M. purpureus* Winter Strain During Pigment and Citrinin Production

To analyze the molecular mechanism of pigment and citrinin production, a transcriptomic analysis of the *M. purpureus* winter strain during pigment production was performed. Due to the rapid increase in pigment production after 8 days of cultivation, total RNA from the mycelia of the *M. purpureus* YY-1 and winter strains was sequenced in independent biological triplicates. The reads per kilobase of exon model per million mapped reads (RPKMs) were calculated as the normalized expression values of each annotated gene. Differential expression analysis was conducted using DESeq (Anders and Huber, 2010). As shown in **Figure 2A**, a high reproducibility of the RNA-seq data from three biological triplicates was observed for the YY-1 and winter strains; the Spearman correlation coefficient was greater than 0.95 for the biological replicates and was less than 0.9 between the different strains (**Figure 2A**). Moreover, the good

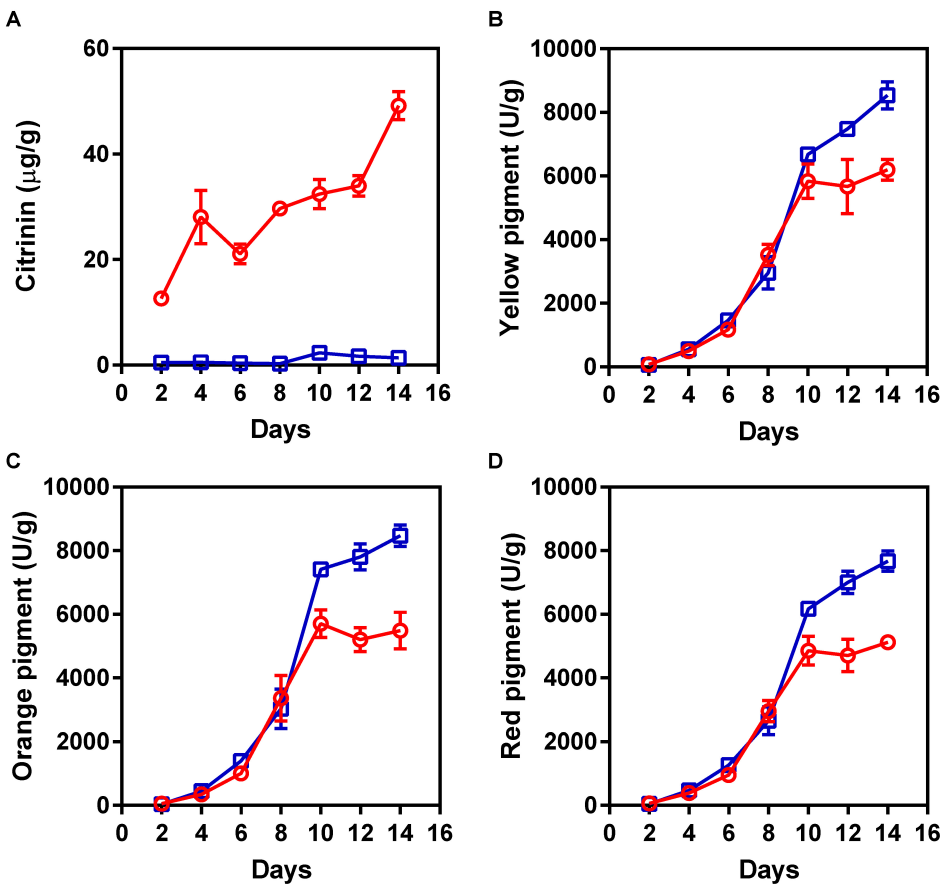


FIGURE 1 | The production of pigments and citrinin by the *Monascus purpureus* YY-1 and winter strains. **(A)** Citrinin, **(B)** yellow pigment, **(C)** orange pigment, and **(D)** red pigment. The red and blue lines represent the YY-1 and winter strains, respectively. The error bars indicate the standard deviations of three independent experiments.

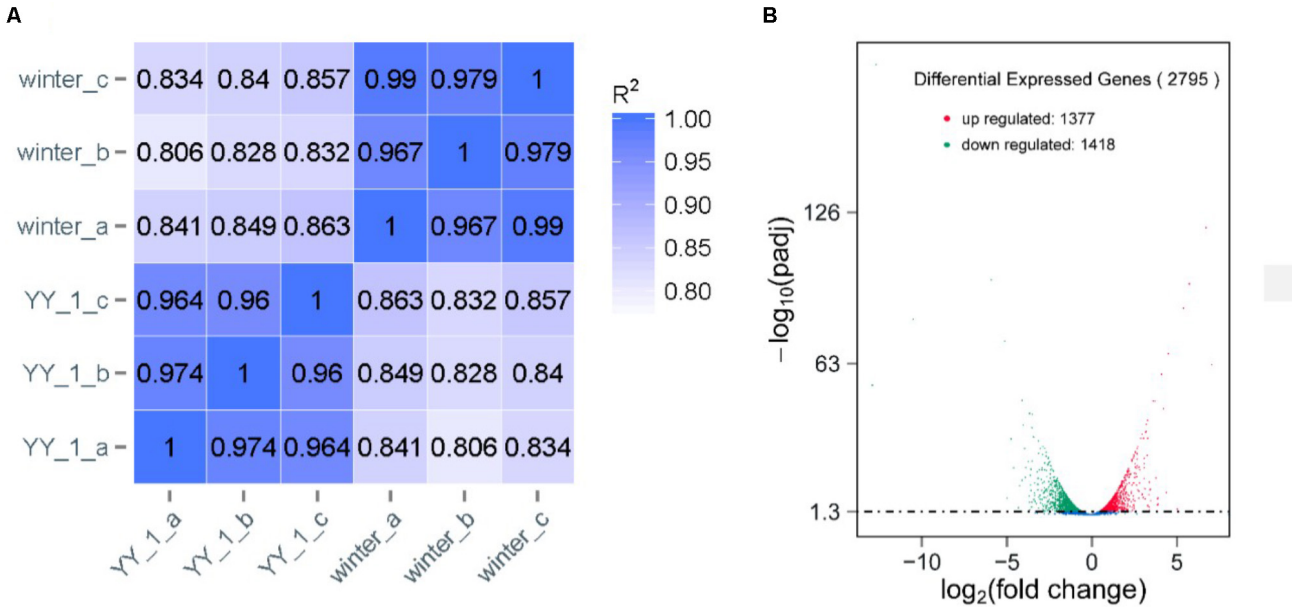


FIGURE 2 | Transcriptomic profiles of the *M. purpureus* YY-1 and winter strains during pigment production. **(A)** Correlation comparison between biological replicates. The Spearman correlation coefficients (R^2) were calculated. **(B)** Differential gene expression. In the statistical analysis of differential gene expression, the transcriptomes derived from the YY-1 strain were used as controls.

agreement with the qPCR results demonstrated the reliability of the RNA-seq data (**Supplementary Figure S1**). To discover genes that were significantly upregulated or downregulated between the conditions tested, only genes with a *P*-value from the DESeq analysis of less than 0.001 were analyzed further (**Supplementary Table S1**).

Compared to the YY-1 strain, 1377 genes in the winter strain showed significantly increased expression levels (**Figure 2B** and **Supplementary Table S1**). Gene Ontology is a standardized system for gene functional classification, which includes biological process, cellular components and molecular functions of gene products (Chicco and Masseroli, 2016). To further elucidate the types of genes with significantly different expression levels, GO enrichment analyses were performed. The result showed that the genes with increased transcription levels in the winter strain were enriched in several biological processes at a *P*-value of less than 0.05, including the small molecule metabolic process, organic acid metabolic process, single-organism metabolic process, oxoacid metabolic process, carboxylic acid metabolic process, and cellular amino acid metabolic process (**Figure 3A** and **Supplementary Table S2**). The oxoacid metabolic process is related to the metabolism of ketone compounds, including pigments and citrinin. Several genes clustered in the functional category of organic acid metabolic processes and involved in two potential pathways for acetyl-CoA biosynthesis, which is a precursor for pigment biosynthesis, showed upregulated expression levels in the winter strain.

In the Gene Ontology analysis of 1418 genes with significantly decreased expression levels, we found that the downregulated genes were involved in three functional classifications, including the biological process, cellular components and molecular functions of gene products. In the biological process enrichment analysis, transmembrane transport and the carbohydrate metabolic process were significantly enriched. In the carbohydrate metabolic process cluster, the transcription levels of 26 genes involved in carbohydrate degradation were significantly decreased (**Supplementary Table S4**), which might result from a reduction in biomass accumulation when secondary metabolites are synthesized.

Thus, a large percentage of genes showing decreased expression levels participated in metabolic- and transporter-related roles. In the cellular component enrichment analysis, the genes enriched were mainly involved in membrane structure. The molecular function enrichment analysis function showed that hydrolase activity, iron ion binding, and anion transmembrane transporter activity were significantly enriched (**Figure 3B** and **Supplementary Table S3**).

Regulation of Citrinin Biosynthesis in *M. purpureus*

Citrinin, a mycotoxin which has nephrotoxic activity in mammals, was isolated from most cultures of *Monascus* strains in 1993 (Endo, 1979). Citrinin targets the kidney, resulting in teratogenicity, carcinogenicity, and mutagenicity. Therefore, decreasing the content of citrinin is becoming a necessity that should be addressed as soon as possible. This study included a comprehensive transcriptomic analysis to elucidate the citrinin synthesis pathway in *M. purpureus*. The citrinin synthesis gene cluster comprised *pksCT*, *ctnA*, *orf1*, *orf3*, *ctnB*, and *ctnC*. According to previous reports, *pksCT*, which encodes a multifunction protein that contains putative domains

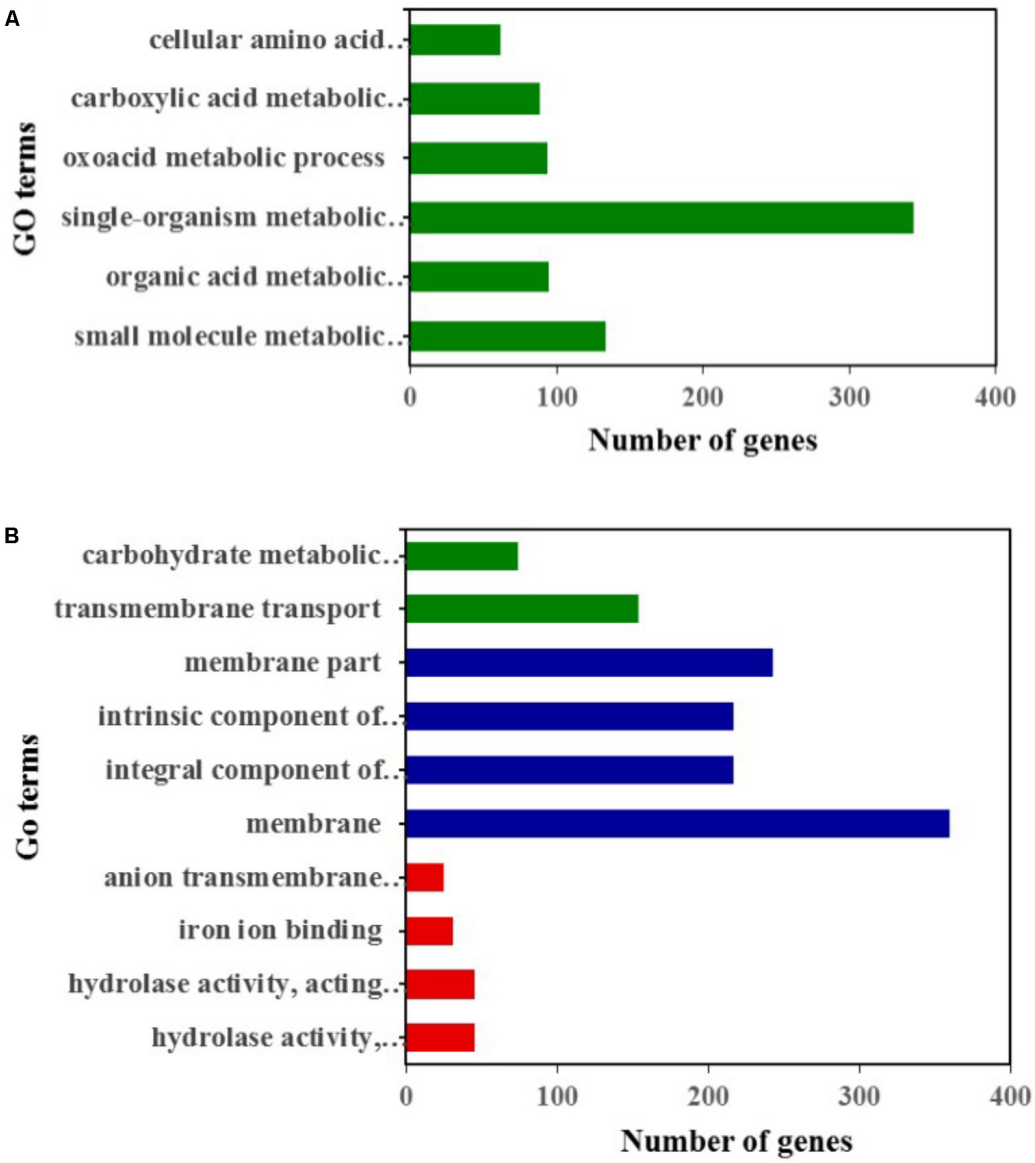


FIGURE 3 | GO classification analysis of genes with significantly upregulated **(A)** and downregulated **(B)** expression levels in the winter strain compared to those in the YY-1 strain. Green, blue, and red represent biological processes, cellular components, and molecular functions, respectively.

for ketosynthase, acyltransferase (AT), acyl carrier protein (ACP), and a rare methyltransferase, was identified by gene disruption as the key factor for citrinin synthesis in *Monascus aurantiacus* and *Monascus purpureus* (Shimizu et al., 2005; Fu et al., 2007). Surprisingly, in spite of the high expression level of the activator *ctnA* gene, *pksCT* exhibited an extremely low expression level; the RPKM values were 6.45 and 2.01 in the *M. purpureus* winter and wild-type strain, respectively (**Figure 4**). This phenomenon was not consistent with the high production of citrinin in the *M. purpureus* YY-1 strain. The result suggested that *pksCT* may not be the only PKS gene involved in citrinin synthesis. To some extent, this possibility was suggested by the study performed by Shimizu et al. (2005). They found that a *pksCT* disruptant was unstable and could revert its phenotype with repeat cultivation. We speculated that *M. purpureus* YY-1 might have evolved isoenzymes of *pksCT*, such as C2.25 and C6.140, which are responsible for citrinin synthesis; the transcription levels of these purported isoenzymes changed by 2.2- and 2.6-fold, respectively, in the *M. purpureus* winter strain.

The expression levels of *orf1*, *ctnA*, *orf3*, and *ctnB* were upregulated in the winter strain relative to that in YY-1 strain, especially *ctnA* and *ctnB*, which exhibited 50- and 98-fold changes, respectively, although inefficient citrinin synthesis was detected in the winter strain (**Figure 4**). This observation is inconsistent with the report that CtnA, which contains a typical Zn(II)2Cys6 DNA binding motif, functions as an activator

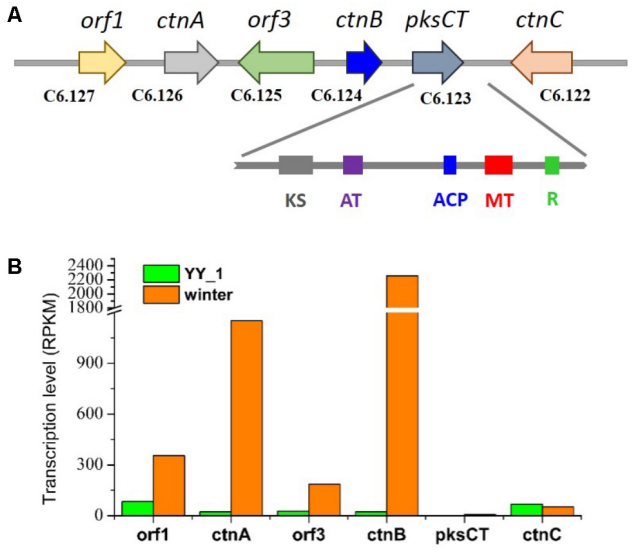


FIGURE 4 | Gene cluster for citrinin synthesis **(A)** and transcription levels **(B)**. KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; MT, methyltransferase; R, reductase.

of *pksCT* and *orf5* and that its disruption leads to the reduction of citrinin production to barely detectable levels (Shimizu et al., 2007).

Transcription factors play a critical role in the regulation of gene expression patterns that control the overall metabolism, including secondary metabolic pathways (Gutierrez et al., 2017; Wong and Matus, 2017). Therefore, the expression levels of transcription factor genes in *M. purpureus* during pigment and citrinin production were analyzed. Based on Pfam annotation, 179 genes were identified as transcription factor-encoding candidates, among which 80 showed significantly different expression levels in the *M. purpureus* winter strain compared to those in the YY-1 strain (**Supplementary Table S5**). Because of transcriptomic fluctuations, 48 transcription factor genes with more than a two-fold change were analyzed further. Of these genes, 37 showed decreased expression levels and 11 exhibited increased transcription levels in the winter strain (**Figure 5**) relative to the levels in the YY-1 strain, including the transcription activator C6.126. In the winter strain, the transcription levels of C5.141 and C3.618, which had respective RPKM values of 19 and 160, decreased by 4.3- and 2.7-fold, respectively, compared to those in the YY-1 strain.

These results indicated that the proteins encoded by *orf1*, *orf3*, and *ctnB* were essential but were not rate-limiting enzymes for citrinin production in *M. purpureus* and that the higher expression levels of these genes in the *M. purpureus* winter strain might result from the decreased expression levels of downstream citrinin synthesis pathway genes. In addition, some transcription factors, including C6.126, C5.141, and C3.618, may be involved in the regulation of citrinin biosynthesis. However, further investigations to elucidate the citrinin biosynthesis pathway in *M. purpureus* are necessary.

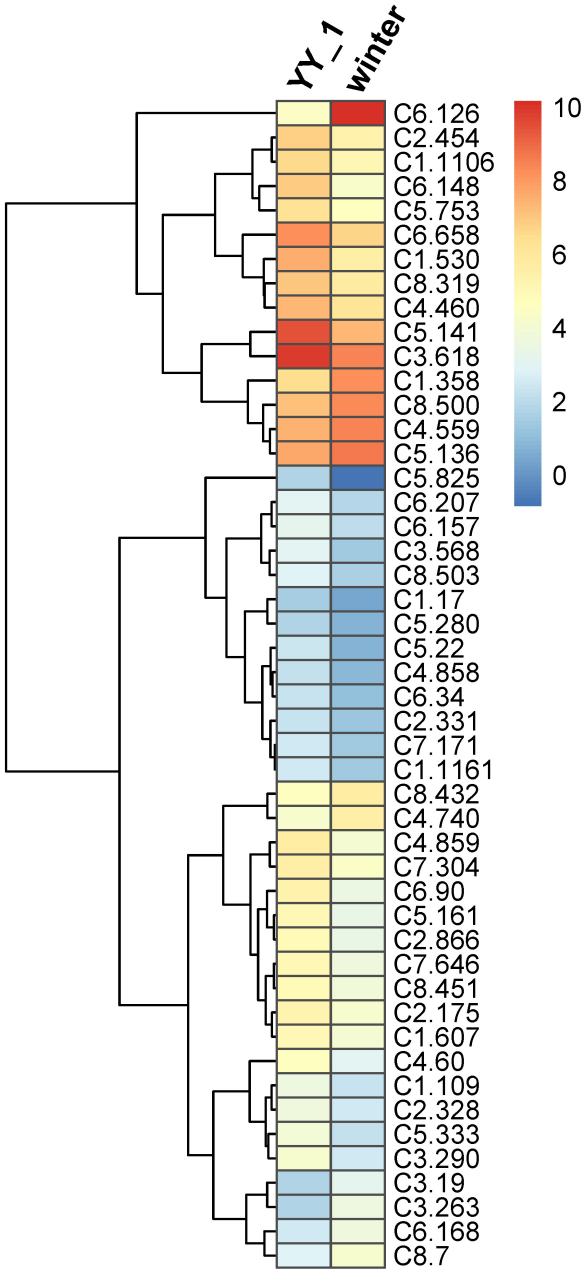


FIGURE 5 | Expression profile of transcription factor genes with differential transcription levels in the *M. purpureus* winter strain compared to those in the YY-1 strain. The tree is a dendrogram of the cluster analysis obtained by the hclust method in R, which is a hierarchical partitioning of data that visually resembles a heat map and reflects the gene expression values in different strains.

A Comprehensive Analysis of the Biosynthesis Pathway of Pigment Hyper-Production

Pigment synthesis pathways in fungi have been attracting more attention, and great progress has been made in this area (Chiang et al., 2009; Mapari et al., 2010). PKS, which

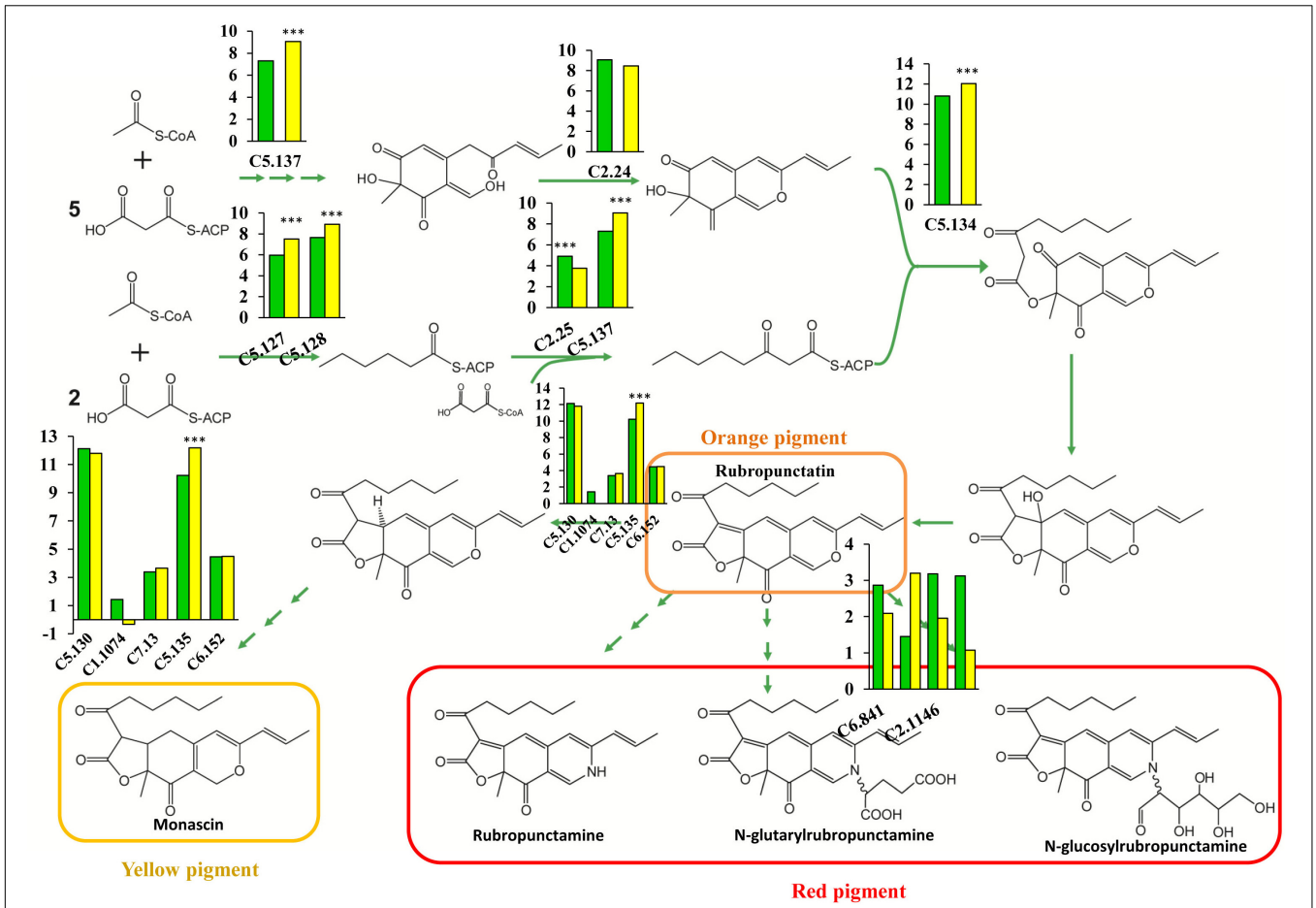


FIGURE 6 | Schematic of pigment biosynthesis pathways in *M. purpureus*, including pathways for orange pigment (rubropunctatin), yellow pigment (monascin), red pigment (rubropunctamine), and water-soluble pigments. The histogram depicts the gene expression levels in the YY-1 (green) and winter (yellow) strains. The *y*-axis represents log2(RPKM), and *** denotes a *P*adj value < 0.001.

closely resembles animal fatty acid synthase (FAS) and is a large, multifunctional polypeptide composed of a set of catalytic domains, including β-ketoacyl synthase (KS), AT, and ACP domains, is the main participant in fungal pigment production (Thomas, 2001; Cox, 2007; Crawford and Townsend, 2010). Several novel enzymes involved in pigment synthesis have been found in *M. purpureus*, yet the importance of genes involved in pigment production has not been investigated, and the rate-limiting enzymes remain to be identified. In this study, we performed a comprehensive analysis of the pigment biosynthesis pathways in *M. purpureus*. PKS (C5.137) serves as a dual-functioning PKS and is responsible for the synthesis of 3-oxoacyl-thioester from acylthioester, showing an increased expression level with a 3.3-fold change in the winter strain (**Figure 6**). A similar reaction is catalyzed by the enzymes encoded by the FAS gene pair (*MpFasA* and *MpFasB*) in *M. azaphilone* (Balakrishnan et al., 2013). In addition, C5.134, a gene that encodes 3-O-acetyltransferase, exhibited high expression levels; its RPKM was 1798 and 4198 in the YY-1 and winter strains, respectively, suggesting the essential role of this gene in pigment production. Furthermore, the 2.33-fold upregulation of C5.134 in the winter strain suggested that C5.134 might be one of the rate-limiting enzymes. The expression levels of the FAS genes (C5.127 and C5.128) were significantly increased by 2.91- and 2.41-fold, respectively, in the winter strain; this result is similar to that in a previous study, which showed that pigment production is involved in fatty acid synthesis (Miazek et al., 2017). Intriguingly, the expression of oxidoreductase (C5.135), which is used to form yellow pigment from orange pigment, was upregulated by 3.9-fold, yet the genes encoding the enzymes involved in the catalytic reaction of rubropunctatin to red pigments showed low expression levels in the winter and parental strains, and the expression levels were not significantly different between the strains.

The large transcriptional differences between the YY-1 and winter strains indicated the complex pigment synthesis pathway in *M. purpureus*. The most remarkable transcriptional changes are related to the acetyl-CoA metabolic network, pigment precursors and a crucial metabolite involved in both central carbon and energy metabolism. Several genes upregulated in the winter strain are involved in two potential pathways

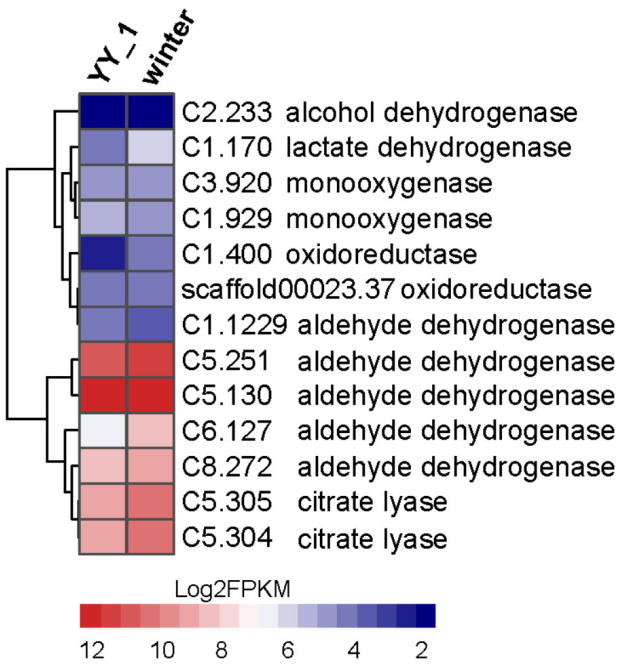


FIGURE 7 | Expression levels of the genes involved in acetyl-CoA biosynthesis in the *M. purpureus* winter and YY-1 strains. The tree is a dendrogram of the cluster analysis obtained by the hclust method in *R*, which is a hierarchical partitioning of data that visually resembles a heat map and reflects the gene expression values in different strains.

for acetyl-CoA biosynthesis (**Figure 7**), such as the genes encoding citrate lyase (C5.304 and C5.305) in the citrate pathway and the genes encoding oxidoreductase activity (C1.400 and C1.170) and aldehyde dehydrogenase (C5.251 and C6.127). Notably, the expression level of C6.127 was increased by 4.2-fold (from 84.06 to 353.68) in the *M. purpureus* winter strain. This observation was consistent with previous reports that fungi responding to carbon starvation stress on poor carbon sources can regulate the biosynthesis of secondary metabolism products (Jorgensen et al., 2009; Yang et al., 2015). In addition, when the KEGG function enrichment analysis of genes upregulated in the *M. purpureus* winter strain was performed, the genes involved in the glycolysis pathway, which provides acetyl-CoA for pigment production, were found to be enriched significantly (**Supplementary Table S6**). In accordance with these possibilities, the transcriptome data showed that many genes involved in the TCA cycle, such as citrate synthase (C4.56), and succinate dehydrogenase (C6.271 and C7.592), are downregulated in the winter strain; this downregulation does not favor the metabolic flux of acetyl-CoA to pigment.

The increased expression levels of key enzymes in the pentose phosphate pathway, 6-phosphogluconate dehydrogenase (C3.60) and glucose-6-phosphate dehydrogenase (C7.667), indicated enhanced NADPH regeneration. This observation suggested that secondary metabolite biosynthesis could be enhanced by an additional supply of NADPH. Oxidoreductase plays an essential role in secondary metabolite synthesis, especially pigment synthesis. In the winter strain, the genes (C5.133, C5.124, and C8.16) encoding oxidoreductase show significantly increased expression levels; these genes exhibited 4.0-, 5.65-, and 2.5-fold changes, respectively. This phenomenon is consistent with the results of previous reports that showed that increasing the expression level of MpigE, an ortholog of C5.133, promoted pigment biosynthesis in *Monascus ruber* M7 (Liu et al., 2014) and that azaH and tropB (orthologs of C5.124 and C8.16, respectively) were proposed to be involved in pyrone ring formation (Zabala et al., 2012; Balakrishnan et al., 2013). These results indicated that oxidoreductase-encoding genes may be candidates for the metabolic engineering of *Monascus* to aid the industrial use of secondary metabolites. Overall, the enhancement of central carbon metabolism, acetyl-CoA biosynthesis and NADPH regeneration led to a large increase in the production of pigments.

CONCLUSION

This is the first report that a strain of *M. purpureus* with low citrinin secretion was obtained with random mutation with protoplast transformation and was analyzed by comparative transcriptomic techniques. We found that the expression level of the key gene involved in citrinin synthesis, *pksCT*, was low in both *M. purpureus* YY-1 and its mutant; this result suggested that citrinin synthesis might be catalyzed by isoenzymes that evolved in the *M. purpureus* YY-1 strain. Furthermore, the expression analysis of the transcription factor genes can promote the study of networks regulating the citrinin synthesis pathway in *M. purpureus*. Moreover, the production of the three pigments (red, yellow, and orange) in the winter mutant during the whole fermentation process was significantly increased. Transcriptome analysis showed that central carbon metabolism, acetyl-CoA biosynthesis and NADPH regeneration were enhanced in the winter strain; these pathways can contribute to the high yield of pigments. This study improves our understanding of pigment and citrinin production in *M. purpureus* and benefits the application of *M. purpureus* in the production of food and pharmaceuticals.

AUTHOR CONTRIBUTIONS

SW and X-JD designed the project. BL carried out the experiments. BL, X-JD, PL, and C-CS participated in the data analysis and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the Ministry of Science and Technology of the People's Republic of China (Project No. 2014BAD04B03), and the International Science

and Technology Cooperation Program of China (Project No. 2014DFR30350).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01374/full#supplementary-material

FIGURE S1 | Correlation comparison between the qRT-PCR results and the RNA-Seq data. The relative expression levels of the selected genes were normalized to their expression in wild-type strains. The Spearman correlation coefficients and P values are shown in the top left of each figure. For the qRT-PCR results (horizontal coordinate), the values represent the means of triplicates.

TABLE S1 | The profiles of RNA-seq reads mapped to the genome of *Monascus purpureus* and the differential expression analysis.

TABLE S2 | Gene ontology analysis of 1377 genes with up-regulated expression levels in *M. purpureus* winter. *p*-values highlighted in yellow indicate functional classifications that were significantly enriched.

TABLE S3 | Gene ontology analysis of 1418 genes with down-regulated expression levels in *M. purpureus* winter. p-values highlighted in yellow indicate functional classifications that were significantly enriched.

TABLE S4 | Genes involved in carbohydrate degradation was significantly decreased in *M. purpureus* winter, belonging to the cluster of carbohydrate metabolic process.

TABLE S5 | Expression profile of transcription factor genes with differential transcriptional levels and more than two-fold charges in *M. purpureus* winter, compared to that in the YY-1 strain.

TABLE S6 | The analysis of KEGG function enrichment of up-regulated genes in M. purpureus winter in *M. purpureus* winter. *p*-values highlighted in yellow indicate functional classifications that were significantly enriched.

TABLE S7 | RT-PCR Primers used in this study.

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Balakrishnan, B., Karki, S., Chiu, S. H., Kim, H. J., Suh, J. W., Nam, B., et al. (2013). Genetic localization and in vivo characterization of a *Monascus azaphilone* pigment biosynthetic gene cluster. *Appl. Microbiol. Biotechnol.* 97, 6337–6345. doi: 10.1007/s00253-013-4745-4749

Blanc, P. J., Laussac, J. P., Le Bars, J., Le Bars, P., Loret, M. O., Pareilleux, A., et al. (1995). Characterization of monascidin A from *Monascus* as citrinin. *Int. J. Food Microbiol.* 27, 201–213. doi: 10.1016/0168-1605(94)00167-5

Chiang, Y. M., Szewczyk, E., Davidson, A. D., Keller, N., Oakley, B. R., and Wang, C. C. (2009). A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. *J. Am. Chem. Soc.* 131, 2965–2970. doi: 10.1021/ja8088185

Chicco, D., and Masseroli, M. (2016). Ontology-based prediction and prioritization of gene functional annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 13, 248–260. doi: 10.1109/TCBB.2015.2459694

Cox, R. J. (2007). Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* 5, 2010–2026. doi: 10.1039/b704420h

Crawford, J. M., and Townsend, C. A. (2010). New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* 8, 879–889. doi: 10.1038/nrmicro2465

Das, A., Gokhale, D. V., and Peberdy, J. F. (1989). Protoplast fusion and genetic-recombination in intrastrain and interstrain crossing in *Aspergillus niger*. *Enzyme Microb. Technol.* 11, 2–5. doi: 10.1016/0141-0229(89)90104-X

Endo, A. (1979). Monacolin K, a new hypocholesterolemic agent produced by a *Monascus* species. *J. Antibiot.* 32, 852–854. doi: 10.7164/antibiotics.32.852

Feng, Y., Shao, Y., and Chen, F. (2012). *Monascus* pigments. *Appl. Microbiol. Biotechnol.* 96, 1421–1440. doi: 10.1007/s00253-012-4504-3

Fu, G., Xu, Y., Li, Y., and Tan, W. (2007). Construction of a replacement vector to disrupt *pksCT gene* for the mycotoxin citrinin biosynthesis in *Monascus aurantiacus* and maintain food red pigment production. *Asia Pac. J. Clin. Nutr.* 16(Suppl. 1), 137–142.

Gutierrez, E., Garcia-Villaraco, A., Lucas, J. A., Gradillas, A., Gutierrez-Manero, F. J., and Ramos-Solano, B. (2017). Transcriptomics, targeted metabolomics and gene expression of blackberry leaves and fruits indicate flavonoid metabolic flux from leaf to red fruit. *Front. Plant Sci.* 8:472. doi: 10.3389/fpls.2017.00472

Hong, M. Y., Seeram, N. P., Zhang, Y., and Heber, D. (2008). Anticancer effects of Chinese red yeast rice versus monacolin K alone on colon cancer cells. *J. Nutr. Biochem.* 19, 448–458. doi: 10.1016/j.jnutbio.2007.05.012

Jorgensen, T. R., Goosen, T., Hondel, C. A., Ram, A. F., and Iversen, J. J. (2009). Transcriptomic comparison of *Aspergillus niger* growing on two different sugars reveals coordinated regulation of the secretory pathway. *BMC Genomics* 10:44. doi: 10.1186/1471-2164-10-44

Lin, C. P., Lin, Y. L., Huang, P. H., Tsai, H. S., and Chen, Y. H. (2011). Inhibition of endothelial adhesion molecule expression by *Monascus purpureus*-fermented rice metabolites, monacolin K, ankaflavin, and monascin. *J. Sci. Food Agric.* 91, 1751–1758. doi: 10.1002/jsfa.4371

Liu, J., Zhou, Y., Yi, T., Zhao, M., Xie, N., Lei, M., et al. (2016). Identification and role analysis of an intermediate produced by a polygenic mutant of *Monascus* pigments cluster in *Monascus ruber* M7. *Appl. Microbiol. Biotechnol.* 100, 7037–7049. doi: 10.1007/s00253-016-7397-8

Liu, Q., Xie, N., He, Y., Wang, L., Shao, Y., Zhao, H., et al. (2014). *MpigE*, a gene involved in pigment biosynthesis in *Monascus ruber* M7. *Appl. Microbiol. Biotechnol.* 98, 285–296. doi: 10.1007/s00253-013-5289-8

Mapari, S. A., Thrane, U., and Meyer, A. S. (2010). Fungal polyketide azaphilone pigments as future natural food colorants. *Trends Biotechnol.* 28, 300–307. doi: 10.1016/j.tibtech.2010.03.004

Miazek, K., Remacle, C., Richel, A., and Goffin, D. (2017). Beech wood *Fagus sylvatica* dilute-acid hydrolysate as a feedstock to support *Chlorella sorokiniana* biomass, fatty acid and pigment production. *Bioresour. Technol.* 230, 122–131. doi: 10.1016/j.biortech.2017.01.034

Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. doi: 10.1371/journal.pone.0030619

Shao, Y. C., Lei, M., Mao, Z. J., Zhou, Y. X., and Chen, F. S. (2014). Insights into *Monascus* biology at the genetic level. *Appl. Microbiol. Biotechnol.* 98, 3911–3922. doi: 10.1007/s00253-014-5608-8

Shi, Y. C., and Pan, T. M. (2011). Beneficial effects of *Monascus purpureus* NTU 568-fermented products: a review. *Appl. Microbiol. Biotechnol.* 90, 1207–1217. doi: 10.1007/s00253-011-3202-x

Shimizu, T., Kinoshita, H., Ishihara, S., Sakai, K., Nagai, S., and Nihira, T. (2005). Polyketide synthase gene responsible for citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* 71, 3453–3457. doi: 10.1128/aem.71.7.3453-3457.2005

Shimizu, T., Kinoshita, H., and Nihira, T. (2007). Identification and in vivo functional analysis by gene disruption of *ctnA*, an activator gene involved in citrinin biosynthesis in *Monascus purpureus*. *Appl. Environ. Microbiol.* 73, 5097–5103. doi: 10.1128/aem.01979-06

Su, Y. C., Wang, J. J., Lin, T. T., and Pan, T. M. (2003). Production of the secondary metabolites gamma-aminobutyric acid and monacolin K by *Monascus*. *J. Ind. Microbiol. Biotechnol.* 30, 41–46. doi: 10.1007/s10295-002-0001-5

Thomas, R. (2001). A biosynthetic classification of fungal and streptomycete fused-ring aromatic polyketides. *Chembiochem* 2, 612–627. doi: 10.1002/1439-7633(20010903)2:9<612::AID-CBIC612>3.0.CO;2-Z

Tian, C., Beeson, W. T., Iavarone, A. T., Sun, J., Marletta, M. A., Cate, J. H., et al. (2009). Systems analysis of plant cell wall degradation by the model filamentous

fungus *Neurospora crassa*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22157–22162. doi: 10.1073/pnas.0906810106

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Wong, D. C. J., and Matus, J. T. (2017). Constructing integrated networks for identifying new secondary metabolic pathway regulators in grapevine: recent applications and future opportunities. *Front. Plant Sci.* 8:505. doi: 10.3389/fpls.2017.00505

Xie, N., Liu, Q., and Chen, F. (2013). Deletion of *pigR* gene in *Monascus ruber* leads to loss of pigment production. *Biotechnol. Lett.* 35, 1425–1432. doi: 10.1007/s10529-013-1219-1

Yang, Y., Liu, B., Du, X., Li, P., Liang, B., Cheng, X., et al. (2015). Complete genome sequence and transcriptomics analyses reveal pigment biosynthesis and regulatory mechanisms in an industrial strain, *Monascus purpureus* YY-1. *Sci. Rep.* 5:8331. doi: 10.1038/srep08331

Zabala, A. O., Xu, W., Chooi, Y. H., and Tang, Y. (2012). Characterization of a silent azaphilone gene cluster from *Aspergillus niger* ATCC 1015 reveals a hydroxylation-mediated pyran-ring formation. *Chem. Biol.* 19, 1049–1059. doi: 10.1016/j.chembiol.2012.07.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Liang, Du, Li, Sun and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 April 2018
doi: 10.3389/fmicb.2018.00621

Identification and Characterization of *als* Genes Involved in D-Allose Metabolism in Lineage II Strain of *Listeria monocytogenes*

*Lu Zhang, Yan Wang, Dongxin Liu, Lijuan Luo, Yi Wang and Changyun Ye**

State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing, China

Listeria monocytogenes, an important food-borne pathogen, causes listeriosis and is widely distributed in many different environments. In a previous study, we developed a novel enrichment broth containing D-allose that allows better isolation of *L. monocytogenes* from samples. However, the mechanism of D-allose utilization by *L. monocytogenes* remains unclear. In the present study, we determined the metabolism of D-allose in *L. monocytogenes* and found that lineage II strains of *L. monocytogenes* can utilize D-allose as the sole carbon source for growth, but lineage I and III strains cannot. Transcriptome analysis and sequence alignment identified six genes (*lmo0734* to *0739*) possibly related to D-allose metabolism that are only present in the genomes of lineage II strains. Recombinant strain ICDC-LM188 containing these genes showed utilization of D-allose by growth assays and Biolog phenotype microarrays. Moreover, *lmo0734* to *0736* were verified to be essential for D-allose metabolism, *lmo0737* and *0738* affected the growth rate of *L. monocytogenes* in D-allose medium, while *lmo0739* was dispensable in the metabolism of D-allose in *L. monocytogenes*. This is the first study to identify the genes related to D-allose metabolism in *L. monocytogenes,* and their distribution in lineage II strains. Our study preliminarily determined the effects of these genes on the growth of *L. monocytogenes*, which will benefit the isolation and epidemiological research of *L. monocytogenes*.

Keywords: *Listeria monocytogenes*, lineage II, D-allose, metabolism, genes

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Beatrix Stessl, Veterinärmedizinische Universität Wien, Austria
Pasquale Russo, University of Foggia, Italy

****Correspondence:***
Changyun Ye
yechangyun@icdc.cn

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *30 November 2017*
Accepted: *16 March 2018*
Published: *04 April 2018*

Citation:
Zhang L, Wang Y, Liu D, Luo L, Wang Y and Ye C (2018) Identification and Characterization of als Genes Involved in D-Allose Metabolism in Lineage II Strain of Listeria monocytogenes. Front. Microbiol. 9:621. doi: 10.3389/fmicb.2018.00621

INTRODUCTION

Listeria monocytogenes is a gram-positive foodborne pathogen of humans that causes listeriosis (Low and Donachie, 1997). Clinical symptoms include meningitis, septicemia, abortion, perinatal infections, and gastroenteritis. The elderly, newborns, pregnant women, and immunocompromised patients are more susceptible (Low and Donachie, 1997; Mead et al., 1999; Kathariou, 2002). Based on the serological reactions of somatic (O) and flagellar antigens (H), *L. monocytogenes* is divided into 13 serotypes, comprising 1/2a, 1/2b, 1/2c, 3a, 3b, 3c, 4a, 4ab, 4b, 4c, 4d, 4e, and 7. Previous study has been reported that a multiplex PCR assay has been developed to identify the serotype, which can separate four major serotype groups (Ward et al., 2008). Furthermore, 1/2b, 3b, 4b, 4d, 4e, and 7 belong to lineage I; serotypes 1/2a, 1/2c, 3a, and 3c belong to lineage II; while serotypes

4a and 4c belong to lineage III (Roberts et al., 2006). Some 4a, 4c, and atypical 4b strains were assigned to lineage IV based on sequence data analysis (Orsi et al., 2011). Lineage I strains are mostly isolated from human listeriosis cases, lineage II mostly exist in food and the environment, and lineage III and IV are rare and mainly found in animal hosts (Liu, 2008). Nevertheless, Zilelidou et al. (2015) have reported that multiply *L. monocytogenes* strains could isolate from a single food sample, suggesting that highly invasive *L. monocytogenes* strains might show stronger growth than low or modestly invasive strains. However, studies did not suggest a link between the competitive advantage and strain origin, serotype, or sequence type (Zilelidou et al., 2015, 2016).

L. monocytogenes is ubiquitous in the environment and is often isolated from contaminated environments and foods. Notably, *L. monocytogenes* has the ability to grow at refrigeration temperatures and survive at high salt concentrations, making it difficult to control (Zhang, 2007). *L. monocytogenes* also can adhere to the surfaces of food processing equipment and resist adverse conditions. It has been reported that *L. monocytogenes* was persistently isolated from raw pork in open markets in China (Luo et al., 2017). *L. monocytogenes* is aerobic, or facultatively anaerobic under some conditions. *L. monocytogenes* can metabolize different carbon sources, including D-glucose. Strains can utilize D-glucose to form lactate, acetate, and acetoin when they are grown aerobically, whereas, acetoin is not formed during anaerobic growth (Pine et al., 1989). In addition, cellobiose, fructose, mannose, galactose, lactose, salicin, maltose, dextrin, and glycerol can also be utilized to produce acid. The growth rate of *L. monocytogenes* increases if fermentable sugars are present (Siddiqi and Khan, 1982).

D-Allose, an aldohexose, is a rare monosaccharide in nature and its physiological functions are varied. D-allose inhibits tumor cell multiplication and active oxygen production (Murata et al., 2003). D-allose is used in a novel enrichment broth to improve the isolation of *L. monocytogenes* and reduce the growth of non-target organisms (Liu et al., 2017). However, the details of the utilization of D-allose in *L. monocytogenes* are unknown. D-allose has been reported to be a carbon source for *Escherichia coli* (Gibbins and Simpson, 1964). It could be converted to fructose-6-phosphate via D-allose-6-phosphate and D-allulose-6-phosphate in *Aerobacter aerogenes* and *E. coli* K12 (Gibbins and Simpson, 1964; Kim et al., 1997). Two operons are involved in D-allose metabolism in *E. coli*, one contains the *alsI* gene, which encodes allose 6-phosphate isomerase, and the other consists of six contiguous genes, *alsR*, *B*, *A*, *C*, *E*, and *K*. *alsR* is a negative regulator for the operon. *alsB, A,* and *C* comprise the transport system of D-allose, *alsB* encodes a D-allose-binding protein, *alsA* encodes an ATP-binding component, and *alsC* encodes a transmembrane protein. The allulose 6-phosphate epimerase is encoded by *alsE* and *alsK* is thought to encode allokinase. Furthermore, the two operons are adjacent in the genome of *E. coli* K12 (Poulsen et al., 1999). In *L. monocytogenes*, operons comprising the genes involved in D-allose are unknown. In this study, we investigated the utilization of D-allose in *L. monocytogenes* and analyzed the associated genes for D-allose metabolism.

METHODS AND MATERIALS

Strains, Plasmids, Media, and Culture Conditions

L. monocytogenes strains EGD-e and ICDC-LM188 were used as D-allose utilizing and D-allose non-utilizing control strains, respectively. A total of 278 *L. monocytogenes* strains, isolated from different areas and sample sources in China, were used for the D-allose utilization assay in this study (Supplementary Table 1). Lineage I and III strains (**Table 1**) were used for gene function analysis. These strains originated from American Type Culture Collection (ATCC, USA) or were isolated in China and stored in State Key Laboratory of Infectious Disease Prevention and Control, China CDC at −80°C (**Table 1**). Plasmid PIMK2 was used to construct the gene expression vectors (**Table 1**; Lauer et al., 2002). Bacterial cells were cultured in Brain Heart Infusion (BHI) Broth (BD, USA) with shaking at 220 rpm. Modified Welshimer's broth (MWB) medium was used in the growth assay (Premaratne et al., 1991), and Luria-Bertani (LB) medium (Oxoid, UK) was used in the transcriptome analysis.

TABLE 1 | Plasmids and strains used in this study.

Strains or plasmids	Genotypes
PLASMIDS	
pIMK2	pIMK rrnB T1 terminators PhelP
pAL1	pIMK2 *lmo0734 0735 0736 0737 0738 0739*
pAL2	pIMK2 *lmo0734 0735 0736 0737 0738*
pAL3	pIMK2 *lmo0734 0735 0736 0737 0739*
pAL4	pIMK2 *lmo0734 0735 0736 0738 0739*
pAL5	pIMK2 *lmo0734 0735 0737 0738 0739*
pAL6	pIMK2 *lmo0734 0736 0737 0738 0739*
pAL7	pIMK2 *lmo0735 0736 0737 0738 0739*
STRAINS (SEROTYPE)	
EGD-e (1/2a)	*lmo0734, 0735, 0736, 0737, 0738, 0739*
ICDC-LM188 (1/2b)	None
RS801 (1/2b)	ICDC-LM188::pAL1
RS802 (1/2b)	ICDC-LM188::pAL2
RS803 (1/2b)	ICDC-LM188::pAL3
RS804 (1/2b)	ICDC-LM188::pAL4
RS805 (1/2b)	ICDC-LM188::pAL5
RS806 (1/2b)	ICDC-LM188::pAL6
RS807 (1/2b)	ICDC-LM188::pAL7
ATCC 19114 (4a)	None
RS19114 (4a)	ATCC19114::pAL1
ATCC 19115 (4b)	None
RS19115 (4b)	ATCC19115::pAL1
ACTC 19116 (4c)	None
RS19116 (4c)	TCC19116::pAL1
ATCC 19117 (4d)	None
RS19117 (4d)	ATCC19117::pAL1
ATCC 19118 (4e)	None
RS19118 (4e)	ATCC19118::pAL1

D-Allose Utilization and Growth Curve Analysis

Single colonies of *L. monocytogenes* strains EGD-e and ICDC-LM188 were incubated in BHI medium with constant shaking at 220 rpm overnight at 37°C, and then a 1% volume of the BHI cultures were transferred separately into BHI medium, and shaken at 220 rpm at 37°C until the OD_{600} values reached 0.6. The strains were then diluted 1:100 into 0.2% D-allose or D-glucose MWB medium (MWB medium containing 2 g/L D-allose or D-glucose as the sole carbohydrate). The strains were grown in a Bioscreen C microbiology reader (Growth Curves Ltd, Helsinki, Finland) at 37°C with shaking. The OD_{600} was monitored at 30 min intervals for 24 h. To obtain the maximum growth rate for each strain (Pontinen et al., 2017), the OD_{600} data were fitted to growth curves using GraphPad Prism software (GraphPad Software Inc., San Diego, CA, USA).

To test the universality of D-allose utilization in *L. monocytogenes*, a single colony of each of 278 *L. monocytogenes* strains was cultured in BHI medium at 37°C with shaking at 220 rpm overnight, and 1% volumes of the cultures were transferred into BHI with shaking 220 rpm at 37°C until the OD_{600} value reached 0.6. Subsequently, the strains were diluted 1:100 into 0.2% D-Allose MWB medium, and the cultures were shaken at 220 rpm for 24 h at 37°C. Finally, the growth status of the bacterial cells were detected.

Transcriptome Analysis and D-Allose Metabolism Related Gene Detection

L. monocytogenes EGD-e was selected as the reference for the transcriptome analysis. One colony of EGD-e was incubated in BHI medium at 37°C with shaking at 220 rpm overnight. Then, 1% volumes of the culture were respectively transferred into 0.2% D-allose and D-glucose LB medium at 37°C with shaking at 220 rpm until the OD_{600} value reached 0.6. The bacteria were then harvested by centrifugation and total RNA was extracted using the Trizol method (Rio et al., 2010). The extracted RNA was used for RNA sequencing with BGISEQ-500, according to the sequencing procedure used by the Beijing Genomics Institute (BGI, China). After filtering out low quantity sequences, clean reads were generated and mapped to the reference genome using HISAT and Bowtie2 tools (Langmead et al., 2009; Kim et al., 2015). The Poisson distribution method was used to analyze differentially expressed genes (DEGs). Clustering analysis of DEGs was performed with cluster and java Treeview software (Michael Eisen, Stanford University, Stanford, CA). Gene Ontology (GO) (http://www.geneontology.org/) annotation was performed for the screened DEGs. Six pairs of specific primers were designed to verify DEGs based on the results of the transcriptome analysis (Supplementary Table 2). DNAs of above 278 *L. monocytogenes* strains were isolated by boiling, detected by PCR using the specific primers, and confirmed by agarose gel electrophoresis.

Plasmids Construction and Electro-Transformation

To construct the D-allose related gene expression vectors, the DNA fragments, being obtained by PCR using the specific primers (**Table 2**), were purified from agarose gel electrophoresis using a DNA Gel Extraction kit (Takara, Shiga, Japan). The purified products were cloned into the *Sal*I and *Bam*HI sites of vector pIMK2 to form plasmids pAL1 to pAL7 (**Figure 1**) via In-Fusion or T4 ligase Cloning. Different lineage strains (**Table 1**) were used to prepare competent cells (Camilli et al., 1990). Plasmid pAL1 was electrotransformed into competent cells of the strains of all lineages except lineage II, and plasmids pAL2, pAL3, pAL4, pAL5, pAL6, and pAL7 were respectively introduced into competent cells of *L. monocytogenes* strain ICDC-LM188 (**Table 1**). The electroporation conditions were 2.5 KV, 400 Ω, 25 μF. Afterwards, the strains were grown on BHI plates containing 50 μg/ml kanamycin in a 37°C incubator for 24 h. The clones were detected by PCR using the specific primers (Supplementary Table 2).

Growth Phenotype and Curve Analysis

The positive clones were incubated in BHI medium at 37°C with shaking at 220 rpm overnight, 1% volumes of cultures were transferred into BHI medium with shaking at 37°C until the OD_{600} value reached 0.6. Afterwards, 1% volumes of inoculum were cultured in 0.2% D-allose MWB medium with shaking 220 rpm at 37°C for 24 h. The cells growth status was monitored (Zwietering et al., 1990; Markkula et al., 2012). Biolog phenotype microarrays (Biolog, USA) were used to analyze the utilization of D-allose by *L. monocytogenes* strain ICDC-LM188 and its recombinant strains (Miller and Rhoden, 1991). Based on the manufacturer's protocol for *Listeria*, the Biolog 96-well micro-plates PM1 and PM2a were used to test carbohydrate metabolism; plate PM2a contains D-allose. Clones were inoculated into 20 ml 1 × Buffer IF-0a, and the turbidity of the suspension was checked until 81% light transmittance was achieved. Then, 1.76 ml of the cell suspension was transferred into 22.24 ml PM1 and PM2a inoculating fluid, and 100 μl/well of the cell suspension was inoculated in 96-well micro-plates, which were incubated in an OmniLog instrument at 37°C for 24 h. Growth curve analysis of the recombinant strains was carried out according to a previous description (Pontinen et al., 2017) using the Bioscreen C microbiology reader (Growth Curves Ltd., Helsinki, Finland.) at 37°C, 0.2% D-allose medium was used as the broth.

RESULTS

D-Allose Utilization and Growth Assay

In our previous study, D-allose was discovered to be a carbon source for *L. monocytogenes* growth, and *L. monocytogenes* could grow in the novel *Listeria* Allose Enrichment Broth (LAEB). In addition to D-allose, tryptone, peptone, and Lablemco powder could contributed a small amount of carbon source and energy in the LAEB medium (Liu et al., 2017). In this study, to identify D-allose utilization in the growth of *L. monocytogenes*, MWB medium was used (Premaratne et al., 1991). D-glucose was replaced by D-allose in MWB

TABLE 2 | Primers used to construct the expression vectors.

Plasmids	Primers (5′-3′)
pAL1	AL1 P1 F
	CGCGGATCCTTAAATTGAGTCTTTTAAAGATAATTCAACGGGAACAATCTCTG
	AL1 P1 R
	CGCGTCGACTCAGTAGTTCAAATCATTTCCATTAG
pAL2	AL2 P1 F
	CCATGGAAAAGGATCCTTAAATTGAGTCTTTTAAAGATAATTCAACGGGAACAATCTCTG
	AL2 P1 R
	AGTTGCTAGGTAACGATGTCTTAATTTTCGC
	AL2 P2 F
	GCGAAAATTAAGACATCGTTACCTAGCAACT
	AL2 P2 R
	ACCCCCTCGAGTCGACTTAAAGTACTGGTGTTAAAAGCGTAT
pAL3	AL3 P1 F
	CCATGGAAAAGGATCCTTAAATTGAGTCTTTTAAAGATAATTCAACGGGAACAATCTCTG
	AL3 P1 R
	TTGCAAATTAAACATTTACGATTTCTGCTTGCCATTCTAAAGTACGTATAAGGT
	AL3 P2 F
	AAGCAGAAATCGTAAATGTTTAATTTGCAAAAAGGTTTTCCAG
	AL3 P2 R
	ACCCCCTCGAGTCGACTCAGTAGTTCAAATCATTTCCATTAG
pAL4	AL4 P1 F
	CCATGGAAAAGGATCCTTAAATTGAGTCTTTTAAAGATAATTCAACGGGAACAATCTCTG
	AL4 P1 R
	TTTTTGATAATCCATTTAATCATTTTCATCTTCAATTCTAGCAATCATTTCGACGCG
	AL4 P2 F
	GATGAAAATGATTAAATGGATTATCAAAAACTAGCTAAAGAG
	AL4 P2 R
	ACCCCCTCGAGTCGACTCAGTAGTTCAAATCATTTCCATTAG
pAL5	AL5 P1 F
	CCATGGAAAAGGATCCTTAAATTGAGTCTTTTAAAGATAATTCAACGGGAACAATCTCTG
	AL5 P1 R
	TACTTCTTGCTCCATTCATCGTAATTCCTCCATCTTTTCTTG
	AL5 P2 F
	GAGGAATTACGATGAATGGAGGCAAGAAGTAACACGACTAAT
	AL5 P2 R
	ACCCCCTCGAGTCGACTCAGTAGTTCAAATCATTTCCATTAG
pAL6	AL6 P1 F
	CCATGGAAAAGGATCCTTAAATTGAGTCTTTTAAAGATAATTCAACGGGAACAATCTCTG
	AL6 P1 R
	AATAGCAATTTTCATATGAAAAAAACATCCATAAAAGATATCGC
	AL6 P2 F
	GGATGTTTTTTTCATATGAAAATTGCTATTGGAAATGATCATGTTGG
	AL6 P2 R
	ACCCCCTCGAGTCGACTCAGTAGTTCAAATCATTTCCATTAG
pAL7	AL7 F
	CGCGGATCCATGAGCAAAAAACTGATTTGTCCTTCTAT
	AL7R
	CGCGTCGACTCAGTAGTTCAAATCATTTCCATTAG

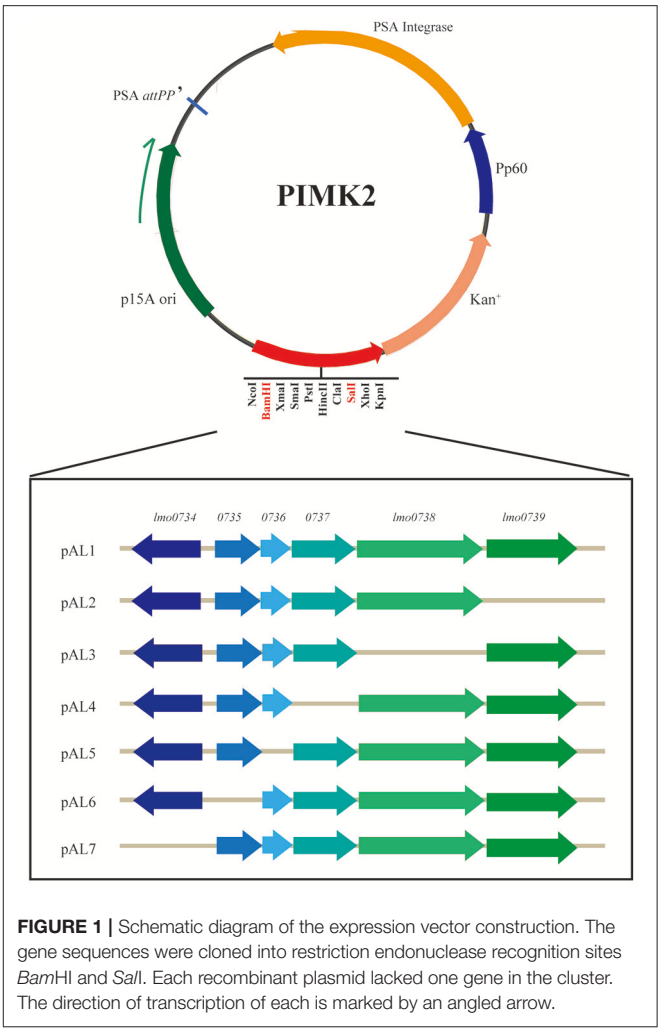


FIGURE 1 | Schematic diagram of the expression vector construction. The gene sequences were cloned into restriction endonuclease recognition sites *Bam*HI and *Sal*I. Each recombinant plasmid lacked one gene in the cluster. The direction of transcription of each is marked by an angled arrow.

medium as the sole carbon and energy source for growth. *L. monocytogenes* strains EGD-e and ICDC-LM188 were used as experimental references, because EGD-e could grow well in LAEB medium, while ICDC-LM188 could not (Liu et al., 2017). Growth rates were calculated (Zwietering et al., 1990; Markkula et al., 2012) and were similar between EGD-e and ICDC-LM 188 in 0.2% D-glucose MWB medium. Moreover, there was no obvious difference in the growth of EGD-e between D-glucose and D-allose medium (**Figure 2**). By contrast, no growth of ICDC-LM 188 was observed in the 0.2% D-allose MWB medium (**Figure 2**). Thus, ICDC-LM188 could not utilize D-allose as carbon and energy source.

In addition, 278 experimental strains were cultured as above to verify their D-allose utilization. The result showed that 171 strains belonging to lineage II could grow in D-allose MWB medium; however, the other 107 strains, including 91 strains of lineage I and 16 strains of lineages III and IV, could not (**Table 3**). The result was analyzed using Fisher's test, which revealed that there was a correlation (Pearson

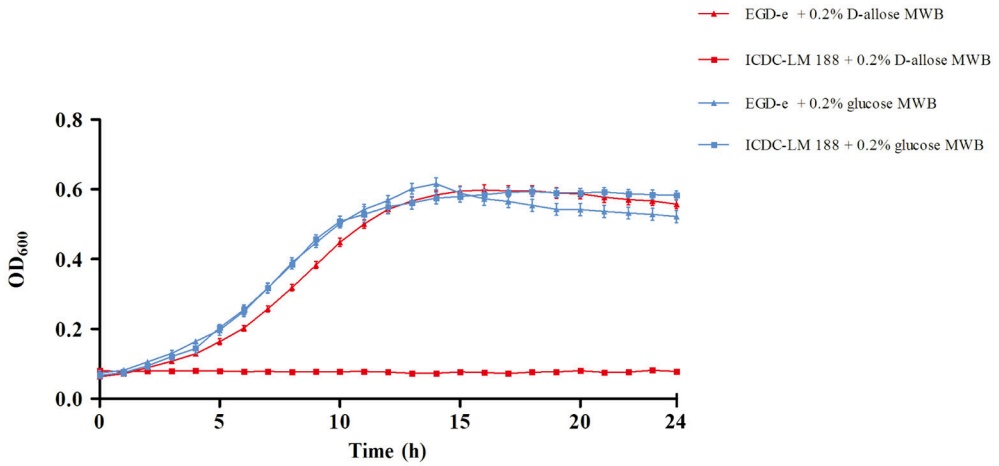


FIGURE 2 | Growth curve of *L. monocytogenes* EGD-e and ICDC-LM188. The red line shows growth in 0.2% D-allose modified Welshimer's broth (MWB) medium, the blue line shows growth in 0.2% D-glucose MWB medium. Triangles represent EGD-e and squares represent ICDC-LM188.

TABLE 3 | D-allose utilization in *L. monocytogenes* strains.

		Growth in D-allose medium		
		+	−	Total
Lineage	I	0	91	91
	II	171	0	171
	III	0	16	16
Total		171	107	278

$r = 0.7071$, $P < 0.001$) between the lineages and D-allose utilization.

Transcriptome Analysis and D-Allose Metabolism Related Gene Detection

Using RNA-Seq technology, more than 23 million clean reads were acquired after filtering out low quality reads. The transcriptomes were successfully sequenced and deposited in GenBank (Accession No. SRR6281666 and SRR6281667). According to the quality control and expression level calculated using the RSEM algorithm and the fragments per kilobase of transcript per million mapped reads (FPKM) method (Li and Dewey, 2011), the differential expression of 15 genes were upregulated and 13 genes were downregulated in D-allose utilizing group compared with the D-glucose utilizing group. Moreover, pathway enrichment analysis of the DEGs was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the functions of the DEGs were classified as nucleotide metabolism and carbohydrate metabolism in the secondary pathway of KEGG (Supplementary Table 3).

Subsequently, whole-genome alignment of different lineage strains was conducted and a gene cassette (*lmo0734* to *0739*) of strain EGD-e was observed to be absent in *L. monocytogenes* lineage I, III, and IV strains (**Figure 3A**). This specific gene cassette was highly expressed under D-allose utilization conditions (**Figure 3B**). DNA was extracted from 278 experimental strains, including different lineages, to verify the presence of the genes in the different lineages. All strains of lineage II contained the gene cassette; however, it was not present in any of the other lineages. Notably, only the strains of lineage II could utilize D-allose as a carbon source for growth. These results were analyzed using Fisher's test, which showed that there was a correlation between the six genes and D-allose utilization in *L. monocytogenes* (Pearson $r = 0.7071$, $P < 0.001$).

Growth and Biolog Characterization of Recombinant Strains

D-allose utilization of the recombinant strains was verified using a growth assay (**Table 1**). D-allose supported the growth in different lineage strains containing pAL1 (containing all six genes; Supplementary Figure 1A). This revealed that the specific gene cassette was involved in D-allose metabolism in different lineage strains of *L. monocytogenes*. To further identify the key genes in the D-allose metabolism of *L. monocytogenes*, plasmids pAL2, pAL3, pAL4, pAL5, pAL6, and pAL7 were constructed according to the schematics shown in **Figure 1**, and were transferred separately into strain ICDC-LM188. Strains RS802, RS803, and RS804 grew on D-allose medium; however, strains RS805, RS806, and RS807 did not (Supplementary Figure 1B). This revealed that genes *lmo0734*, *lmo0735*, and *lmo0736* were essential for D-allose metabolism of *L. monocytogenes*.

The recombinant strains RS801, RS802, RS803, and RS804 showed different growth rates in MWB medium with 0.2% D-allose (**Figure 4**). The growth curves were analyzed in Graphpad Prism, the specific growth rates were calculated (**Table 4**). According to *t*-test analysis (Zwietering et al., 1990; Pontinen et al., 2017), the growth rates of RS801 and RS802 were similar, while RS803 and RS804 were similar. Moreover, the growth rates of RS801 and RS802 were higher than those of RS803 and RS804 ($P < 0.01$).

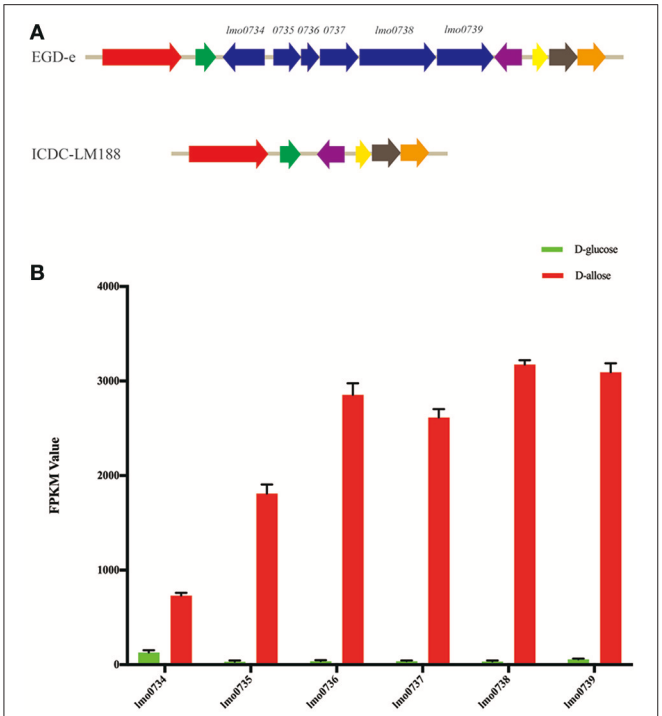


FIGURE 3 | Genome analysis and gene expression levels. **(A)** The cluster of whole-genome alignment between *L. monocytogenes* EGD-e and ICDC-LM188. The same color angled arrows represent the same genes. Six genes of EGD-e were found to be absent in ICDC-LM188 (between the green and purple angled arrows). **(B)** Comparison of the expression level comparisons of five different genes between D-glucose and D-allose ($P < 0.001$). FPKM, fragments per kilobase of transcript per million mapped reads.

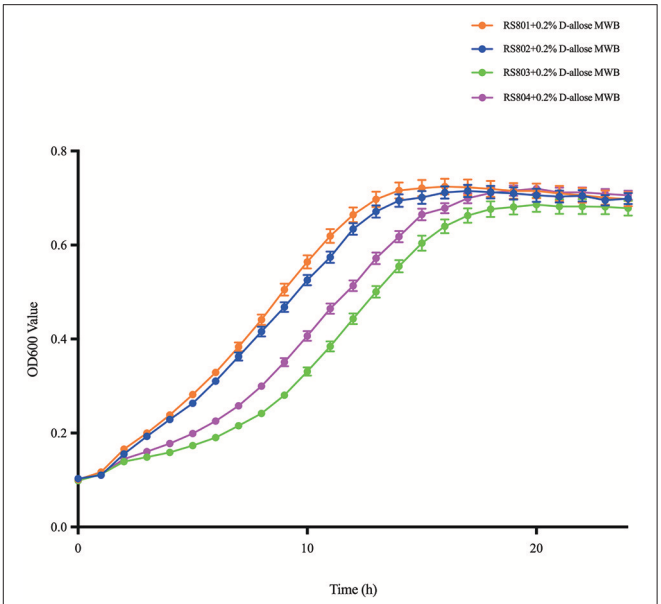


FIGURE 4 | Growth curve of RS801, RS802, RS803, and RS804. Graph showing growth in 0.2% D-allose modified Welshimer's broth (MWB). Orange line: RS801; blue line: RS802; green line: RS803; purple line: RS804.

TABLE 4 | Growth rates of recombinant strains in D-allose.

Strains	Growth rate ± *SD* (OD600 units/h)
RS801	0.1504 ± 0.006
RS802	0.1486 ± 0.005
RS803	0.1055 ± 0.005
RS804	0.1082 ± 0.004

Biolog phenotype microarrays showed that recombinant strains RS801, RS802, RS803, and RS804 could grow on D-allose basal broth while ICDC-LM188 could not (**Figure 5**). In addition to metabolism of D-allose, the recombinant strains and ICDC-LM188 showed similar utilization of other carbohydrates. This revealed that the six genes (*lmo0734-0739*) represented a specific cassette involved in the carbohydrate metabolism of D-allose.

DISCUSSION

L. monocytogenes has been described as a resistant and ubiquitous bacterium that can survive in harsh conditions. Under nutrient-limited conditions, some subgroup lineage II strains can grow faster than some subgroups of lineage I strains (Tang et al., 2015), and most strains of lineage II were reported to be isolated more frequently from food samples than from patients (Chenal-Francisque et al., 2011; Montero et al., 2015). Some lineage II strains have been reported to form biofilms to adapt to harsh environments (Wong, 1998; Møretrø and Langsrud, 2004). In the previous study of LAEB, we found that some *L. monocytogenes* strains could utilize D-allose for growth (Liu et al., 2017). In the present study, we further confirmed that only strains of lineage II could utilize D-allose as a carbon source, and the other lineages could not, including 91 lineage I strains and 17 lineage III strains. This suggested the presence of special carbohydrate metabolisms in different lineage strains, which might play important roles in the growth and isolation of *L. monocytogenes* in different environments. The hypothesis needs to be studied further in the future.

In our study, a specific gene cassette associated with D-allose was present in lineage II strains, which allowed them to utilize D-allose. The cassette has been reported to be unique to lineage II and atypical 4b, but is absent in other lineages (Doumith et al., 2004; Milillo et al., 2009; Lee et al., 2012). In particular, *lmo0737* is regarded as a specific gene to identify lineage II (Ward et al., 2008). Additionally, the present of the specific gene cassette has been reported to have no impact on invasion and growth in nutrient-deprived conditions in knockout mutants of EGD-e (Milillo et al., 2009). Nonetheless, in this study, we verified that the gene cassette presents in lineage II strains and is related to D-allose metabolism.

Gene *lmo0734* is defined as a Lac I regulator in genome of EGD-e, and it is a member of the D-allose metabolism operon of *L. monocytogenes*. Transcriptional fusions in the *als* operon were Lac^+ in the original background in D-allose metabolism of *E. coli* K12 (Kim et al., 1997). Genes *lmo0735*

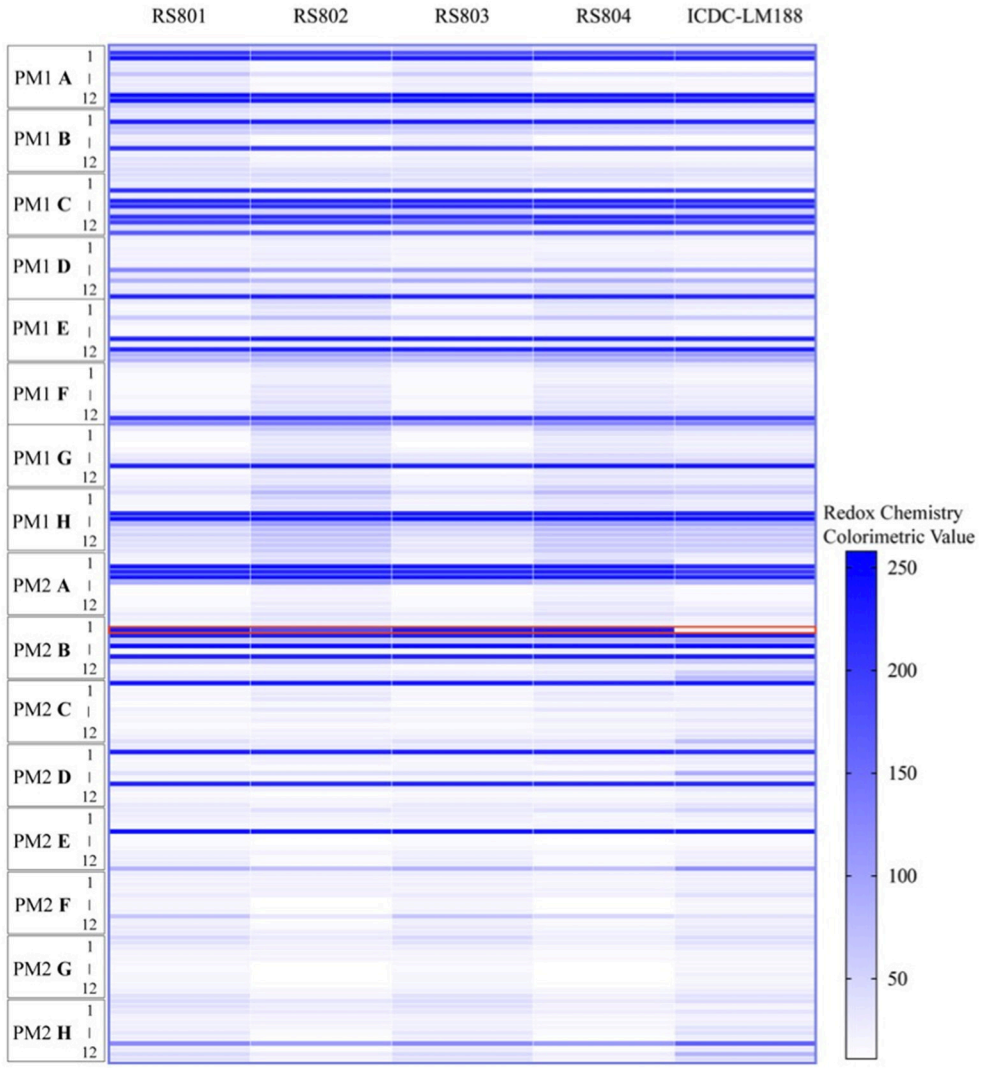


FIGURE 5 | Phenotype microarray analysis of the utilization of carbohydrate by ICDC-LM188, RS801, RS802, RS803, and RS804. All plates were incubated at 37°C for 24 h. PM represent a PM plate, A–H and 1–12, respectively, represent the column and row of the PM plate. Each horizontal line in the heat map shows a type of carbohydrate metabolism. The red border marking the line represents the utilization of D-allose.

and *lmo0736* are regarded to have *alsE* and *alsI* functions. The three genes have been reported to be essential in D-allose metabolism of *L. monocytogenes*. In previous studies, *alsE* was required in D-allose metabolism of *E. coli* K12 and *A. aerogenes*, while *alsI* proved to be essential in *A. aerogenes* and dispensable in *E. coli K12* (Gibbins and Simpson, 1964; Kim et al., 1997; Poulsen et al., 1999). Gene *lmo0737* was annotated as a hypothetical protein belonging to a member of TIM phosphate binding superfamily. Phosphorylase kinase is one of the TIM family (Nagano et al., 1999). D-allose is phosphorylated to D-allulose by the product of *alsK*, a phosphorylase kinase. Gene *lmo0737* might be the equivalent to *alsK*, being responsible for phosphorylation for D-allose. Gene *lmo0738* is annotated as encoding a PTS beta-glucoside transporter subunit IIABC in EGD-e, which has the same function as the products of *alsA*, *B*, and *C*, namely transferring D-allose across the membrane. Genes *lmo0737* and *lmo0738* have been reported to be dispensable in D-allose metabolism in other bacteria. Nonetheless, absence of *lmo0737* and *lmo0738* decreased the growth rate in D-allose MWB medium. The requirement for *lmo0734* to *lmo0736* and the dispensability of *lmo0738* to *lmo0739* suggested that *lmo0734* to *lmo0736* are irreplaceable, while other genes might compensate for the activities of *lmo0737* and *lmo0738*. Interestingly, ribose has been reported as an analog of D-allose, and D-allose can be metabolized via some proteins in the D-ribose metabolic pathway (Kim et al., 1997). Gene *lmo0739* encodes 6-phospho-beta-glucosidase in *L. monocytogenes*. Currently, there are no studies to indicate 6-phospho-beta-glucosidase plays a role in D-allose metabolism. In addition, *lmo0739* is absent in genome of

Listeria ivanovii PAM55, which can utilize D-allose (Liu et al., 2017).

In the transcriptome analysis, there were other DEGs in addition to genes *lmo0734* to *lmo0739*. Among the downregulated genes, genes *lmo0096* and *lmo0097* mainly participate in mannose metabolism. D-allose metabolism is a part of mannose metabolism (KEGG Pathway map00051). D-allose metabolism might have an affect on mannose metabolism. Genes *lmo2761* to *lmo2765* have functions in D-cellobiose metabolism, which is associated with D-glucose (KEGG Pathway map00050), which is isomer of D-allose. However, a clear relationship between D-allose and D-cellobiose metabolism has not been demonstrated. As for the correlation between these genes and D-allose metabolism, it need to be studied in future.

CONCLUSION

Our previous study demonstrated that the utilization of D-allose in a novel enrichment broth could increase the isolation rate of *L. monocytogenes*. The present study firstly proposes D-allose metabolism in *L. monocytogenes* and identified the key genes involved in this metabolism. We determined the distribution of these genes in different lineage strains. We provide preliminarily evidence for the utilization of D-allose in the novel enrichment broth for *L. monocytogenes* lineage II strains. Our results will serve as a reference for the optimization of novel broth (LAEB) to increase the isolation of other lineages strains in the future and for further research into the carbohydrate metabolism of *L. monocytogenes*.

AUTHOR CONTRIBUTIONS

LZ: Designed the project, analyzed data, and wrote manuscript; DL, LL, and YaW: Analyzed data; YiW and CY: Carried out the experiments.

ACKNOWLEDGMENTS

This work was supported by the Mega Project of Research on the Prevention and Control of HIV/AIDS, Viral Hepatitis Infectious Diseases, Ministry of Science and Technology of China [grant number 2013ZX10004-101 to CY]; the State Key Laboratory of Infectious Disease Prevention and Control [grant number 2015SKLID507 to CY]; and the National Institute for Communicable Disease Control and Prevention, China [grant number 2016ZZKTB09 to CY].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00621/full#supplementary-material

REFERENCES

Camilli, A., Portnoy, A., and Youngman, P. (1990). Insertional mutagenesis of *Listeria monocytogenes* with a novel Tn917 derivative that allows direct cloning of DNA flanking transposon insertions. *J. Bacteriol.* 172, 3738–3744. doi: 10.1128/jb.172.7.3738-3744.1990

Chenal-Francisque, V., Lopez, J., Cantinelli, T., Caro, V., Tran, C., Leclercq, A., et al. (2011). Worldwide distribution of major clones of *Listeria monocytogenes*. *Emerg. Infect. Dis.* 17, 1110–1112. doi: 10.3201/eid/1706.101778

Doumith, M., Cazalet, C., Simoes, N., Frangeul, L., Jacquet, C., Kunst, F., et al. (2004). New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays. *Infect. Immun.* 72, 1072–1083. doi: 10.1128/IAI.72.2.1072-1083.2004

Gibbins, L. N., and Simpson, F. J. (1964). The incorporation of D-allose into the glycolytic pathway by *Aerobacter aerogenes*. *Can. J. Microbio.* 10, 829–836. doi: 10.1139/m64-108

Kathariou, S. (2002). *Listeria monocytogenes* virulence and pathogenicity, a food safety perspective. *J. Food Prot.* 65, 1811–1829. doi: 10.4315/0362-028X-65.11.181

Kim, C., Song, S., and Park, C. (1997). The D-allose operon of *Escherichia coli* K-12. *J. Bacteriol.* 179, 7631–7637. doi: 10.1128/jb.179.24.7631-7637.1997

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Lauer, P., Chow, M. Y., Loessner, M. J., Portnoy, D. A., and Calendar, R. (2002). Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* 184, 4177–4186. doi: 10.1128/JB.184.15.4177-4186.2002

Lee, S., Ward, T. J., Graves, L. M., Wolf, L. A., Sperry, K., Siletzky, R. M., et al. (2012). Atypical *Listeria monocytogenes* serotype 4b strains harboring a lineage II-specific gene cassette. *Appl. Environ. Microbio.* 78, 660–667. doi: 10.1128/AEM.06378-11

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Liu, D. (2008). *Handbook of Listeria monocytogenes*. Boca Raton, FL: Taylor & Francis CRC Press.

Liu, D., Wang, Y., Wang, Y., Zhang, L., Luo, L., Liu, K., et al. (2017). Development of a novel *Listeria* enrichment broth for the isolation of pathogenic *Listeria*. *J. Food Prot.* 80, 1768–1776. doi: 10.4315/0362-028X.JFP-16-529

Low, J. C., and Donachie, W. (1997). A review of *Listeria monocytogenes* and listeriosis. *Vet. J.* 153, 9–29. doi: 10.1016/S1090-0233(97)80005-6

Luo, L., Zhang, Z., Wang, H., Wang, P., Lan, R., Deng, J., et al. (2017). A 12-month longitudinal study of *Listeria monocytogenes* contamination and persistence in pork retail markets in China. *Food Control* 76, 66–73. doi: 10.1016/j.foodcont.2016.12.037

Markkula, A., Lindstrom, M., Johansson, P., Bjorkroth, J., and Korkeala, H. (2012). Roles of four putative DEAD-box RNA helicase genes in growth of *Listeria monocytogenes* EGD-e under heat, pH, osmotic, ethanol, and oxidative stress conditions. *Appl. Environ. Microbiol.* 78, 6875–6882. doi: 10.1128/AEM.01526-12

Mead, P. S., Slutsker, L., Dietz, V., Mccaig, L. F., Bresee, J. S., Shapiro, C., et al. (1999). Food-related illness and death in the United States. *Emerg. Infect. Dis.* 5, 607–625. doi: 10.3201/eid0506.990624

Milillo, S. R., Badamo, J. M., and Wiedmann, M. (2009). Contributions to selected phenotypic characteristics of large species-and lineage-specific genomic regions in *Listeria monocytogenes*. *Food Microbiol.* 26, 212–223. doi: 10.1016/j.fm.2008.08.010

Miller, J. M., and Rhoden, D. L. (1991). Preliminary evaluation of Biolog, a carbon source utilization method for bacterial identification. *J. Clin. Microbiol.* 29, 1143–1147.

Montero, D., Bodero, M., Riveros, G., Lapierre, L., Gaggero, A., Vidal, R. M., et al. (2015). Molecular epidemiology and genetic diversity of *Listeria monocytogenes*

isolates from a wide variety of ready-to-eat foods and their relationship to clinical strains from listeriosis outbreaks in Chile. *Front. Microbiol.* 6:384. doi: 10.3389/fmicb.2015.00384

Møretrø, T., and Langsrud, S. (2004). *Listeria monocytogenes*: biofilm formation and persistence in food-processing environments. *Biofilms* 1, 107–121. doi: 10.1017/S1479050504001322

Murata, A., Sekiya, K., Watanabe, Y., Yamaguchi, F., Hatano, N., Izumori, K., et al. (2003). A novel inhibitory effect of D-allose on production of reactive oxygen species from neutrophils. *J. Biosci. Bioeng.* 96, 89–91. doi: 10.1016/S1389-1723(03)90104-6

Nagano, N., Hutchinson, E. G., and Thornton, J. M. (1999). Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci.* 8, 2072–2084. doi: 10.1110/ps.8.10.2072

Orsi, R. H., den Bakker, H. C., and Wiedmann, M. (2011). *Listeria monocytogenes* lineages: genomics, evolution, ecology, and phenotypic characteristics. *Int. J. Med. Microbiol.* 301, 79–96. doi: 10.1016/j.ijmm.2010.05.002

Pine, L., Malcolm, G. B., Brooks, J. B., and Daneshvar, M. I. (1989). Physiological studies on the growth and utilization of sugars by *Listeria* species. *Can. J. Microbiol.* 35, 245–254. doi: 10.1139/m89-037

Pontinen, A., Lindstrom, M., Skurnik, M., and Korkeala, H. (2017). Screening of the two-component-system histidine kinases of *Listeria monocytogenes* EGD-e. LiaS is needed for growth under heat, acid, alkali, osmotic, ethanol and oxidative stresses. *Food Microbiol.* 65, 36–43. doi: 10.1016/j.fm.2017.01.018

Poulsen, T. S., Chang, Y. Y., and Hovejensen, B. (1999). d-Allose Catabolism of *Escherichia coli*: involvement of alsI and regulation of als regulon expression by allose and ribose. *J. Bacteriol.* 181, 7126–7130.

Premaratne, R. J., Lin, W. J., and Johnson, E. A. (1991). Development of an improved chemically defined minimal medium for *Listeria monocytogenes*. *Appl. Environ. Microbiol.* 57, 3046–3048.

Rio, D. C., Ares, M., Hannon, G. J., and Nilsen, T. W. (2010). Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harb. Protoc.* 2010:pdb.prot5439. doi: 10.1101/pdb.prot5439

Roberts, A., Nightingale, K., Jeffers, G., Fortes, E., Kongo, J. M., and Wiedmann, M. (2006). Genetic and phenotypic characterization of *Listeria monocytogenes* lineage III. *Microbiol.* 152(Pt 3), 685–693. doi: 10.1099/mic.0.28503-0

Siddiqi, R., and Khan, M. A. (1982). Vitamin and nitrogen base requirements for *Listeria monocytogenes* and haemolysin production. *Zentralblatt Für Bakteriologie Mikrobiologie Und Hygiene.abt.originale.a Medizinische Mikrobiologie Infektionskrankheiten Und Parasitologie* 253, 225–235.

Tang, S., Wiedmann, M., Gardner, A. L., Brown, A. M., Boor, K. J., and Bergholz, T. M. (2015). Clonal clustering using 10-gene multilocus sequence typing reveals an association between genotype and *Listeria monocytogenes* maximum growth rate in defined medium. *Foodborne Pathog. Dis.* 12, 972–982. doi: 10.1089/fpd.2015.2019

Ward, T. J., Ducey, T. F., Usgaard, T., Dunn, K. A., and Bielawski, J. P. (2008). Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates. *Appl. Environ. Microbiol.* 74, 7629–7642. doi: 10.1128/AEM.01127-08

Wong, A. C. L. (1998). Biofilms in food processing environments. *J. Dairy Sci.* 81, 2765–2770. doi: 10.3168/jds.S0022-0302(98)75834-5

Zhang, W. (2007). *Listeria*, *Listeriosis, and Food Safety, 3rd Edn*. Boca Raton, FL: CRC Press.

Zilelidou, E. A., Rychli, K., Manthou, E., Ciolacu, L., Wagner, M., and Skandamis, P. N. (2015). Highly invasive *Listeria monocytogenes* strains have growth and invasion advantages in strain competition. *PLoS ONE* 10:e0141617. doi: 10.1371/journal.pone.0141617

Zilelidou, E., Karmiri, C. V., Zoumpopoulou, A. G., Mavrogonatou, B. E., Kletsas, C. D., Tsakalidou, C. E., et al. (2016). *Listeria monocytogenes* strains underrepresented during selective enrichment with an ISO method might dominate during passage through simulated gastric fluid and *in vitro* infection of Caco-2 cells. *Appl. Environ. Microbiol.* 82, 6846–6858. doi: 10.1128/AEM.02120-16

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., and Van't Riet, K. (1990). Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.* 56, 1875–1881.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhang, Wang, Liu, Luo, Wang and Ye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 May 2018
doi: 10.3389/fmicb.2018.00951

Distinct Microbial Signatures Associated With Different Breast Cancer Types

*Sagarika Banerjee[1], Tian Tian[2], Zhi Wei[2], Natalie Shih[3], Michael D. Feldman[3], Kristen N. Peck[1], Angela M. DeMichele[4], James C. Alwine[5] and Erle S. Robertson[1]**

[1] Tumor Virology Program, Department of Otorhinolaryngology-Head and Neck Surgery and Microbiology, Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [2] Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, United States, [3] Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [4] Division of Hematology Oncology, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [5] Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

A dysbiotic microbiome can potentially contribute to the pathogenesis of many different diseases including cancer. Breast cancer is the second leading cause of cancer death in women. Thus, we investigated the diversity of the microbiome in the four major types of breast cancer: endocrine receptor (ER) positive, triple positive, Her2 positive and triple negative breast cancers. Using a whole genome and transcriptome amplification and a pan-pathogen microarray (PathoChip) strategy, we detected unique and common viral, bacterial, fungal and parasitic signatures for each of the breast cancer types. These were validated by PCR and Sanger sequencing. Hierarchical cluster analysis of the breast cancer samples, based on their detected microbial signatures, showed distinct patterns for the triple negative and triple positive samples, while the ER positive and Her2 positive samples shared similar microbial signatures. These signatures, unique or common to the different breast cancer types, provide a new line of investigation to gain further insights into prognosis, treatment strategies and clinical outcome, as well as better understanding of the role of the micro-organisms in the development and progression of breast cancer.

Keywords: microbiome, endocrine receptor positive breast cancer, triple negative breast cancer, triple positive breast cancer, HER2 positive breast cancer

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
César López-Camarillo, Universidad Autónoma de la Ciudad de México, Mexico
Michele Maltz, University of Connecticut Mansfield, United States

***Correspondence:**
Erle S. Robertson
erle@pennmedicine.upenn.edu

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *18 October 2017*
Accepted: *24 April 2018*
Published: *15 May 2018*

Citation:
Banerjee S, Tian T, Wei Z, Shih N, Feldman MD, Peck KN, DeMichele AM, Alwine JC and Robertson ES (2018) Distinct Microbial Signatures Associated With Different Breast Cancer Types. Front. Microbiol. 9:951. doi: 10.3389/fmicb.2018.00951

INTRODUCTION

Breast cancer, the second leading cause of cancer death in women, is responsible for the death of 1 in 52 women below 50 years of age (American Cancer Society, 2017). The American Cancer Society estimated that there will be 255,180 new breast cancer cases (2,470 men and 252,710 women) in the US by 2017 (American Cancer Society, 2017). Based on the immuno histochemical classification of hormone receptor status in the cancerous breast cells, there are 4 major groups of breast cancers: endocrine receptor (estrogen or progesterone receptor) positive (abbreviated in the study as BRER), human epidermal growth factor receptor 2 (HER2) positive (abbreviated in the study as BRHR), triple positive (estrogen, progesterone and HER2 receptor positive) (abbreviated in the study as BRTP) and triple negative (absence of estrogen, progesterone and HER2 receptors) (abbreviated in the study as BRTN) (Schnitt, 2010; American Cancer Society, 2017). These four types have specific prognoses and responses to therapy. Specifically, the hormone receptor positive breast cancers

(BRER, BRTP) respond to endocrine therapy and show better prognosis, while the hormone receptor negative types (BRHR, BRTN) are more aggressive, non-responsive to endocrine therapy and have poor prognosis (Schnitt, 2010). BRTN cancer is seen in 15–20% of breast cancer patients, is the most aggressive of all the breast cancers, is unresponsive to treatment, highly angiogenic, proliferative and has the lowest survival rate (Siegel et al., 2016).

However in the recent times the global gene expression studies have identified breast tumors further into distinct molecular classes based on the expression level of endocrine receptors, proliferative genes, oncogenes and other genes; luminal A (ER+/PR+ and Ki67 high), luminal B (ER+/PR+, Ki67 low or, ER+/PR+/HE R2+), HER2+, basal (ER-/PR-/basal myoepithelial markers high/ EGFR+), and normal breast-like (ER-/PR-/basal myoepithelial markers-/EGFR-) (Yersal and Barutca, 2014).

Among the risk factors to develop cancer in general, infectious agents are known to be the third highest after tobacco usage and obesity, contributing 15–20% of cancer incidence (Morales-Sanchez and Fuentes-Panana, 2014). Age and genetic pre-disposition are also known cancer risk factors; however, the majority of cancers have unknown etiology (Madigan et al., 1995). Recent studies of microbiome dysbiosis in human health suggest specific changes in the microbiome in a number of disease states (Turnbaugh et al., 2006; Xuan et al., 2014; Chen and Wei, 2015), including cancer (Sheflin et al., 2014; Xuan et al., 2014). Further, studies have suggested the association of particular micro-organisms with specific cancers (Banerjee et al., 2015, 2017a,b; Chen and Wei, 2015). Thus, a distinct microbiome may contribute to the cause or development of cancer. Conversely, the tumor micro-environment may provide a specialized niche in which these viruses and microorganisms may persist. In either case, cancer-type specific microbial signatures may provide clues for early diagnosis, prognosis and the design of treatment strategies.

We have recently identified a distinct microbial signature associated with triple negative breast cancer (Banerjee et al., 2015). In the present study we asked whether the microbial signatures associated with BRTN are shared by other breast cancer types, or do different breast cancer types have unique signatures. To study this we screened BRTN, BRTP, BRER, and BRHR samples using PathoChip, a pan-pathogen array containing oligonucleotide probes for the detection of all known, sequenced viruses, as well as known human bacterial, parasitic, and fungal pathogens. Additionally, PathoChip contains viral family specific conserved probes that allow for detection of uncharacterized members of the viral families (Baldwin et al., 2014). The PathoChip screen includes a whole genome and transcriptome amplification step that allowed detection of very low copy number of both DNA and RNA viruses and micro-organisms from the cancer tissues (Baldwin et al., 2014; Banerjee et al., 2015). Our analyses now show distinct microbial signatures for BRTN and BRTP samples, while the BRER and BRHR samples had similar microbial signatures.

MATERIALS AND METHODS

Study Samples

The study was approved by the institutional review board at the University of Pennsylvania (Protocol number 819358). All methods were performed in accordance with the relevant guidelines and regulations and reviewed by resident pathologists at the UPENN hospital. In the present study, 50 endocrine receptor (estrogen or progesterone receptor) positive (abbreviated as BRER in the study), 34 human epidermal growth factor receptor 2 (HER2) positive (abbreviated as BRHR in the study), 24 triple positive (estrogen, progesterone and HER2 receptor positive) (abbreviated as BRTP in the study) and 40 triple negative (absence of estrogen, progesterone and HER2 receptors, abbreviated as BRTN in the study) breast cancer tissues were included along with 20 breast control samples from healthy individuals. Due to HIPAA regulations, we could not obtain any information regarding the type of treatment these breast cancer patients received, or, if they were new patients. These tissues were obtained as de-identified archived samples. Tumors needing macro-dissection were received in the form of 10 μm sections on glass slides with marked guiding H&E slides, while tumors that did not require macro-dissection were received as 10 μm paraffin rolls. The 20 non-matched control tissues were derived from breast reduction surgeries and obtained as 10 μm paraffin rolls. Our resident pathologist reviewed case history, confirmed the tumor types and demarcated the cancer cells. All the samples were de-identified FFPE (formalin fixed paraffin embedded) samples of breast tumors or controls, and were received from the Abramson Cancer Center Tumor Tissue and Biosample Core. Extreme care was taken to avoid contamination during cutting of the FFPE sections (Banerjee et al., 2015). For each samples, microtome and other equipments were cleaned with 70% ethanol. Further, a new blade was used to prepare and cut each sample, and the area was also de-contaminated before cutting each sample (Banerjee et al., 2015).

Pathochip Design, Sample Preparation, and Microarray Processing

The PathoChip Array design has been previously described in detail (Baldwin et al., 2014). The PathoChip probes were generated *in silico* using the genome sequences of all known viruses, as well as known human bacterial, parasitic and fungal pathogens. The PathoChip comprises 60,000 probe sets manufactured as SurePrint glass slide microarrays (Agilent Technologies Inc.), containing 8 replicate arrays per slide. Each probe is a 60-nt DNA oligomer that targets multiple genomic regions of the micro-organisms, for example, 18S rRNA gene, 5.8S rRNA gene, 28S rRNA gene, ITS1 and ITS2 for parasite detections, 16S rRNA gene for bacteria detections, 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2 and 26S rRNA gene to detect fungi, and conserved and specific viral genes to detect viral families and specific viruses. PathoChip screening was done using both DNA and RNA extracted from formalin-fixed paraffin-embedded (FFPE) tumor tissues as described previously (Baldwin et al., 2014; Banerjee et al., 2015). The quality of

the extracted nucleic acids was determined by agarose gel electrophoresis and the A260/280 ratio. The extracted RNA and DNA were subjected to whole genome and transcriptome co-amplification (WTA) as previously described (Banerjee et al., 2015). A non-template control (RNAase/DNAase free water) was used during the WTA step, to determine if any contamination was present during the amplification step. The quality of the WTA products was determined by agarose gel electrophoresis. Human reference RNA and DNA were also extracted from the human B cell line, BJAB and were used for WTA as previously described (Banerjee et al., 2015). The WTA products were purified, (PCR purification kit, Qiagen, Germantown, MD, USA); the WTA products from the cancers were labeled with Cy3 and those from the human reference DNA were labeled with Cy5 (SureTag labeling kit, Agilent Technologies, Santa Clara, CA). The labeled DNAs were purified and hybridized to the PathoChip as described previously (Banerjee et al., 2015). Post-hybridization, the slides were washed, scanned and visualized using an Agilent SureScan G4900DA array scanner (Banerjee et al., 2015).

The question of potential contamination of FFPE blocks or during processing is always a concern. In these experiments all samples were handled and processed in the pathology laboratory using standard aseptic conditions. Likewise the preparation of the DNA and RNA from the samples was done in a dedicated laboratory under established condition designed to minimize laboratory contamination.

Microarray Data Extraction and Statistical Analysis

Agilent Feature Extraction software (Baldwin et al., 2014; Banerjee et al., 2015) was used to extract the raw data from the microarray images. We used the R program for normalization and data analyses (R Core Team, 2015). We calculated scale factor using the signals of green (Cy3) and red (Cy5) channels for human probes. Scale factors are the sum of green/sum of red signal ratios of human probes. Then we used scale factors to obtain normalized signals for all other probes. For all probes except human probes, normalized signal is log2 transformed of green signals/scale factors modified red signals (log2 g – scale factor * log2 r). On the normalized signals, one-sided *t*-test is applied to select probes significantly present in cancer samples by comparing cancer samples vs. controls. The significance cut-off was log2 fold change of signal $\geq$1 and adjusted *p*-value (all *p*-value were adjusted via Benjamini–Hochberg procedure for controlling FDR) $\leq$0.01, control prevalence $\leq$25%, case prevalence $\geq$40%. Prevalence is calculated as the detection of the microbial signatures in the cancer and in the control samples as percentages. For a particular microbial signature with multiple probes, we calculated the prevalence of that signature by calculating the maximum number of samples that contained even one of the probes of that signature.

The cancer samples were also subjected to hierarchical clustering, based on the detection of microbial signatures in the samples. We used hierarchical clustering technic (Euclidean distance, complete linkage, normalized hybridization signals not scaled) to cluster samples which were represented as heatmaps (Kolde, 2015). Then clusters were further validated by CH-index (Calinski and Harabasz index) which is implemented in the R package as NbClust (Charrad et al., 2014). CH- index is a cluster index that maximizes inter-cluster distances and minimizes intra-cluster distances. We calculated the possible cluster solution that would maximize the index values to achieve the best clustering of the data. Statistical significance between different groups was determined using the two-sided *t*-test.

Based on the clinical outcomes of the different breast cancer patients, the cases for each breast cancer types were divided into two groups: alive and deceased (with severe outcomes) (Supplementary Table S4). We calculated the proportion of the two groups in each of the hierarchical cluster/sub-cluster of the 4 breast cancer types. One sided *t*-test was also done to compare the differences of average hybridization signals of organisms in these two groups. Nominal *p*-values along with log fold change were calculated. Microbial signatures that were detected with significantly (nominal *p*-value $<$ 0.05) higher average hybridization signal in the deceased cases or in the patients that survived were selected to do box plots for representation of the data. Also differences in the detection of some signatures which were not statistically significant between the different outcomes, but showed some trend were plotted as well. Where, the *p*-value $>$ 0.05, we can only suggest that higher detection of those microbial signatures with either of the outcome, is only seen as a trend.

PCR Validation of Pathochip Results

PCR primers from the conserved and/or specific regions of the micro-organisms detected by PathoChip screen were used. The PCR amplification reaction mixtures for each reaction contained 200-400 ng of WTA product and 20 pM each of forward and reverse primers (**Table 7**), 300 μM of dNTPs and 2.5U of LongAmp Taq DNA polymerase (NEB). DNA was denatured at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, different annealing temperature for different set of primers for 30–45 s, and 65°C for 30 s. The PCR conditions for each of the primer sets are mentioned in **Table 7**.

RESULTS

Microbial Signatures Associated With Different Breast Cancer Types

Unique and common microbial signatures associated with different breast cancer types have been listed in **Table 1** and are represented in **Figures 1A**, **2B**, **3C,F**. To establish the microbial signatures in the cancers we compared the average hybridization signal for each probe in the cancer samples vs. the controls. Those probes that detected significant higher hybridization signals in the cancer samples (*p*-value $<$ 0.05, log2 fold change in hybridization signal $>$ 1), present in atleast 40% of the cancer samples, and $\leq$25% of the controls were considered in the present study. A stringent cut-off criteria of microbial signature detections only in the cancers and not (0% prevalence) in the controls lead mostly to detect less number of probes for a

TABLE 1 | Unique and common microbial signatures in 4 breast cancer types; the endocrine receptor positives (BRER), human epidermal growth factor receptor 2 positives (BRHR), triple positives (BRTP) and the triple negatives (BRTN).

Cancer types	Viral signatures	Bacterial signatures	Fungal signatures	Parasitic signatures
BRER		Arcanobacterium, Bifidobacterium, Cardiobacterium, Citrobacter, Escherichia	Filobasidilla, Mucor, Trichophyton	Brugia, Paragonimus
BRHR	Nodaviridae	Streptococcus	Epidermophyton, Fonsecaea, Pseudallescheria	Balamuthia
BRTP	Birnaviridae, Hepeviridae	Bordetella, Campylobacter, Chlamydia, Chlamydophila, Legionella, Pasteurella	Penicillium	Ancylostoma, Angiostrongylus, Echinococcus, Sarcocystis, Trichomonas, Trichostrongylus
BRTN		Aerococcus, Arcobacter, Geobacillus, Orientia, Rothia	Alternaria, Malassezia, Piedraia, Rhizomucor	Centrocestus, Contracaecum, Leishmania, Necator, Onchocerca, Toxocara, Trichinella, Trichuris
BRTN+BRER		Mycoplasma		
BRTN+BRTP				Babesia, Mansonella, Schistosoma
BRER+BRTN			Fusarium	
BRER+BRHR	Caliciviridae	Acinetobacter, Alcaligenes, Anaplasma, Eikenella, Fusobacterium, Kingella, Lactococcus, Salmonella		Ascaris
BRHR+BRTP	Arteriviridae	Borrelia, Klebsiella	Coccidioides	
BRHR+BRTN				Strongyloides
BRER+BRTP	Hepadnaviridae	Helicobacter, Neisseria, Pediococcus, Prevotella, Propionibacterium, Treponema	Cunninghamella, Geotrichum	Hartmannella, Hymenolepis, Macracanthorhynchus
BRTN+BRER+BRTP		Stenotrophomonas	Pleistophora, Rhodotorula	Naegleria
BRTN+BRER+BRHR		Brucella, Caulobacter, Peptoniphilus		
BRHR+BRTN+BRTP		Haemophilus		
BRHR+BRTP+BRER	Astroviridae, Circoviridae, Orthomyxoviridae, Polyomaviridae, Togaviridae	Agrobacterium, Clostridium, Corynebacterium, Erysipelothrix, Lactobacillus, Listeria, Shigella, Staphylococcus	Ajellomyces, Aspergillus, Candida, Trichosporon	Entamoeba, Plasmodium, Thelazia
BRER+BRHR+BRTP +BRTN	Adenoviridae, Anelloviridae, Arenaviridae, Bunyaviridae, Coronaviridae, Filoviridae, Flaviviridae, Herpesviridae, Iridoviridae, Papillomaviridae, Paramyxoviridae, Parvoviridae, Picornaviridae, Poxviridae, Reoviridae, Retroviridae, Rhabdoviridae	Actinomyces, Bartonella, Brevundimonas, Coxiella, Mobiluncus, Mycobacterium, Rickettsia, Sphingomonas		

particular signature for some signatures, not for all, but not that the majority signatures detected with our accepted cut-off was lost (Supplementary Figures S2a–d).

We further averaged the hybridization signals of all the significant probes for each microbial genera and viral families, shown in the **Figures 1–3**. Supplementary Table S1 shows the average hybridization signals of the probes of microorganisms significantly detected in the cancers vs. the controls, with respective adjusted *p*-values with multiple corrections. Supplementary Table S2 shows the proportion of probes that were detected significantly in each of the breast cancer types vs. the controls. Supplementary Figure S1 shows the average fold change in hybridization signal intensity for the significantly detected probes of each of the signatures detected in the different breast cancer types over their respective signals in the control breast samples. Additionally, we calculated the percent prevalence of the significant microbial signatures in the cancer samples, which indicate how prevalent a significant virus or microorganism signature is in the cancer samples regardless of the hybridization intensity.

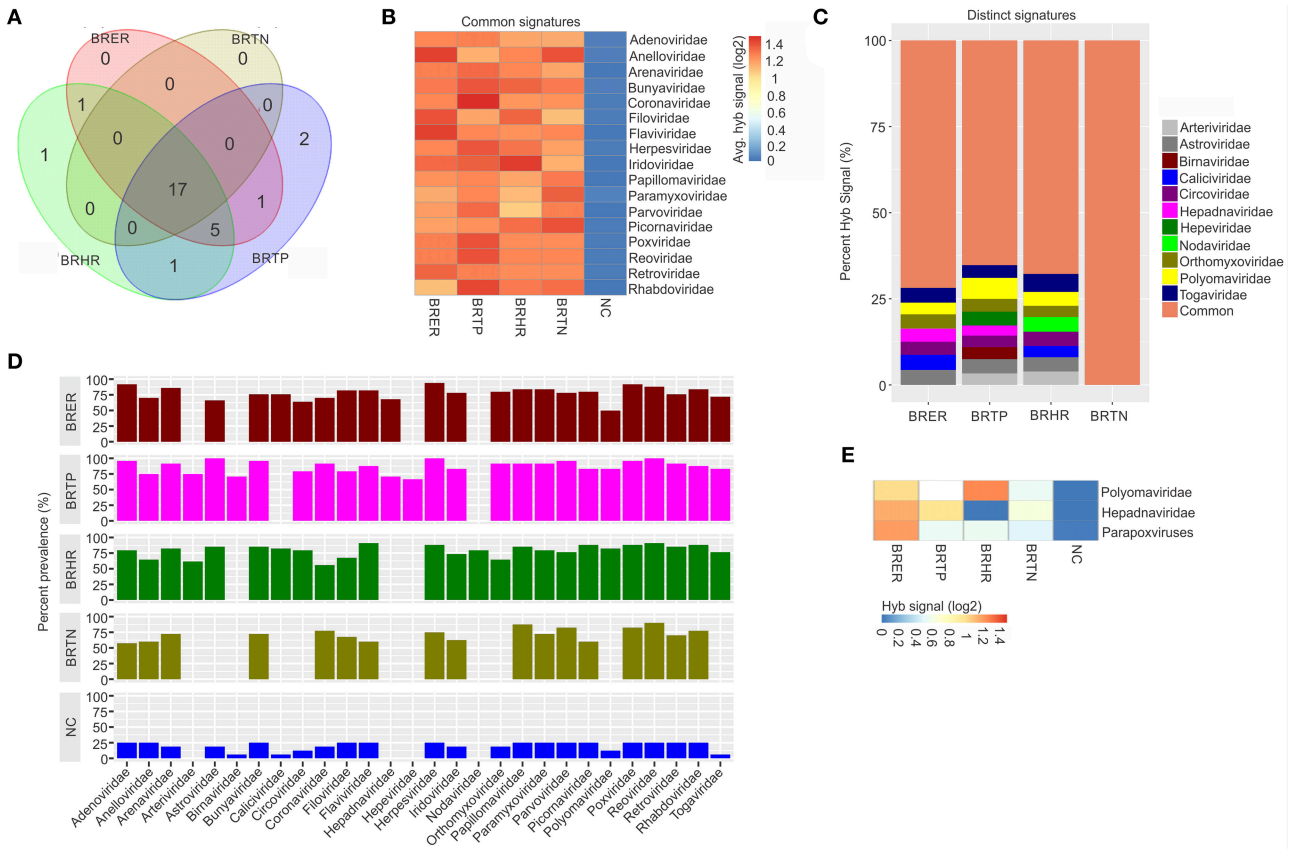


FIGURE 1 | Viral signatures associated with different breast cancer types. Among the breast cancer types, the endocrine receptor (estrogen/progesterone) positives are abbreviated as BRER, human epidermal growth factor receptor 2 positives are abbreviated as BRHR, triple positives (estrogen, progesterone and HER2 receptor positive) are abbreviated as BRTP and the triple negatives (absence of estrogen, progesterone and HER2 receptors) are abbreviated as BRTN. The normal breast control samples obtained from healthy individuals are abbreviated as NC. **(A)** Venn diagram showing the common and unique viral signatures in the 4 types of breast cancers. **(B)** The heat map of common viral signatures in the 4 breast cancer types. **(C)** Relative hybridization signals of viral probes detected in breast cancer types. For example, hybridization signals for Polyomaviridae probes were 4, 6, and 3% of the total hybridization signals detected in BRER, BRTP, and BRHR respectively. **(D)** Prevalence of viral signatures in 4 breast cancer types. Since the hybridization signals for Polyomaviridae, Hepadnaviridae and Parapoxviridae were lower than the cut-off (log2 fold change in hybridization signal >1) in one or more breast cancer types they are depicted as negative in this figure. However, **(E)** shows the heat map of hybridization signals for those viral signatures to be still significantly higher in the cancers when compared to the control.

Viral Signatures Associated With Different Breast Cancer Types

Significant hybridization (described above), at levels above the controls, was detected for 28 viral families among the four breast cancer types (**Figures 1A,D**). Of these, 17 viral families were detected with significantly higher hybridization signals in greater than 50% of the samples representing all 4 breast cancer types, as compared to the controls (**Figures 1B,D**). They include signatures of Adenoviridae, Anelloviridae, Arenaviridae, Bunyaviridae, Coronaviridae, Filoviridae, Flaviviridae, Herpesviridae, Iridoviridae, Papillomaviridae, Paramyxoviridae, Parvoviridae, Picornaviridae, Poxviridae, Reoviridae, Retroviridae, and Rhabdoviridae (**Figure 1B**). Importantly, in examining the percent hybridization signal (**Figure 1C**) and percent prevalence (**Figure 1D**) we find that there were a number of viral families significantly detected only in a subset of breast cancer types. Specifically, the signatures for Birnaviridae and Hepeviridae were only detected in BRTP; and Nodaviridae only in BRHR (**Figures 1C,D**). Further examination of the percent prevalence (**Figure 1D**), shows that BRTN samples show low or no prevalence of Arteriviridae, Astroviridae, Birnaviridae, Caliciviridae, Circoviridae, Hepadnaviridae, Nodaviridae, Orthomyxoviridae, Polyomaviridae, and Togaviridae; BRHR samples show low or no prevalence of Birnaviridae, Hepadnaviridae, and Hepeviridae; BRTP samples show low or no prevalence of Caliciviridae and Nodaviridae; and BRER samples show low or no prevalence of Arteriviridae, Birnaviridae, Hepeviridae, and Nodaviridae.

Hybridization signal intensity offers an additional way to compare the data. Here we noted marked differences for specific viral families between the different breast cancer types. For example, probes for polyomaviridae were detected with the highest hybridization signal in the BRHRs, followed by BRERs and BRTPs (**Figure 1E**). Polyomaviridae were detected

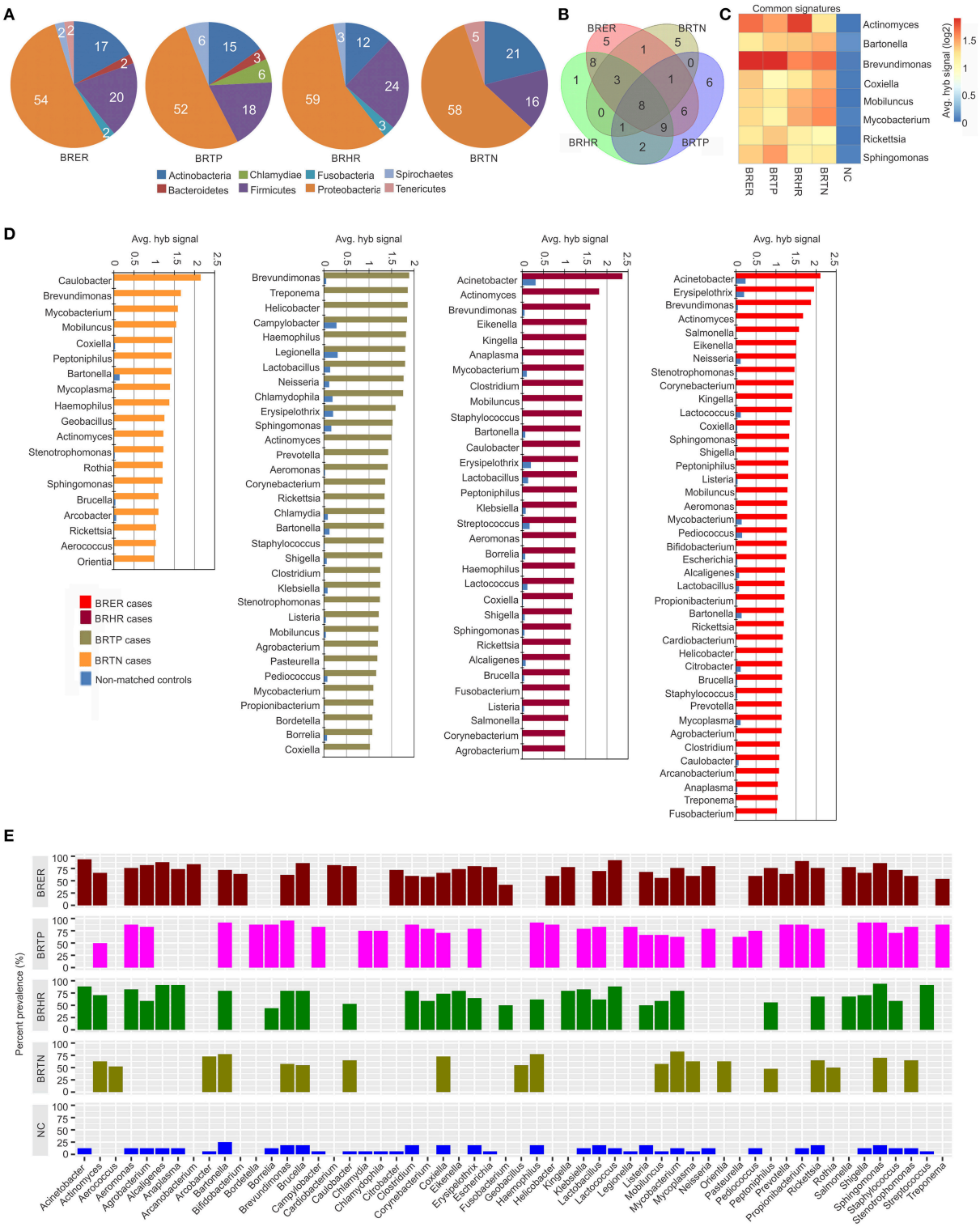


FIGURE 2 | Bacterial signatures associated with different breast cancer types. Among the breast cancer types, the endocrine receptor (estrogen/progesterone) positives are abbreviated as BRER, human epidermal growth factor receptor 2 positives are abbreviated as BRHR, triple positives (estrogen, progesterone and HER2 receptor positive) are abbreviated as BRTP and the triple negatives (absence of estrogen, progesterone and HER2 receptors) are abbreviated as BRTN. The normal breast control samples obtained from healthy individuals are abbreviated as NC. **(A)** Bacterial phyla associated with breast cancer types. **(B)** Venn diagram showing the common and unique bacterial signatures in the 4 types of breast cancers **(C)**. The heat map of common viral signatures in the 4 breast cancer types. **(D)** Hybridization signals of bacterial probes detected in breast cancer types. **(E)** Prevalence of bacterial signatures in 4 breast cancer types.

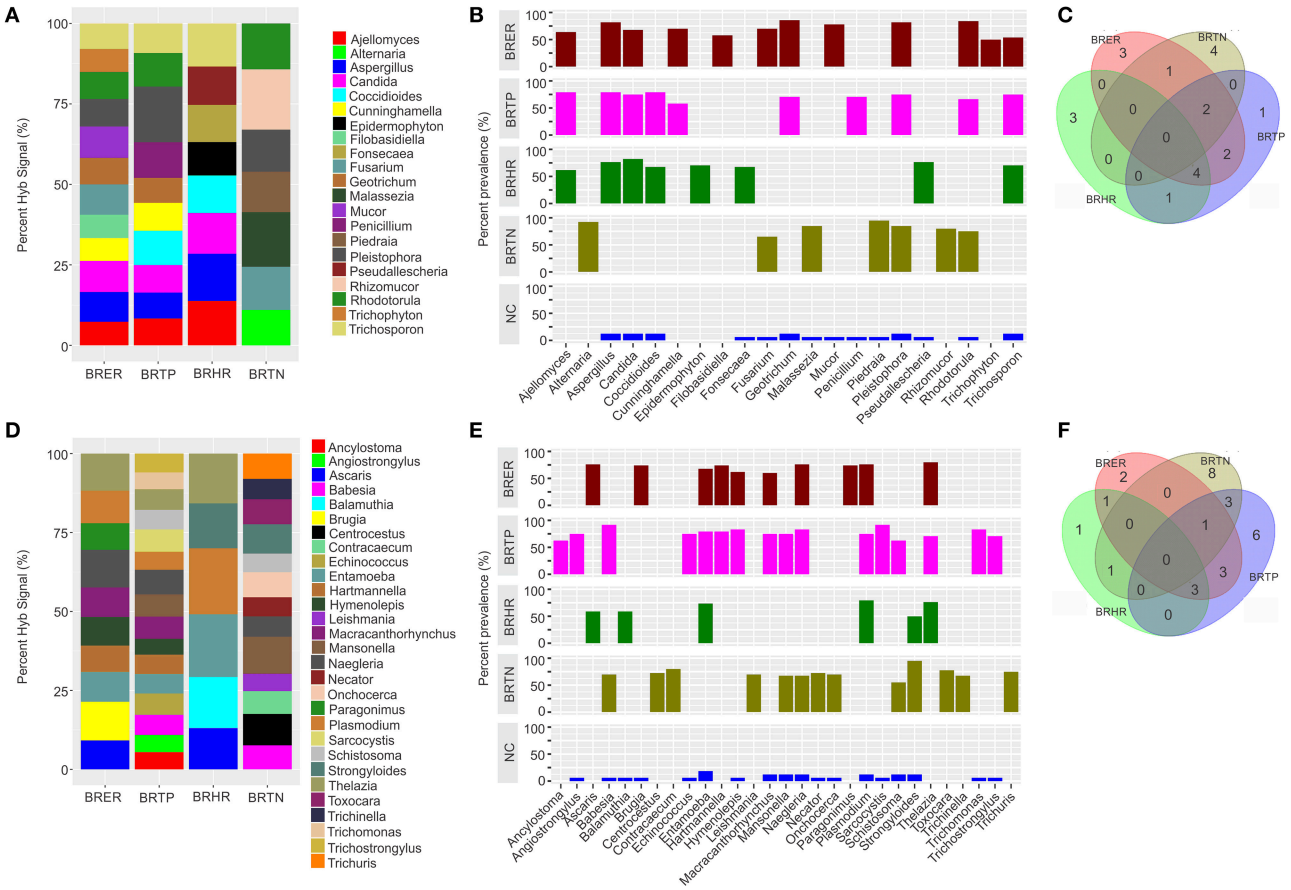


FIGURE 3 | Fungal and parasitic signatures associated with different breast cancer types. Among the breast cancer types, the endocrine receptor (estrogen/progesterone) positives are abbreviated as BRER, human epidermal growth factor receptor 2 positives are abbreviated as BRHR, triple positives (estrogen, progesterone, and HER2 receptor positive) are abbreviated as BRTP and the triple negatives (absence of estrogen, progesterone, and HER2 receptors) are abbreviated as BRTN. The normal breast control samples obtained from healthy individuals are abbreviated as NC. **(A)** Relative hybridization signals of fungal probes detected in breast cancer types. For example hybridization signals for *Ajellomyces* were 7, 8, and 14% of the total hybridization signals detected in BRER, BRTP, and BRHR respectively, and that of *Rhizomucor* is 19% of the hybridization signals detected in BRTN. **(B)** Prevalence of viral signatures in 4 breast cancer types. **(C)** Venn diagram showing the common and unique fungal signatures in the 4 types of breast cancers. **(D)** Relative hybridization signals of parasitic probes detected in breast cancer types. For example hybridization signals for *Plasmodium* were 10, 6, and 21% of the total hybridization signals detected in BRER, BRTP, and BRHR respectively, and that of *Mansonella* is 7 and 12% of the hybridization signals detected in BRTP and BRTN respectively. **(E)** Prevalence of parasitic signatures in 4 breast cancer types. **(F)** Venn diagram showing the common and unique parasitic signatures in the 4 types of breast cancers.

in the BRTNs compared to the controls; however, at a lower hybridization signal (log2 fold change in hybridization signal = 0.4–1; **Figure 1E**) which is below the cut-off to consider the signal positive, thus polyomaviridae are not shown to be present in the BRTNs in **Figure 1C** or **Figure 1D**. Similarly, probes of Hepadnaviridae were significantly detected with low hybridization signal in the BRTNs (**Figure 1E**), while detected with higher hybridization signal intensity (log2 fold change in hybridization signal >1) in the BRERs and BRTPs (**Figure 1E**).

Signatures of Herpesviridae, Adenoviridae and Poxviridae were detected in >90% of the BRER samples screened (**Figure 1D**), while the highest hybridization signal was detected for Anelloviridae and Flaviviridae (**Figure 1B**). Signatures of Astroviridae, Herpesviridae, Reoviridae were detected in all of the BRTP samples tested (**Figure 1D**), with the highest hybridization signal detected for Polyomaviridae signatures (**Figure 1C**). For BRHR samples, signatures of Reoviridae and Flaviviridae were detected in >90% of the samples screened (**Figure 1D**), with signatures of Togaviridae showing the highest hybridization signal (**Figure 1C**). Among the BRTN samples, we detected signatures of Reoviridae in 90% of the samples screened (**Figure 1D**), with signatures of Picornaviridae and Anelloviridae with the highest hybridization signal (**Figure 1B**).

Probes of Poxviridae family were detected significantly in >80% of all the breast cancer types analyzed. Interestingly, probes of Parapoxviridae were detected significantly with high hybridization signal intensity in BRER cancers vs. the controls (**Figure 1E**). Probes of Parapoxviridae were also detected significantly in the other 3 types of breast cancers compared to the controls, but showed much lower hybridization signal intensity

for those probes (log2 fold change in hybridization signal ∼0.5) (**Figure 1E**).

The data show that the cancer samples as a whole have a robust viral signature. However, there are significant and defining differences between the four types with BRTN having the least complex viral signature.

In the healthy control breast tissues, signatures of the viral families Arteriviridae, Hepadnaviridae, Hepeviridae, and Nodaviridae were not detected which were detected in one or more of the cancer types (**Figure 1D**).

Bacterial Signatures Associated With Different Breast Cancer Types

Figures 2A–E shows the analysis of bacterial signatures in the 4 breast cancer types. Significant hybridization, above the levels of the controls, was detected for 56 bacterial genera; the majority (50–60%) was proteobacteria, the major group of gram negative bacteria. These phyla partitioned into bacterial signatures unique to each cancer types, as well as signatures that were common to multiple breast cancer types (**Table 1**, **Figures 2B–D**). Significant hybridization signals common to all 4 breast cancer types were detected for *Actinomyces, Bartonella, Brevundimonas, Coxiella, Mobiluncus, Mycobacterium, Rickettsia,* and *Sphingomonas* (**Figures 2B,C**).

The marked diversity in bacterial signatures between the breast cancer types are shown in **Figure 2B**. We identified distinct bacterial signatures uniquely associated with each type of breast cancer analyzed. In this regard BRTN had the least complex bacterial signature, while BRER is the most complex (**Figures 2D,E**). Signals for *Arcanobacterium, Bifidobacterium, Cardiobacterium, Citrobacter, Escherichia* were significantly detected in the BRER samples compared to the controls, while those of *Bordetella, Campylobacter, Chlamydia, Chlamydophila, Legionella,* and *Pasteurella* were significantly associated with the BRTPs. Signals for *Streptococcus* were detected significantly in the BRHRs, whereas, *Aerococcus, Arcobacter, Geobacillus, Orientia,* and *Rothia* were found associated with the BRTNs.

Hybridization signal intensity again provides an additional view of the complexity of the bacterial community and its diversity among the different breast cancers (**Figures 2C,D**). Signals for *Brevundimonas* were detected with higher average hybridization signals in the endocrine receptor positive BRER and BRTP compared to the endocrine receptor negative BRHR and BRTN (**Figures 2C,D**). Hybridization signals of *Mobiluncus* and *Mycobacterium* were predominantly detected in the endocrine receptor negative samples.

Bacterial signatures of *Actinomyces* were detected in all 4 cancer types; however their hybridization signal intensity was markedly lower in the BRTN samples (**Figure 2C**). Similarly, *Bartonella* was significantly detected in all cancer types, but its hybridization signal intensity was markedly lower in the BRER samples compared to the others (**Figure 2C**). The bacterial probes detected with the highest hybridization signals were those for *Acinetobacter* in BRER and BRHR samples, *Brevundimonas* in BRTP samples and *Caulobacter* in BRTN samples (**Figure 2D**). As in the case of the viruses our data show that the cancer samples have a robust bacterial signature with significant and defining differences between the four breast cancer types. The healthy control samples did not have some of the bacterial signatures that were detected in one or more of the cancer types, namely, *Actinomyces, Aerococcus, Arcanobacterium, Bifidobacterium, Bordetella, Cardiobacterium, Corynebacterium, Eikenella, Fusobacterium, Geobacillus, Helicobacter, Kingella, Orientia, Pasteurella, Peptinophilus, Prevotella, Rothia, Salmonella,* and *Treponema* (**Figure 2E**).

Fungal Signatures Associated With Different Breast Cancer Types

Significant hybridization, above the levels of the controls, was detected for 21 different genera of fungi among the 4 types of breast cancer (**Figures 3A,B**). Interestingly, none of these families were detected in all four cancer types (**Figures 3B,C**). In fact the fungi signatures for each type of breast cancer were relatively unique; only 7 fungal families (*Aspergillus, Candida, Coccidioides, Cunninghamella, Geotrichum, Pleistophora,* and *Rhodotorula*) were detected in more than one type of breast cancer. The receptor positive cancer samples (BRER and BRTPs) showed much more complex fungal diversity than the BRTN samples (**Figures 3A,B**). **Table 1** and **Figure 3C** show the unique fungal signatures associated with different breast cancer types. Fungal signatures of *Filobasidiella, Mucor*, and *Trichophyton* were found to be significantly associated with BRER samples, *Penicillium* with BRTP samples, *Epidermophyton, Fonsecaea, Pseudallescheria* with BRHR samples and *Alternaria, Malassezia, Piedraia,* and *Rhizomucor* with BRTN samples. Fungal signatures of *Ajellomyces, Alternaria, Cunninghamella, Epidermophyton, Filobasidiella, Rhizomucor,* and *Trichophyton* detected in one or more breast cancer types were not detected in the healthy controls (**Figure 3B**).

Parasitic Signatures Associated With Different Breast Cancer Types

Significant hybridization, above the levels of the controls, was detected for 29 different genera of parasites among the 4 types of breast cancer (**Figures 3D,E**). As in the case of the fungi, no single genus of parasite was significantly detected in all four breast cancer types (**Figures 3E,F**). Each cancer showed a relatively distinct parasitic signature pattern, with BRHR showing the least diverse signatures. **Table 1** and **Figure 3F** shows the unique and common parasitic signatures among the different breast cancer types.

Analysis of hybridization signal intensity in **Figure 3D** shows that *Plasmodium* was detected with the highest hybridization signal in the BRHR samples and also detected in the BRER samples and BRTP samples but not in BRTN samples. In BRTN the highest hybridization signal intensity was detected for the probes of *Mansonella* followed by *Centrocestus*, whereas *Strongyloides* was detected in almost all of the BRTN samples. *Naegleria* was detected with the highest hybridization signal intensity in BRTP (**Figure 3D**) while *Sarcocystis* and *Babesia* were detected in 92% of BRTP samples (**Figure 3E**). Among the BRER samples, *Brugia* showed the highest hybridization

signal intensity (**Figure 3D**), while *Thelazia* showed the highest prevalence (**Figure 3E**). Signatures of *Brugia* and *Paragonimus* were only detected in BRER samples (**Table 1**, **Figures 3D,E**). *Ancylostoma, Angiostrongylus, Echinococcus, Sarcocystis, Trichomonas, Trichostrongylus* were found uniquely associated with BRTP samples (**Table 1**, **Figures 3D–F**). *Balamuthia* signatures were associated significantly with BRHR samples, and that of *Centrocestus, Contracaecum, Leishmania, Necator, Onchocerca, Toxocara, Trichinella,* and *Trichuris* were detected significantly only with BRTN samples (**Table 1**, **Figures 3D–F**). Signatures of *Ancylostoma, Ascaris, Centrocestus, Contracaecum, Hartmanella, Leishmania, Paragonimus, Thelazia, Toxocara, Trichinella, Trichuris* detected in one or more cancer types were not detected in the healthy controls (**Figure 3E**).

Hierarchical Clustering of the Breast Cancer Samples Based on the Detection of Microbial Signatures

Using the hierarchical clustering analysis based on the detection of microbial signatures associated with the 4 breast cancer types we determine if the breast cancer types fell into any unique and identifiable clusters. While this analysis identified distinct clusters in each of the breast cancer types based on the detection of their microbial signature patterns (**Figures 4A–D**), it also defined the distinct microbial signature pattern found in BRTNs and BRTPs whereas, BRER and BRHR shared similar microbial signatures (**Figure 4E**).

Individually, the different BC types fell into distinct microbial signature clusters. BRER samples fell into 2 distinct clusters 1ER

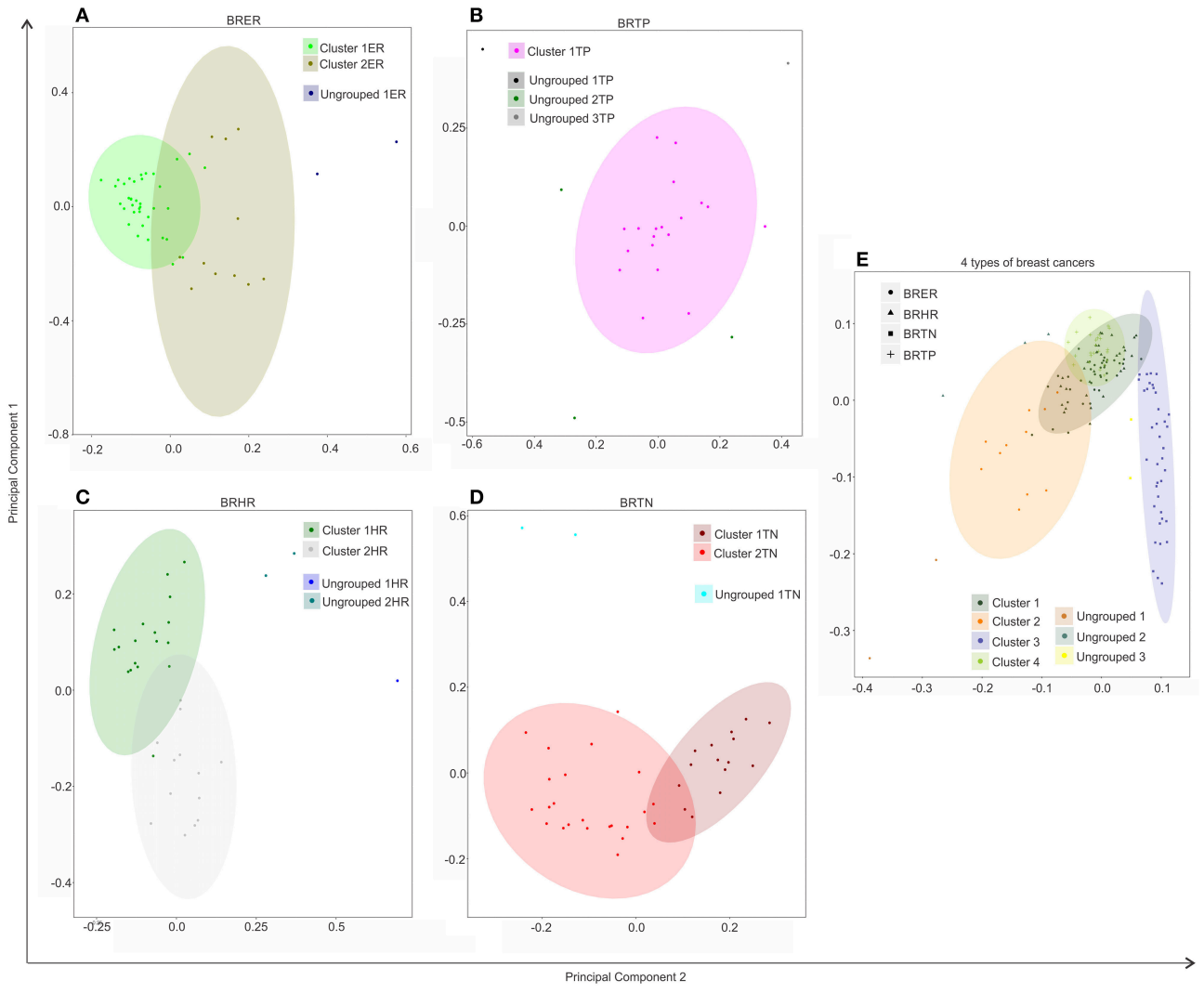


FIGURE 4 | Hierarchical clustering of BRER **(A)**, BRTP **(B)**, BRHR **(C)**, BRTN **(D)**, and all 4 breast cancer types **(E)** based on their microbial signature detection pattern. The endocrine receptor (estrogen/progesterone) positive breast cancers are abbreviated as BRER, human epidermal growth factor receptor 2 positive breast cancers are abbreviated as BRHR, triple positive (estrogen, progesterone, and HER2 receptor positive) breast cancers are abbreviated as BRTP and the triple negative (absence of estrogen, progesterone, and HER2 receptors) breast cancers are abbreviated as BRTN.

and 2ER, along with 2 ungrouped samples (ungrouped 1ER) (**Figure 4A**). Samples grouped in Cluster 1ER and 2ER differed significantly based on the higher detection of mostly bacterial and viral and certain fungal and parasitic signatures in the samples of Cluster 2ER (**Table 2**). The ungrouped BRER samples (ungrouped 1ER) were significantly different from clusters 1ER and 2ER (**Table 2**).

Majority of the BRTP samples had similar microbial detections and grouped together into 1 major cluster (cluster 1TP), while few samples remained ungrouped (**Figure 4B**).

The BRHR samples formed 2 major clusters (cluster 1HR and cluster 2HR) (**Figure 4C**), and they differed from each other in having higher detection of certain bacterial and viral signatures in cluster 2HR compared to samples in cluster 1HR (**Table 3**). Bacterial signatures of *Kingella, Brevundimonas, Eikenella, Bartonella, Acinetobacter, Nodaviridae, Actinomyces, Aeromonas, Mobiluncus, Fusobacterium, Alcaligenes, Brucella,* and *Staphylococcus*; viral signatures of Orthomyxoviridae, Parvoviridae, Papillomaviridae, Nodaviridae, and Astroviridae and fungal signatures of *Aspergillus* showed significant higher detection in cluster 2HR. The 3 BRHRs that could not be grouped (ungrouped 1HR and 2HR) showed higher detection of certain microbial signatures listed in **Table 3** compared to the clustered BRHR samples; in particular, included the parasitic signature of *Entamoeba* and bacterial signatures of *Listeria* and *Corynebacterium*.

The BRTN samples formed two distinct clusters (cluster 1TN and 2TN) with 2 samples that did not cluster into distinct group (ungrouped 1TN) (**Figure 4D**). Cluster 1TN differed from Cluster 2TN in having higher detection of bacterial probes of *Caulobacter, Brevundimonas, Peptoniphilus, Rothia, Geobacillus, Aerococcus, Mobiluncus, Actinomyces, Bartonella*, fungal probes of *Malassezia, Piedraia, Rhodotorula, Rhizomucor* and parasitic signatures of *Leishmania, Toxocara, Contracaecum, Centrocestus, Trichuris, Strongyloides* (**Table 4**). Whereas, samples in Cluster 2TN had significant higher hybridization signal intensity for viral signatures of Poxviridae, Paramyxoviridae, Reoviridae, Parvoviridae, Arenaviridae, bacterial signatures of *Sphingomonas, Brucella, Orientia, Stenotrophomonas*, fungal signatures of *Pleistophora* and parasitic signatures of *Trichinella*. The ungrouped samples differed from the grouped samples in having significantly higher detection of certain viral probes of Anelloviridae, Retroviridae, Poxviridae, and Arenaviridae compared to Cluster 1TN and Cluster 2TN samples (**Table 4**).

Figure 4E shows the comparison of the microbial signatures from all four breast cancer types together in the clustering analysis. The data show that the different breast cancers grouped into 4 major clusters plus a few ungrouped BRER (2 samples), BRHR (3 samples), and BRTN (2 samples) samples (ungrouped 1, 2, and 3 respectively). Most of the BRTNs were very distinct in their microbial signature pattern association, and they clustered together (cluster 3). Similarly all the BRTPs screened clustered together to form a distinct cluster 4. Conversely, most of the BRER samples shared a similar microbial signature pattern with all of the BRHR samples forming the distinct cluster 1, while the remaining 11 BRER samples formed cluster 2. The BRERs in cluster 2 differed from those in Cluster 1 in having significant

TABLE 2 | Significant differences in microbial signatures between the hierarchical clusters of the endocrine receptor positive breast cancers (BRER).

Organism	T statistic	*P*-value	LogFC
BRER CLUSTER 1ER VS. CLUSTER 2ER IN FIGURE 4A			
Propionibacterium	−7.211	7.39E−07	−1.557
Arcanobacterium	−7.194	2.25E−06	−1.616
Brucella	−6.512	2.76E−05	−1.613
Adenoviridae	−5.232	9.03E−05	−0.901
Brevundimonas	−5.243	0.000166	−3.275
Caulobacter	−5.296	0.0002	−1.858
Pleistophora	−4.763	0.000346	−1.617
Eikenella	−4.538	0.000435	−2.023
Poxviridae	−4.644	0.000443	−0.823
Filoviridae	−3.886	0.00056	−0.906
Alcaligenes	−4.288	0.000854	−1.525
Kingella	−4.275	0.000879	−1.954
Reoviridae	−4.019	0.001078	−0.733
Flaviviridae	−4.083	0.001118	−1.037
Retroviridae	−3.926	0.001159	−1.087
Erysipelothrix	−3.928	0.00116	−1.730
Herpesviridae	−4.253	0.001288	−1.083
Orthomyxoviridae	−4.033	0.001308	−1.030
Papillomaviridae	−4.112	0.001326	−1.076
Shigella	−3.685	0.001537	−1.190
Astroviridae	−3.349	0.001945	−0.795
Citrobacter	−3.552	0.002306	−1.063
Macracanthorhynchus	−3.680	0.002931	−1.784
Rickettsia	−3.409	0.004689	−1.169
Coxiella	−3.490	0.004739	−2.031
Trichophyton	−3.566	0.00481	−2.754
Acinetobacter	−3.281	0.004871	−1.544
Escherichia	−3.366	0.005736	−1.493
Mycoplasma	−3.250	0.006671	−1.627
Staphylococcus	−3.098	0.007102	−0.751
Paramyxoviridae	−3.059	0.007333	−0.621
Aeromonas	−3.024	0.007923	−0.984
Pediococcus	−3.129	0.008223	−1.645
Mucor	−3.220	0.008331	−2.103
Brugia	−3.065	0.009069	−1.711
Paragonimus	−2.766	0.014456	−1.021
Polyomaviridae	−2.750	0.015486	−1.258
Rhabdoviridae	−2.825	0.016266	−0.913
Stenotrophomonas	−2.808	0.016458	−1.617
Parvoviridae	−2.594	0.017333	−0.465
Rhodotorula	−2.723	0.018918	−0.915
Cardiobacterium	−2.726	0.020259	−1.386
Prevotella	−2.521	0.020957	−0.911
Treponema	−2.660	0.022602	−1.820
Circoviridae	−2.587	0.02349	−1.135
Bunyaviridae	−2.544	0.025934	−0.772
Hymenolepis	−2.547	0.026229	−1.566
Hartmannella	−2.452	0.027852	−0.837
Plasmodium	−2.334	0.029775	−0.646

(Continued)

TABLE 2 | Continued

Organism	T statistic	*P*-value	LogFC
Sphingomonas	−2.440	0.033178	−1.495
Corynebacterium	−2.314	0.036828	−1.463
Lactobacillus	−2.284	0.040908	−0.887
Anelloviridae	−2.246	0.041582	−0.867
Aspergillus	−2.232	0.046187	−1.222
BRER CLUSTER 1ER VS. UNGROUPED 1ER IN FIGURE 4A			
Pleistophora	−28.365	4.78E−13	−4.105
Brugia	−7.866	1.87E−08	−1.653
Escherichia	−30.534	6.54E−08	−4.933
Neisseria	−6.310	4.77E−07	−1.466
Retroviridae	−13.206	0.000123	−2.407
Coxiella	−11.147	0.000156	−2.297
Togaviridae	−4.362	0.000578	−0.742
Papillomaviridae	−9.237	0.011952	−1.470
Citrobacter	−12.427	0.011967	−3.997
Brucella	−12.872	0.013747	−2.140
Corynebacterium	−7.594	0.015709	−3.346
Fusobacterium	−8.502	0.040745	−5.272
Rickettsia	−2.767	0.045012	−0.457
BRER CLUSTER 2ER VS. UNGROUPED 1ER IN FIGURE 4A			
Retroviridae	−4.822	0.00084	−1.320
Citrobacter	−7.684	0.003969	−2.934
Corynebacterium	−2.730	0.027196	−1.883
Escherichia	−7.864	8.52E−06	−3.439
Fusobacterium	−6.841	0.017391	−4.806
Neisseria	−2.315	0.042091	−1.067
Pleistophora	−7.693	1.07E−05	−2.489

higher hybridization signals for certain bacterial signatures like *Brevundimonas, Sphingomonas, Erysipelothrix, Mycoplasma, Brucella, Prevotella, Arcanobacterium, Staphylococcus, Rickettsia, Propionibacterium, Lactobacillus, Shigella*, viral signatures of Polyomaviridae, Circoviridae, Herpesviridae, Papillomaviridae, Retroviridae, Orthomyxoviridae, Flaviviridae, Iridoviridae, Poxviridae, Reoviridae, fungal signatures of *Trichophyton, Mucor, Rhodotorula, Geotrichum, Pleistophora* and parasitic signatures of *Paragonimus, Macracanthorhynchus, Hartmannella* (**Table 5**).

Thus, we identified specific microbial signature patterns associated with different breast cancer types. It will be interesting to see if such distinct microbial signature pattern associated with different breast cancer types, correlate to differences in pathogenesis and clinical outcome.

Association of Microbial Signatures With Clinical Outcomes in the Four Breast Cancer Types

The samples we used in this study were de-identified samples. Thus due to HIPPA regulations we were able to procure only limited sub-set of data from the Tumor Registry. This included outcome, specifically whether the patient was alive or dead since

TABLE 3 | Significant differences in microbial signatures between the hierarchical clusters of the human epidermal growth factor receptor 2 positive breast cancers (BRHR).

Organism	T statistic	*P*-value	LogFC
BRHR CLUSTER 1HR VS. CLUSTER 2HR IN FIGURE 4C			
Kingella	−8.239	8.7E−09	−2.185
Brevundimonas	−9.967	2.3E−08	−2.411
Eikenella	−6.587	4.2E−07	−2.008
Bartonella	−7.715	5.8E−07	−1.759
Acinetobacter	−5.766	9.3E−06	−2.858
Nodaviridae	−5.598	1.4E−05	−2.066
Actinomyces	−5.389	1.8E−05	−2.445
Aeromonas	−6.501	1.9E−05	−2.143
Mobiluncus	−4.808	9.1E−05	−2.027
Fusobacterium	−4.200	0.00059	−1.741
Alcaligenes	−4.352	0.00065	−1.164
Brucella	−3.442	0.00319	−0.891
Orthomyxoviridae	3.186	0.00422	1.180
Parvoviridae	−2.993	0.00666	−0.516
Aspergillus	−2.517	0.02244	−1.169
Papillomaviridae	−2.427	0.02385	−0.449
Staphylococcus	−2.073	0.04818	−1.147
Astroviridae	−2.128	0.04831	−0.642
BRHR CLUSTER 1HR VS. UNGROUPED(1HR+2HR), IN FIGURE 4C			
Lactobacillus	−8.722	5E-07	−3.149
Paramyxoviridae	−8.701	7E-07	−1.295
Entamoeba	−14.273	4E-05	−6.228
Anaplasma	−9.105	2E-04	−1.783
Astroviridae	−7.437	0.003	−1.825
Staphylococcus	−3.735	0.004	−1.758
Brucella	−6.840	0.011	−2.152
Mobiluncus	−5.976	0.011	−3.399
Rhabdoviridae	−6.586	0.013	−2.423
Klebsiella	−3.389	0.017	−2.348
Actinomyces	−3.952	0.02	−2.016
Candida	−4.990	0.025	−2.924
Flaviviridae	−5.242	0.026	−2.677
Caliciviridae	−3.043	0.027	−1.220
Ascaris	−3.655	0.03	−2.466
Arenaviridae	−4.007	0.035	−2.163
Corynebacterium	−4.706	0.038	−4.514
Lactococcus	−2.652	0.04	−0.769
Bartonella	−4.347	0.042	−1.957
Listeria	−4.206	0.043	−3.924
Thelazia	−3.931	0.045	−3.126
Papillomaviridae	−3.901	0.048	−1.585
Anelloviridae	−2.520	0.048	−1.859
Poxviridae	−4.155	0.049	−2.570
BRHR CLUSTER 2HR VS. UNGROUPED(1HR+2HR) IN FIGURE 4C			
Entamoeba	−15.712	0.0001	−6.501
Paramyxoviridae	−4.736	0.0003	−1.116
Lactobacillus	−4.561	0.0005	−2.908
Anaplasma	−4.400	0.0007	−1.343

(Continued)

TABLE 3 | Continued

Organism	T statistic	*P*-value	LogFC
Nodaviridae	4.850	0.0032	2.267
Astroviridae	−3.418	0.0068	−1.183
Rhabdoviridae	−4.586	0.008	−1.980
Candida	−5.384	0.0193	−3.190
Klebsiella	−4.734	0.0197	−2.750
Brucella	−3.320	0.0209	−1.261
Flaviviridae	−4.883	0.021	−2.601
Listeria	−4.088	0.026	−4.122
Arenaviridae	−3.471	0.0344	−1.991
Ascaris	−3.922	0.0368	−2.499
Anelloviridae	−3.885	0.0375	−2.349
Corynebacterium	−4.125	0.0408	−4.068

diagnosis and treatment; the cause of death and length of survival were not available. These data provide only indications of trends which will have to be statistically verified in future studies using samples with associated clinical data.

For these analyses the hierarchical clustering for each of the four different breast cancer types were further grouped into sub-clusters based on microbial detections (**Figure 5A**). In the BRTNs the cases of sub-cluster 2b had the highest (63%) proportion of the patients who had died, followed by that of Cluster 1 (33%); while sub-clusters 2a and 2c had a higher number of surviving patients (**Figure 5B**). The shared feature of sub-clusters 1 and 2b is a higher detection of fungal and parasitic signatures (**Figure 5A**). BRTP samples did not fall into discrete sub-clusters (**Figure 5A**), but overall BRTP showed 82% surviving patients. For BRER samples, sub-clusters 1a, b and c had similar numbers of patients who had died (25, 22, and 33%, respectively), while these numbers were much lower for sub-clusters 2a and 2b. Sub-clusters 2a and 2b are notable in that they have an overall more robust and diverse microbial signature. Examining the sub-clusters for BRHR shows a high number of surviving patients in all sub-clusters (1a, 1b, and 2; 75, 86, and 85%, respectively). Within the limits of the data these analyses suggest that the specific microbial signatures may correlate with outcome especially in the case of BRTN.

Using the survival data, we also examined variation in average hybridization signal for microbial signatures between the breast cancer types (**Figure 6** and **Table 6**). Interestingly, these analyses showed that high hybridization signals of specific viruses and microbes in a particular breast cancer type may trend with patients who had died, others trended with surviving patients. For example, in BRTP Herpesviridae signatures were detected significantly higher in BRTP patients who had died. Similarly, BRTN patients who had died had significant higher hybridization signals for certain fungal (*Malassezia*, *Rhizomucor*, *Rhodotorula*) and parasitic (*Centrocestus*, *Strongyloides*, *Trichuris*, *Contracaecum*, *Leishmania*) signatures. In the BRERs we found a trend of higher detection of the bacterium *Peptinophilus* signatures in the deceased cases. Similarly, we

TABLE 4 | Significant differences in microbial signatures between the hierarchical clusters of the triple negative breast cancers (BRTN).

Organism	T statistic	*P*-value	LogFC
BRTN CLUSTER 1TN VS CLUSTER 2TN IN FIGURE 4D			
Caulobacter	14.725	5E-16	4.006
Brevundimonas	16.431	5E-15	3.496
Peptoniphilus	12.519	9E-11	3.326
Rothia	10.847	4E-10	2.647
Geobacillus	11.926	3E-09	2.917
Aerococcus	12.298	3E-09	2.544
Mobiluncus	8.106	1E-08	2.793
Leishmania	6.619	2E-07	1.256
Actinomyces	6.417	2E-07	1.837
Malassezia	6.015	1E-06	1.598
Toxocara	5.734	4E-06	1.387
Contracaecum	5.814	4E-06	1.176
Piedraia	5.368	6E-06	0.815
Rhodotorula	5.298	6E-06	1.402
Centrocestus	5.325	2E-05	1.720
Rhizomucor	5.023	3E-05	1.413
Trichuris	4.913	3E-05	1.267
Strongyloides	4.814	4E-05	1.110
Bartonella	4.524	2E-04	1.901
Poxviridae	−3.959	3E-04	−0.504
Paramyxoviridae	−3.773	6E-04	−1.022
Sphingomonas	−3.686	9E-04	−1.264
Pleistophora	3.640	0.001	1.097
Reoviridae	−2.987	0.005	−0.539
Trichinella	2.916	0.006	1.062
Arenaviridae	−2.845	0.008	−0.713
Brucella	−2.748	0.01	−1.098
Orientia	−2.521	0.018	−0.942
Parvoviridae	−2.294	0.028	−0.701
Stenotrophomonas	−2.231	0.032	−0.533
BRTN CLUSTER 1TN VS CLUSTER 1TN IN FIGURE 4D			
Anelloviridae	−18.960	4E-05	−8.958
Retroviridae	−20.048	6E-11	−6.108
Poxviridae	−25.133	1E-13	−1.989
Arenaviridae	−10.652	4E-08	−1.201
Iridoviridae	3.061	0.008	0.972
Mycoplasma	2.912	0.011	1.102
Trichinella	4.806	3E-04	1.146
Rickettsia	2.916	0.011	1.156
Adenoviridae	2.848	0.013	1.178
Filoviridae	3.956	0.001	1.226
Actinomyces	6.995	4E-05	1.605
Babesia	7.967	1E-06	1.648
Aerococcus	6.799	0.014	2.342
Toxocara	28.244	1E-13	2.495
Rothia	13.228	3E-09	2.874
Centrocestus	13.652	0.036	2.912
Peptoniphilus	14.486	8E-10	3.518

(Continued)

TABLE 4 | Continued

Organism	T statistic	*P*-value	LogFC
BRTN CLUSTER 2TN VS. UNGROUPED 1TN IN FIGURE 4D			
Anelloviridae	−23.294	0.0007	−9.497
Retroviridae	−18.299	2E-12	−5.681
Poxviridae	−14.532	5E-13	−1.485
Arenaviridae	−2.111	0.0459	−0.488
Rickettsia	3.576	0.0017	1.064
Filoviridae	3.714	0.0012	1.100
Toxocara	4.920	6E-05	1.108
Centrocestus	3.124	0.0122	1.192
Adenoviridae	2.834	0.0097	1.249
Iridoviridae	4.805	8E-05	1.341
Babesia	5.584	1E-05	1.556
Mycoplasma	4.207	0.0004	1.701

found a trend of higher detection of certain bacteria (*Listeria*, *Lactobacillus*, *Borrelia*) in the BRHR cases with severe outcome.

Conversely, high hybridization signals for Paramyxoviridae, Astroviridae, and Polyomaviridae were found with greater frequency, respectively, in the BRTN, BRTP, and BRER cancer patients who survived. Additionally, high hybridization signals for the bacteria *Sphingomonas* and the fungus *Aspergillus* were detected in the BRHR patients who survived. Again within the limits of the clinical data these finding suggest that the qualitative and quantitative nature of the microbial signatures associated with a patient's cancer may provide diagnostic and prognostic information.

Validation of Pathochip Screen Results by PCR

We selected several viruses and microorganisms detected in the BC samples for verification by non-quantitative PCR and sequencing, these included several viral families and individual viruses (Herpesvirus, Polyoma, Papilloma, Parapox, and MMTV), as well as a prevalent bacterium (*Brevundimonas*), and fungus (*Pleistophora*). The primers used were either previously published (**Table 7**) or were designed based on sequences from the conserved and specific regions of the microorganisms. For detection of parasites we used pan-parasite diagnostic PCR primers enabling exhaustive detection of non-human eukaryotic species-specific small subunit rDNA in human clinical samples. For the validation experiments we used the WTA prepared and used for the initial screening. The PCR amplification showed the expected amplicons for the PathoChip-detected viruses, as well as the selected bacterium, fungus and parasite (**Figure 7**). Sequencing of the PCR products verified the detection of the appropriate virus or other microorganism (Supplementary Table S3, Supplementary Figure S3).

DISCUSSIONS

The human microbiome is comprised of mutualistic, pathogenic, transient and residential viruses and microorganisms. Many

TABLE 5 | Significant differences in microbial signatures between the hierarchical clusters of the endocrine receptor positive breast cancers of cluster 1 vs. 2 in **Figure 4E**.

Organism	T statistic	*P*-value	LogFC	Types
Trichophyton	−4.156	0.002	−3.067	Fungal
Mucor	−4.269	0.001	−2.400	Fungal
Brevundimonas	−2.357	0.033	−1.966	Bacterial
Sphingomonas	−3.025	0.012	−1.781	Bacterial
Erysipelothrix	−2.685	0.018	−1.725	Bacterial
Mycoplasma	−3.094	0.009	−1.598	Bacterial
Polyomaviridae	−4.103	7E-04	−1.511	Viral
Paragonimus	−3.974	0.001	−1.448	Parasitic
Macracanthorhynchus	−3.362	0.004	−1.437	Parasitic
Brucella	−4.795	3E-04	−1.419	Bacterial
Circoviridae	−2.985	0.011	−1.292	Viral
Prevotella	−2.306	0.04	−1.291	Bacterial
Hartmannella	−3.531	0.003	−1.183	Parasitic
Rhodotorula	−2.323	0.04	−1.116	Fungal
Herpesviridae	−2.669	0.022	−1.088	Viral
Geotrichum	−2.352	0.034	−1.015	Fungal
Arcanobacterium	−2.949	0.011	−1.013	Bacterial
Pleistophora	−2.163	0.047	−0.989	Fungal
Papillomaviridae	−2.913	0.013	−0.917	Viral
Staphylococcus	−3.037	0.009	−0.909	Bacterial
Retroviridae	−2.452	0.027	−0.876	Viral
Orthomyxoviridae	−3.343	0.005	−0.863	Viral
Rickettsia	−2.321	0.037	−0.852	Bacterial
Flaviviridae	−2.814	0.014	−0.851	Viral
Propionibacterium	−2.354	0.033	−0.834	Bacterial
Iridoviridae	−2.539	0.022	−0.830	Viral
Poxviridae	−2.732	0.019	−0.787	Viral
Lactobacillus	−2.271	0.04	−0.781	Bacterial
Shigella	−2.171	0.042	−0.735	Bacterial
Reoviridae	−2.579	0.023	−0.655	Viral

recent studies have suggested that the body's microbiome dramatically affects health, where perturbation of the microbiome leads to altered physiology and pathology, including cancer. However, the reverse may also be true, that different human diseases create disease microenvironments amenable to the persistence of a differential microbiome, with or without a direct effect on the establishment or progression of the disease. Such differential microbiomes could be specific to each such disease. Using our in-house metagenomic array technology (PathoChip), we previously established distinct microbial signatures in triple negative breast cancers (BRTNs) (Banerjee et al., 2015). In the present study we determined the microbial signatures that were significantly higher in the 4 major breast cancer types (BRTN, BRTP, BRER, BRHR) compared to the healthy breast control tissues, and also determined whether the microbial signatures associated with the BRTNs was a specific feature of BRTNs, or a generic feature shared with other types of breast cancers.

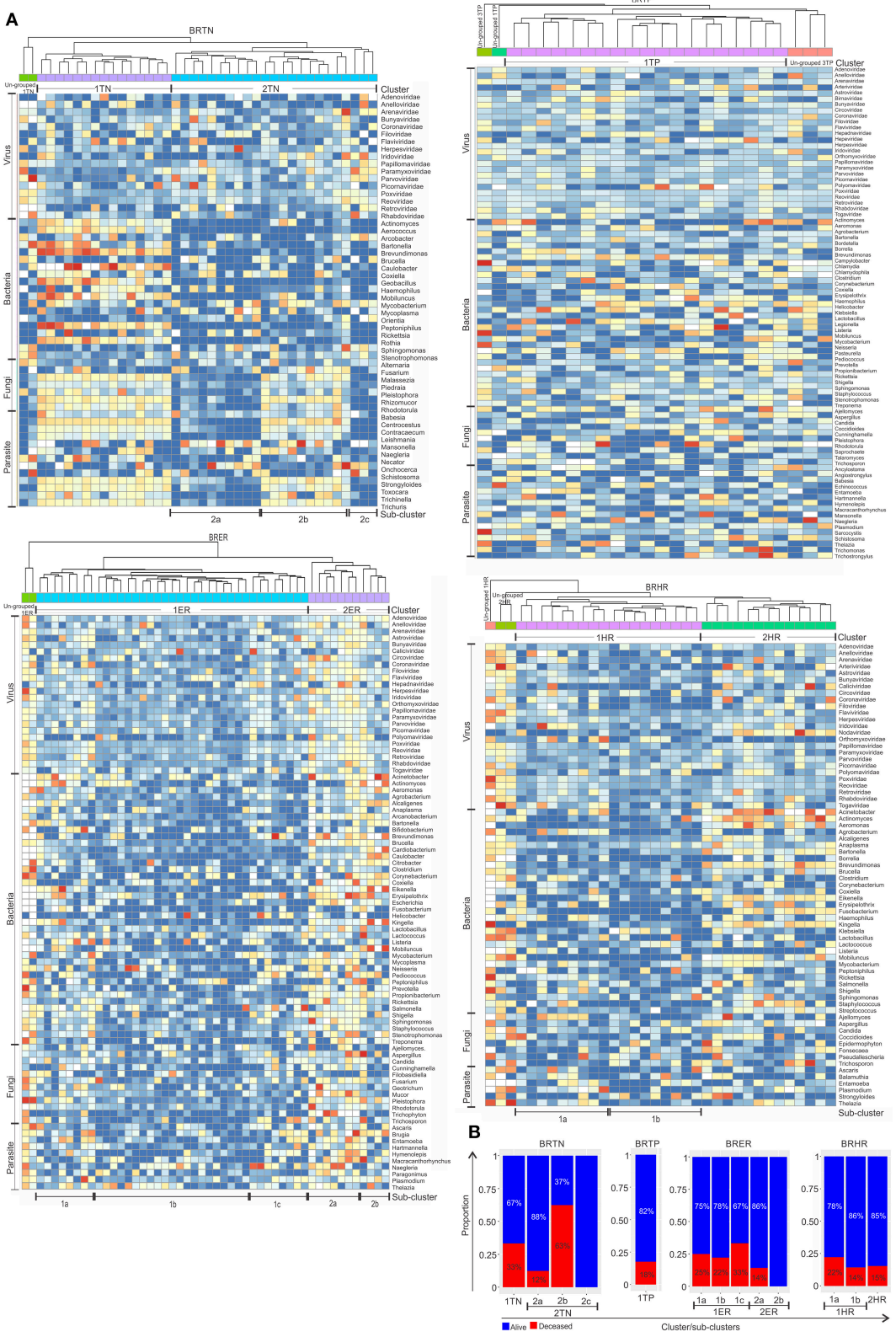


FIGURE 5 | Heat map of hierarchical clustering of the 4 types of breast cancers **(A)**, and the proportion of patients with and without severe outcome (death) in each cluster/sub-cluster **(B)**. Among the breast cancer samples, the endocrine receptor (estrogen/progesterone) positives are abbreviated as BRER, human epidermal growth factor receptor 2 positives are abbreviated as BRHR, triple positives (estrogen, progesterone, and HER2 receptor positive) are abbreviated as BRTP and the triple negatives (absence of estrogen, progesterone, and HER2 receptors) are abbreviated as BRTN.

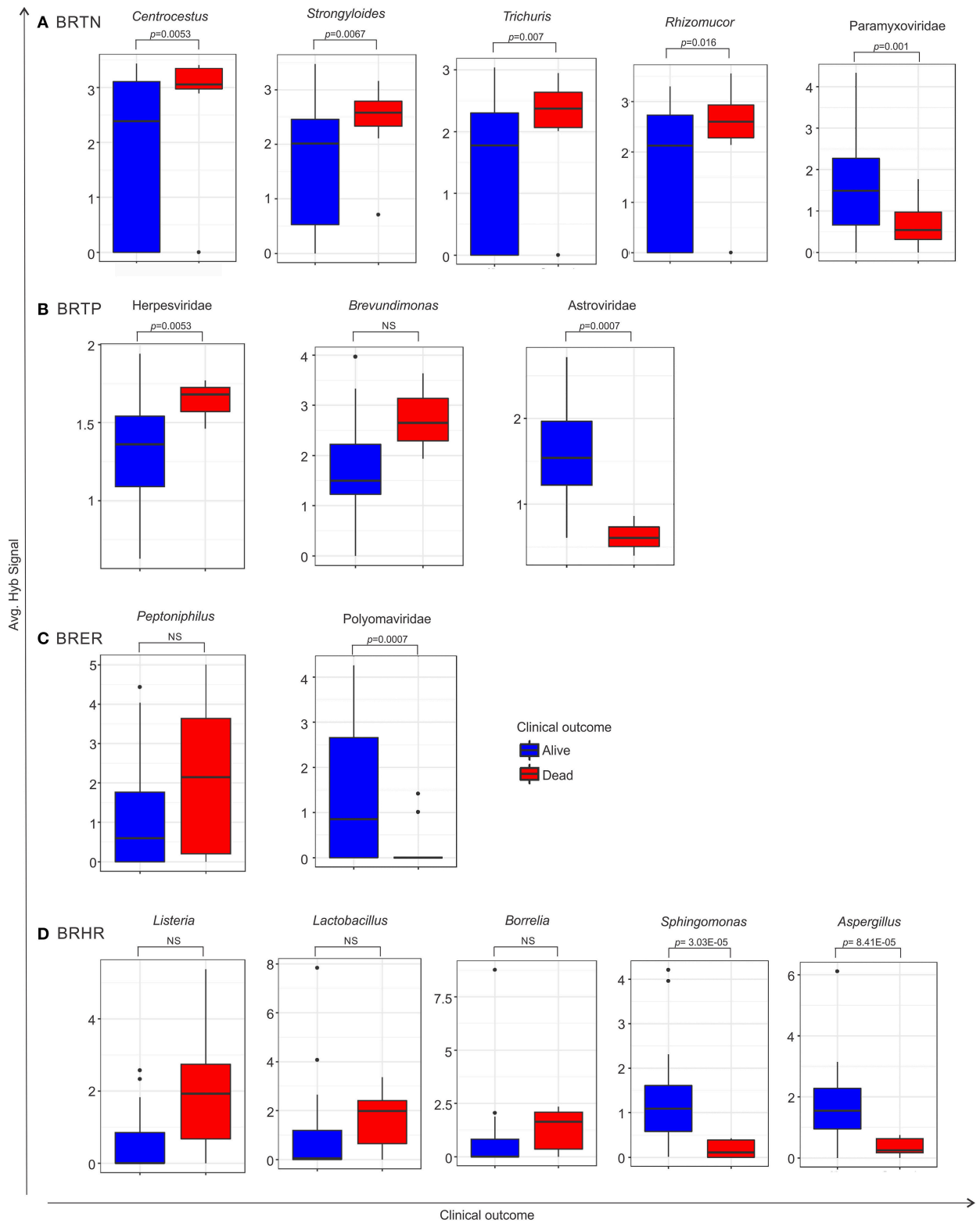


FIGURE 6 | Box plots representing either significant or trend of higher detection of microbial signatures in BRTN **(A)**, BRTP **(B)**, BRER **(C)**, and BRHR **(D)** cases with low (alive) or severe (dead) clinical outcomes, compared by one sided *t*-test. The *p*-value of the tests is shown in the figure if significant. NS represents non-significant test, however still a trend cannot be ignored. Among the breast cancer types, the endocrine receptor (estrogen/progesterone) positives are abbreviated as BRER, human epidermal growth factor receptor 2 positives are abbreviated as BRHR, triple positives (estrogen, progesterone, and HER2 receptor positive) are abbreviated as BRTP and the triple negatives (absence of estrogen, progesterone, and HER2 receptors) are abbreviated as BRTN.

TABLE 6 | One sided *t*-test of microbial signature detection in different breast cancer types [endocrine receptor positives (BRER), human epidermal growth factor receptor 2 positives (BRHR), triple positives (BRTP) and the triple negatives (BRTN)], with their clinical outcome.

Organism	*p*-value	Adjust *p*-value	logFC
BRTN_DECEASED_VS_ALIVE_T_TEST			
Centrocestus	0.0053	0.129	0.752
Strongyloides	0.00667	0.129	0.546
Trichuris	0.00708	0.129	0.745
Malassezia	0.0119	0.129	0.701
Contracaecum	0.01253	0.129	0.664
Leishmania	0.0138	0.129	0.752
Rhizomucor	0.01642	0.131	0.631
Rhodotorula	0.01881	0.132	0.79
BRTN_ALIVE_VS_DECEASED_T_TEST			
Paramyxoviridae	0.00131	0.0733	1.251
Filoviridae	0.00586	0.1642	1.578
Onchocerca	0.01925	0.282	1.466
Orientia	0.02014	0.282	1.809
Schistosoma	0.03627	0.4062	1.393
Arcobacter	0.04846	0.4191	1.147
BRTN_DECEASED_VS_ALIVE_T_TEST			
Herpesviridae	0.02802	0.9685	0.282
BRTN_ALIVE_VS_DECEASED_T_TEST			
Astroviridae	0.00072	0.0512	1.339
Anelloviridae	0.00179	0.0512	3.241
Campylobacter	0.00181	0.0512	2.478
Coccidioides	0.01974	0.4195	2.257
Hepeviridae	0.0313	0.5018	1.44
Angiostrongylus	0.0368	0.5018	2.033
Adenoviridae	0.04206	0.5018	0.327
BRTN_DECEASED_VS_ALIVE_T_TEST			
Peptoniphilus	0.10238	0.9997	0.879
BRTN_ALIVE_VS_DECEASED_T_TEST			
Eikenella	0.00027	0.0154	2.179
Kingella	0.00047	0.0154	2.111
Caulobacter	0.00071	0.0154	2.23
Polyomaviridae	0.00073	0.0154	2.26
Geotrichum	0.00089	0.0154	1.362
Alcaligenes	0.00146	0.0211	1.735
Ajellomyces	0.00206	0.0256	2.114
Escherichia	0.00239	0.026	1.465
Circoviridae	0.00405	0.0391	1.203
Cardiobacterium	0.00603	0.0485	1.659
Helicobacter	0.00715	0.0485	1.869
Anaplasma	0.00758	0.0485	1.697
Arcanobacterium	0.00762	0.0485	1.135
Herpesviridae	0.00781	0.0485	0.696
BRTN_DECEASED_VS_ALIVE_T_TEST			
Listeria	0.08097	1	2.008
Lactobacillus	0.17059	1	0.811
Borrelia	0.20093	1	0.762
BRTN_ALIVE_VS_DECEASED_T_TEST			
Sphingomonas	3.03*E*−05	0.0021	2.825
Aspergillus	8.41*E*−05	0.0029	2.239
Coxiella	0.00086	0.02	1.865
Candida	0.00188	0.0328	1.154
Epidermophyton	0.00306	0.0428	2.234

Nominal p-value (p-value) and p-value with multiple correction (adjust p-value) for each microbial signature detection along with the log2 fold change (logFC) for the t-tests are mentioned.

Our data showed that the various breast cancers have robust and varied micro-organisms with aspects that are unique to each type as well as shared components. The data suggest that breast cancer microbial signatures may provide type-specific communities of organisms unique to each breast cancer type. We also point out that our control FFPE samples, processed in the same way as tumor samples, had different signatures, generally with much lower hybridization signals, arguing against gross contamination.

Examining viral signatures we found that the majority of the viral families detected were associated with all 4 breast cancer types. However, several important viruses were differentially detected; for example, among known oncogenic viruses the signatures of Polyomaviridae were detected with high significance and high signal intensity in the BRER and BRHR samples and with low signal intensity in the other breast cancer types. Signatures of Hepadnaviridae were similarly detected in BRER and BRTPs with high signal intensity, but with very low signal intensity in the other two cancer types. It is intriguing that signatures for Parapoxviridae family were found in all the breast cancers with BRERs showing the highest level of detection. Parapox viruses are known to have homologs to human genes responsible for angiogenesis (Ueda et al., 2003; Delhon et al., 2004).

There were a number of bacterial families shared by all four breast cancer types. For example, all four breast cancer types had dominant signatures for Proteobacteria followed by Firmicutes. The presence of these two bacterial phyla in the breast cancer tissues has been reported (Urbaniak et al., 2014, 2016; Hieken et al., 2016), and suggested to be a result of adaptation to the fatty acid environment and metabolism in the tissue (Urbaniak et al., 2014). Another study found a positive correlation between Proteobacteria and the metabolic by-products of fatty acid metabolism, along with host-derived genes involved in fatty acid biosynthesis (El Aidy et al., 2013). In particular, the signature of the proteobacteria *Brevundimonas* genus was detected with high hybridization signal and prevalence in all four breast cancer types. *Brevundimonas* causes bacteremia and has been found associated with immunocompromised and/or cancer patients in other studies (Han and Andrade, 2005; Lee et al., 2011; Banerjee et al., 2015). Additionally, the *Mobiluncus* family was detected in all four types, it is mostly known to be associated with bacterial

TABLE 7 | Primers used for PCR validation of PathoChip screen.

Micro-organism	Primers	Sequence (5′-3′)	Annealing temp and time	Extension temp and time	Amplicon size (bp)	Detail
Herpes	FP 1 RP 2	GAA GAC GCT GAT GAA CCA CG AAG CAC CTG GTG TAC TTT CAC	51°C for 45 s	65°C for 20 s	96	Self-designed
Mouse mammary tumor virus (MMTV)	SN FP 5 (gag) RP 6 (gag)	ACT CAG AAG GAA ACC CCT GCC TC ATC TCC TTT TTC CCT GGC CTC TGC	57°C for 30 s	65°C for 30 s	70	Self-designed
HPV	HPV GP5 HPV gp6	TTTGTTACTGTGGTAGATACTA GAAAAATAAACTGTAAATCATATTC	43°C for30 s	65°C for 30 s	104–141	Self-designed
Polyoma	PYV.for PYV.rev	GGAAAGTCTTTAGGGTCTTCTACC TAGGTGCCAACCTATGGAACAGA	53°C for 30 s	65°C for 30 s	178–183	N Engl J Med 1992; 326:988–993 April 9, 1992
Parapox	FP RP	ATC TTC ACG GGC GCA GTC G CTC TTC GAC GAC GAC GGG AAC	56°C for 30 s	65°C for 30 s	286	Self-designed
Bacteria *Brevundimonas*	FP 17 RP 18	TTG CAG AGG ACA ATC CGA ACT GAG AAC TGC CTT TGA TAC TGG CGA TC	52°C for 30 s	65°C for 60 s	667	Self-designed
Fungus *Pleistophora*	FP 19 RP 20	AGG TCT CCT AGG TGA ATA GCC CCG TGC TTA CAG TTA TTT CCT C	48°C for 30 s	65°C for 30 s	219	Self-designed
Parasite	G3FI G3RI	GCCAGCAGCCGCGGTAATTC ACATTCTTGGCAAATGCTTTCGCAG	48°C for 30 s	65°C for 30 s	404	Patents WO 2014071946 A1

vaginosis (Gatti, 2000); however, the association of breast cancers may correlate with recent studies showing an association with breast abscesses and extragenital infections (Glupczynski et al., 1984; Sturm, 1989). We also detected Actinomyces signatures in all four breast cancers, especially in BRHRs where it was detected with very high signal intensity. Previous studies have reported Actinomycosis in the breast tissues of breast cancer patients (Aamir and Bokhari, 2005; Abdulrahman and Gateley, 2015; Banerjee et al., 2015), as primary (Salmasi et al., 2010), or secondary infections (Brunner et al., 2000) in breast, and in breast abscess (Attar et al., 2007). Additionally, each type of breast cancer held signatures for unique bacterial genera, and may provide an ability to detect specific breast cancer types.

Fungal infections in cancer patients are common. Among the fungal signatures we detected were yeasts like *Candida, Geotrichum, Rhodotorula, Trichosporon* as well as fungi causing Mucormycosis, Aspergillosis (cutaneous infections) and dermatophytes like *Epidermophyton* and *Trichophyton* are commonly known to be associated with cancers (Mays et al., 2006; Ansari et al., 2015; Banerjee et al., 2015; Jung et al., 2015; Rodríguez-Gutiérrez et al., 2015; Berkovits et al., 2016). Also, we detected *Fonsecaea* infection, which is seen to predispose squamous cell carcinoma development (Azevedo et al., 2015).

Possibly the most intriguing and unexpected result of the PathoChip screening is the detection of parasite signatures in different breast cancer types. These signatures were quite unique to the different breast cancer types with no signal parasite being prevalently found in all four. Many parasite signatures were distinctly detected in only one type of breast cancer. It should be kept in mind that our sensitive detection approach allows us to detect low abundance organisms, as well as unknown members of parasite families. However, the association of specific parasites with cancer is known. Among the parasites detected, *Trichinella* (detected in BRTN) has been found in a patient with recurrent ductal invasive breast carcinoma (Kristek et al., 2005). *Schistosoma* (detected in BRTN, BRTP) has been linked to bladder cancer (Samaras et al., 2010; Benamrouz et al., 2012); additionally we detected signatures of *Ascaris* (BRHR, BRER) and *Trichuris* (BRTN) which have been associated with pediatric cancers (Menon et al., 1999). Similarly, *Strongyloides* (BRTN, BRHR) has been associated with adult cancer patients (Guarner et al., 1997). Other signatures detected, *Leishmania* (BNTN) and *Plasmodium* (BRHR, BRTP, BRER), induce the inhibition of apoptosis (Heussler et al., 2001), which may promote oncogenesis (Lowe and Lin, 2000).

It was interesting to further investigate if detection of certain microbial signatures in breast cancers differed among patients who survived or died. We noticed higher detection of certain parasitic and fungal signatures in BRTN patients who died. Of particular interest in these analyses was the finding of high hybridization signals of specific viruses and microbes in a particular breast cancer type that may trend with patients who died, while others trended with surviving patients. Within the limits of the clinical data that could be provided, our

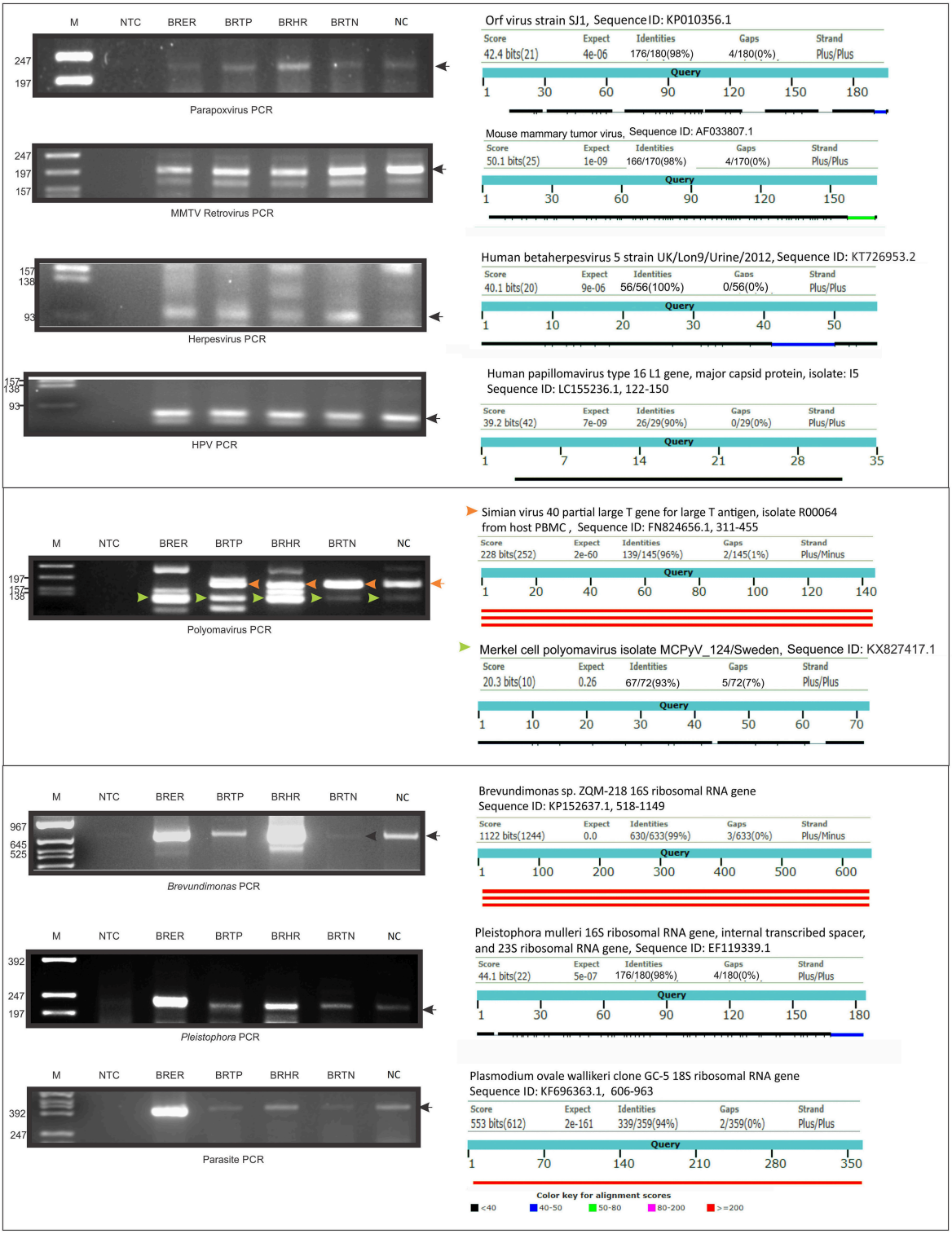


FIGURE 7 | PCR validation of microbial signatures in the 4 types of breast cancers and healthy control, using the primers from **Table 7**. Among the breast cancer types, the endocrine receptor (estrogen/progesterone) positives are abbreviated as BRER, human epidermal growth factor receptor 2 positives are abbreviated as BRHR, triple positives (estrogen, progesterone, and HER2 receptor positive) are abbreviated as BRTP and the triple negatives (absence of estrogen, progesterone, and HER2 receptors) are abbreviated as BRTN. The breast control samples obtained from healthy individuals are abbreviated as NC. The **left** shows the cropped gel pictures of EtBr stained amplicons run on agarose gel, where M is DNA ladder of RsaI digested ϕX/174, NTC is non-template control. The sequenced amplicons were subjected to nucleotide blast program in NCBI, and the results are shown in the **right**. In the Polyomavirus PCR gel picture, the orange and the green arrow heads signify Simian virus 40 and Merkel cell polyomavirus amplicons respectively, the electropherogram of the sequences of which are marked with the same arrow heads in Supplementary Figure S3.

findings suggest that the qualitative and quantitative nature of the microbial signatures associated with a patient's cancer may provide diagnostic and prognostic information.

Our findings suggest that the micro-organisms in breast cancers are diverse, extensive and have unique aspects that differentiate the four different breast cancers tested. We represented the microbial signatures that were significantly higher in the breast tumor microenvironment, when compared to healthy breast tissues. Some of these tumor microbial signatures overlapped with the reported skin microbiome (Findley and Grice, 2014; Hieken et al., 2016). For example: Bacteria like, *Lactobacillus, Prevotella, Staphylococcus, Lactococcus, Streptococcus* have been reported earlier as healthy breast skin flora (Hieken et al., 2016; Urbaniak et al., 2016), *Propionibacterium, Corynebacterium* bacteria, and *Malassezia* fungi has been reported to be common skin commensals (Grice and Segre, 2011). Although the detection of those common skin/healthy breast floras in the breast tumor microenvironment in the current study is not surprising, there still exists a breast tumor specific microbiome, which was also reported by other studies (Urbaniak et al., 2014; Xuan et al., 2014).

Many of the microbial signatures that were detected in one or more of the breast cancer types were not detected in the healthy controls, as mentioned in the results section. Most of those micro-organisms were found in earlier studies to be associated with cancer and/or immunocompromised patients (Kontoyianis et al., 1994; Menon et al., 1999; Narikiyo et al., 2004; Aamir and Bokhari, 2005; Kristek et al., 2005; Ramanan et al., 2014; Abdulrahman and Gateley, 2015; Banerjee et al., 2015, 2016).

It is possible that micro-organisms in the breast cancer could contribute to the origin, potentiation or modulation of oncogenesis. However, it is equally possible that the tumor microenvironment provides favorable conditions for specific micro-organisms to persist more readily than in the normal tissue microenvironment. Moreover, due to HIPAA regulations we could not get any information on the type of treatment these breast cancer patients received. Thus, while we can only assume that the samples from some of the patients could be obtained before treatment, others could be receiving treatment already at the time of sample procurement. Especially patients already receiving treatment could be immunocompromised, which further exposes them to a higher infection rate, and thus detecting higher number of micro-organisms from those samples is not surprising.

Our data demonstrate for the first time that the microbial signatures of BRTN and BRTPs are distinct and significantly different from the microbial signatures largely shared by BRER and BRHR. Furthermore, the unique characteristics of the breast cancer associated microbial signatures potentially provide certain tools for specific diagnosis and treatment of these cancers. These findings are hypothesis-generating and needs further investigation to identify a microbial risk signature for the different breast cancer types and potential microbial-based prevention therapies. A complete review of the microbiome in these breast cancers and healthy controls would open up more insight into answering those questions.

AUTHOR CONTRIBUTIONS

ER and JA conceptualized the study; SB and ER planned the experiments; SB performed the experiments, analyzed part of the data, made tables and figures for the manuscript, wrote the manuscript, with contributions from ER and JA; ZW and TT analyzed the micro-array data; KP provided technical assistance during the experiments; NS and MF were the pathologists who provided and evaluated the samples for identification of breast cancer and controls; AD identified the patients with different breast cancer types for inclusion in the study.

ACKNOWLEDGMENTS

The study was supported by Avon Foundation Grant no. Avon-02-2012-053 (to ER), and from the Abramson Cancer Center Director's fund. We thank Wallace Wormley and Noah Goodman for their support on the clinical correlation study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00951/full#supplementary-material

REFERENCES

Aamir, S., and Bokhari, N. (2005). Actinomycosis of breast. *Int. J. Pathol* 3:102.

Abdulrahman, G. O. Jr., and Gateley, C. A. (2015). Primary actinomycosis of the breast caused by *Actinomyces turicensis* with associated Peptoniphilus harei. *Breast Dis.* 35, 45–47. doi: 10.3233/BD-140381

American Cancer Society, W. C. O. (2017). *Cancer Facts & Figures 2017*. Atlanta, GA: American Cancer Society. Available online at: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf

Ansari, Sh., Shirzadi, E., and Elahi, M. (2015). The Prevalence of fungal infections in children with hematologic malignancy in Ali-Asghar Children Hospital between 2005 and 2010. *Iran J. Ped. Hematol. Oncol.* 5, 1–10.

Attar, K. H., Waghorn, D., Lyons, M., and Cunnick, G. (2007). Rare species of actinomyces as causative pathogens in breast abscess. *Breast J.* 13, 501–505. doi: 10.1111/j.1524-4741.2007.00472.x

Azevedo, C. M., Marques, S. G., Santos, D. W., Silva, R. R., Silva, N. F., Santos, D. A., et al. (2015). Squamous cell carcinoma derived from chronic chromoblastomycosis in Brazil. *Clin. Infect. Dis.* 60, 1500–1504. doi: 10.1093/cid/civ104

Baldwin, D. A., Feldman, M., Alwine, J. C., and Robertson, E. S. (2014). Metagenomic assay for identification of microbial pathogens in tumor tissues. *MBio* 5:e01714–14. doi: 10.1128/mBio.01714-14

Banerjee, S., Peck, K. N., Feldman, M. D., Schuster, M. G., Alwine, J. C., and Robertson, E. S. (2016). Identification of fungal pathogens in a patient with acute myelogenic leukemia using a pathogen detection array technology. *Cancer Biol. Ther.* 17, 339–345. doi: 10.1080/15384047.2015.1121349

Banerjee, S., Tian, T., Wei, Z., Peck, K. N., Shih, N., Chalian, A. A., et al. (2017a). Microbial signatures associated with oropharyngeal and oral squamous cell carcinomas. *Sci. Rep.* 7:4036. doi: 10.1038/s41598-017-03466-6

Banerjee, S., Tian, T., Wei, Z., Shih, N., Feldman, M. D., Alwine, J. C., et al. (2017b). The ovarian cancer oncobiome. *Oncotarget* 8, 36225–36245. doi: 10.18632/oncotarget.16717

Banerjee, S., Wei, Z., Tan, F., Peck, K. N., Shih, N., Feldman, M., et al. (2015). Distinct microbiological signatures associated with triple negative breast cancer. *Sci. Rep.* 5:15162. doi: 10.1038/srep15162

Benamrouz, S., Conseil, V., Creusy, C., Calderon, E., Dei-Cas, E., and Certad, G. (2012). Parasites and malignancies, a review, with emphasis on digestive cancer induced by Cryptosporidium parvum (Alveolata: Apicomplexa). *Parasite* 19, 101–115. doi: 10.1051/parasite/2012192101

Berkovits, C., Toth, A., Szenzenstein, J., Deak, T., Urban, E., Gacser, A., et al. (2016). Analysis of oral yeast microflora in patients with oral squamous cell carcinoma. *Springerplus* 5:1257. doi: 10.1186/s40064-016-2926-6

Brunner, S., Graf, S., Riegel, P., and Altwegg, M. (2000). Catalase-negative *Actinomyces neuii* subsp. neuii isolated from an infected mammary prosthesis. *Int. J. Med. Microbiol.* 290, 285–287. doi: 10.1016/S1438-4221(00)80128-9

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: an r package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36. doi: 10.18637/jss.v061.i06

Chen, Y., and Wei, J. (2015). Identification of pathogen signatures in prostate cancer using RNA-seq. *PLoS ONE* 10:e0128955. doi: 10.1371/journal.pone.0128955

Delhon, G., Tulman, E. R., Afonso, C. L., Lu, Z., De La Concha-Bermejillo, A., Lehmkuhl, H. D., et al. (2004). Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *J. Virol.* 78, 168–177. doi: 10.1128/JVI.78.1.168-177.2004

El Aidy, S., Derrien, M., Merrifield, C. A., Levenez, F., Dore, J., Boekschoten, M. V., et al. (2013). Gut bacteria-host metabolic interplay during conventionalisation of the mouse germfree colon. *ISME J.* 7, 743–755. doi: 10.1038/ismej.2012.142

Findley, K., and Grice, E. A. (2014). The skin microbiome: a focus on pathogens and their association with skin disease. *PLoS Pathog.* 10:e1004436. doi: 10.1371/journal.ppat.1004436

Gatti, M. (2000). Isolation of Mobiluncus species from the human vagina. *Zentralbl. Bakteriol.* 289, 869–878. doi: 10.1016/S0934-8840(00)80017-1

Glupczynski, Y., Labbe, M., Crokaert, F., Pepersack, F., Van Der Auwera, P., and Yourassowsky, E. (1984). Isolation of Mobiluncus in four cases of extragenital infections in adult women. *Eur. J. Clin. Microbiol.* 3, 433–435. doi: 10.1007/BF02017365

Grice, E. A., and Segre, J. A. (2011). The skin microbiome. *Nat. Rev. Microbiol.* 9, 244–253. doi: 10.1038/nrmicro2537

Guarner, J., Matilde-Nava, T., Villasenor-Flores, R., and Sanchez-Mejorada, G. (1997). Frequency of intestinal parasites in adult cancer patients in Mexico. *Arch. Med. Res.* 28, 219–222.

Han, X. Y., and Andrade, R. A. (2005). Brevundimonas diminuta infections and its resistance to fluoroquinolones. *J Antimicrob Chemother.* 55, 853–859. doi: 10.1093/jac/dki139

Heussler, V. T., Kuenzi, P., and Rottenberg, S. (2001). Inhibition of apoptosis by intracellular protozoan parasites. *Int. J. Parasitol.* 31, 1166–1176. doi: 10.1016/S0020-7519(01)00271-5

Hieken, T. J., Chen, J., Hoskin, T. L., Walther-Antonio, M., Johnson, S., Ramaker, S., et al. (2016). The microbiome of aseptically collected human breast tissue in benign and malignant disease. *Sci. Rep.* 6:30751. doi: 10.1038/srep30751

Jung, D. S., Farmakiotis, D., Jiang, Y., Tarrand, J. J., and Kontoyiannis, D. P. (2015). Uncommon candida species fungemia among cancer patients, Houston, Texas, USA. *Emerg. Infect. Dis.* 21, 1942–1950. doi: 10.3201/eid2111.150404

Kolde, R. (2015). *pheatmap: Pretty Heatmaps.* R package version 1.0.2.

Kontoyianis, D. P., Vartivarian, S., Anaissie, E. J., Samonis, G., Bodey, G. P., and Rinaldi, M. (1994). Infections due to Cunninghamella bertholletiae in patients with cancer: report of three cases and review. *Clin. Infect. Dis.* 18, 925–928. doi: 10.1093/clinids/18.6.925

Kristek, J., Marjanovic, K., Dmitrovic, B., Krajinovic, Z., and Sakic, K. (2005). Trichinella spiralis and breast carcinoma–a case report. *Coll. Antropol.* 29, 775–777.

Lee, M. R., Huang, Y. T., Liao, C. H., Chuang, T. Y., Lin, C. K., Lee, S. W., et al. (2011). Bacteremia caused by Brevundimonas species at a tertiary care hospital in Taiwan, 2000-2010. *Eur. J. Clin. Microbiol. Infect. Dis.* 30, 1185–1191. doi: 10.1007/s10096-011-1210-5

Lowe, S. W., and Lin, A. W. (2000). Apoptosis in cancer. *Carcinogenesis* 21, 485–495. doi: 10.1093/carcin/21.3.485

Madigan, M. P., Ziegler, R. G., Benichou, J., Byrne, C., and Hoover, R. N. (1995). Proportion of breast cancer cases in the United States explained by well-established risk factors. *J. Natl. Cancer Inst.* 87, 1681–1685. doi: 10.1093/jnci/87.22.1681

Mays, S. R., Bogle, M. A., and Bodey, G. P. (2006). Cutaneous fungal infections in the oncology patient: recognition and management. *Am. J. Clin. Dermatol.* 7, 31–43. doi: 10.2165/00128071-200607010-00004

Menon, B. S., Abdullah, M. S., Mahamud, F., and Singh, B. (1999). Intestinal parasites in Malaysian children with cancer. *J. Trop. Pediatr.* 45, 241–242. doi: 10.1093/tropej/45.4.241

Morales-Sanchez, A., and Fuentes-Panana, E. M. (2014). Human viruses and cancer. *Viruses* 6, 4047–4079. doi: 10.3390/v6104047

Narikiyo, M., Tanabe, C., Yamada, Y., Igaki, H., Tachimori, Y., Kato, H., et al. (2004). Frequent and preferential infection of *Treponema denticola*, *Streptococcus mitis*, and *Streptococcus anginosus* in esophageal cancers. *Cancer Sci.* 95, 569–574. doi: 10.1111/j.1349-7006.2004.tb02488.x

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramanan, P., Barreto, J. N., Osmon, D. R., and Tosh, P. K. (2014). Rothia bacteremia: a 10-year experience at Mayo Clinic, Rochester, Minnesota. *J. Clin. Microbiol.* 52, 3184–3189. doi: 10.1128/JCM.01270-14

Rodríguez-Gutiérrez, G., Carrillo-Casas, E. M., Arenas, R., Garcia-Mendez, J. O., Toussaint, S., Moreno-Morales, M. E., et al. (2015). Mucormycosis in a non-hodgkin lymphoma patient caused by Syncephalastrum racemosum: case report and review of literature. *Mycopathologia* 180, 89–93. doi: 10.1007/s11046-015-9878-1

Salmasi, A., Asgari, M., Khodadadi, N., and Rezaee, A. (2010). Primary actinomycosis of the breast presenting as a breast mass. *Breast Care (Basel).* 5, 105–107. doi: 10.1159/000301599

Samaras, V., Rafailidis, P. I., Mourtzoukou, E. G., Peppas, G., and Falagas, M. E. (2010). Chronic bacterial and parasitic infections and cancer: a review. *J. Infect. Dev. Ctries.* 4, 267–281. doi: 10.3855/jidc.819

Schnitt, S. J. (2010). Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod. Pathol.* 23(Suppl. 2), S60–S64. doi: 10.1038/modpathol.2010.33

Sheflin, A. M., Whitney, A. K., and Weir, T. L. (2014). Cancer-promoting effects of microbial dysbiosis. *Curr. Oncol. Rep.* 16:406. doi: 10.1007/s11912-014-0406-0

Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics, 2016. *CA Cancer J. Clin.* 66, 7–30. doi: 10.3322/caac.21332

Sturm, A. W. (1989). Mobiluncus species and other anaerobic bacteria in non-puerperal breast abscesses. *Eur. J. Clin. Microbiol. Infect. Dis.* 8, 789–792. doi: 10.1007/BF02185846

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031. doi: 10.1038/nature05414

Ueda, N., Wise, L. M., Stacker, S. A., Fleming, S. B., and Mercer, A. A. (2003). Pseudocowpox virus encodes a homolog of vascular endothelial growth factor. *Virology* 305, 298–309. doi: 10.1006/viro.2002.1750

Urbaniak, C., Cummins, J., Brackstone, M., Macklaim, J. M., Gloor, G. B., Baban, C. K., et al. (2014). Microbiota of human breast tissue. *Appl. Environ. Microbiol.* 80, 3007–3014. doi: 10.1128/AEM.00242-14

Urbaniak, C., Gloor, G. B., Brackstone, M., Scott, L., Tangney, M., and Reid, G. (2016). The microbiota of breast tissue and its association with breast cancer. *Appl. Environ. Microbiol.* 82, 5039–5048. doi: 10.1128/AEM.01235-16

Xuan, C., Shamonki, J. M., Chung, A., Dinome, M. L., Chung, M., Sieling, P. A., et al. (2014). Microbial dysbiosis is associated with human breast cancer. *PLoS ONE* 9:e83744. doi: 10.1371/journal.pone.0083744

Yersal, O., and Barutca, S. (2014). Biological subtypes of breast cancer: prognostic and therapeutic implications. *World J. Clin. Oncol.* 5, 412–424. doi: 10.5306/wjco.v5.i3.412

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Banerjee, Tian, Wei, Shih, Feldman, Peck, DeMichele, Alwine and Robertson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 June 2017
doi: 10.3389/fmicb.2017.01163

Oropharyngeal and Sputum Microbiomes Are Similar Following Exacerbation of Chronic Obstructive Pulmonary Disease

Hai-Yue Liu[1,2†], Shi-Yu Zhang[1†], Wan-Ying Yang[1], Xiao-Fang Su[3], Yan He[2], Hong-Wei Zhou[2] and Jin Su[1,3*]*

[1] Department of Environmental Health, School of Public Health, Southern Medical University, Guangzhou, China, [2] State Key Laboratory of Organ Failure Research, Division of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, China, [3] Chronic Airways Diseases Laboratory, Department of Respiratory and Critical Care Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, China

Growing evidence suggests that the airway microbiota might be involved in acute exacerbation of chronic obstructive pulmonary disease (AECOPD). Understanding this relationship requires examination of a large-scale population for a long duration to accurately monitor changes in the microbiome. This type of longitudinal study requires an appropriate sampling strategy; two options are the collection of sputum or oropharyngeal swabs. Comparative analysis of the changes that occur in these two specimen types has not been previously performed. This observational study was conducted to explore oropharyngeal microbial community dynamics over time and to examine the relationship between oropharyngeal swabs and sputum. A total of 114 samples were collected from four patients suffering from severe AECOPD. Bacterial and fungal communities were evaluated using 16S rRNA and ITS sequencing. Inter-individual differences were found in bacterial community structure, but the core genera were shared by both sample types and included 32 lineages. Most of the core genera were members of the phyla Proteobacteria, Firmicutes and Ascomycota. Although the oropharyngeal samples showed higher bacterial alpha diversity, the two sample types generated rather similar taxonomic profiles. These results suggest that the sputum microbiome is remarkably similar to the oropharyngeal microbiome. Thus, oropharyngeal swabs can potentially be used instead of sputum samples for patients with exacerbation of COPD.

Keywords: microbiome, 16S rRNA, ITS, AECOPD, sputum, oropharyngeal swab

OPEN ACCESS

Edited by:
Florence Abram, NUI Galway, Ireland

Reviewed by:
Nicola Segata, University of Trento, Italy
David William Waite, University of Queensland, Australia

****Correspondence:***
Jin Su
drsujin@126.com
Hong-Wei Zhou
biodegradation@gmail.com

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *12 April 2017*
Accepted: *07 June 2017*
Published: *22 June 2017*

Citation:
Liu H-Y, Zhang S-Y, Yang W-Y, Su X-F, He Y, Zhou H-W and Su J (2017) Oropharyngeal and Sputum Microbiomes Are Similar Following Exacerbation of Chronic Obstructive Pulmonary Disease. Front. Microbiol. 8:1163. doi: 10.3389/fmicb.2017.01163

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a common and slowly progressive disease characterized by sustained and irreversible airflow limitations that lead to the gradual loss of lung function (Einarsson et al., 2016). As symptoms progress, patients experience severe shortness of breath and extreme dyspnea. This makes the collection of sputum samples challenging (Kim et al., 2011), as sputum examination requires the use of a high-permeability 5% saltwater spray, and patients often find it difficult to tolerate atomization, resulting in failure to induce sputum

production for sampling. According to the World Health Organization, COPD will become the third most common cause of death globally by 2020 and will be the fifth most economically burdensome disease. The acute exacerbation of chronic obstructive pulmonary disease (AECOPD) is a major cause of this burden (Montes de Oca and Perez-Padilla, 2017).

Airway microbial communities associated with COPD have long been studied. AECOPD is mainly caused by infection with bacteria, viruses, and fungi, with more than 50% of cases being caused by bacterial infection (Sethi and Murphy, 2008). AECOPD is typically associated with the overgrowth of pathogenic bacteria, especially *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Streptococcus pneumoniae*, and *Moraxella catarrhalis*. In the airway, the presence of specific bacteria or fungi, such as *Moraxella* or *Aspergillus*, can increase the levels of inflammatory factors in the sputum and in turn exacerbate and increase the mortality associated with COPD (Hill et al., 2000; Huerta et al., 2014). Changes in airway microbial composition can also lead to exacerbation of COPD (Dy and Sethi, 2016). Therefore, studies of the changes that occur in the microbiome during AECOPD are crucial. A sputum-based longitudinal survey examining the function of the lung microbiome and its potential role in the disease etiology of AECOPD provides good examples to help understand the potential of the lung microbiome as a target for future respiratory therapeutics to manage COPD (Huang et al., 2014; Wang et al., 2016). Improved understanding of the microbial underpinnings of AECOPD will allow for the identification of therapeutic targets and the development of improved treatment options (Mammen and Sethi, 2016).

The anatomical structure of the airway is composed of a series of continuous channels: gas enters through the oral cavity or nose, passes through the pharynx and larynx, enters the trachea, and then gradually passes through the dendritic bronchi and bronchioles to the terminal bronchioles, respiratory bronchioles, and alveoli (Dmitrieva, 2013). Research has shown that the composition of the microbial community in the lower airways in healthy people is the same as that in the upper airways (Charlson et al., 2011). Additional research has further supported this opinion, showing similar microbial compositions in the upper and lower airways and also demonstrating that the lower airway possesses higher relative abundances of *Enterobacteriaceae* and *Haemophilus* (Morris et al., 2013). *Prevotella*, *Sphingomonas*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Veillonella*, *Staphylococcus,* and *Streptococcus* are commonly detected bacteria in healthy individuals, as well as in COPD airway microbiota (Zakharkina et al., 2013). Further, a previous study has reported that the airway microbiome of COPD patients harbors more *Pseudomonas* spp. of Proteobacteria and *Lactobacillus* spp. of Firmicutes than that of healthy individuals (Park et al., 2014). The microbial community present in COPD patients has been clearly described based on research using either oropharyngeal swabs or sputum samples. This research has shown that the oropharyngeal microbiota of COPD patients is mainly composed of Proteobacteria, Bacteroidetes, Firmicutes, and Actinobacteria, similar to the sputum microbiota (Park et al., 2014; Huang and Boushey, 2015). In pulmonary fibrosis, tuberculosis, and other diseases, oropharyngeal samples can be used in place of sputum samples to determine the structure of the airway microbial community (Botero et al., 2014; Zemanick et al., 2015).

Growing evidence suggests that it is important to examine large-scale populations for longitudinal studies attempting to identify therapeutic targets (Huang et al., 2014; Wang et al., 2016). This type of longitudinal study requires an appropriate sampling strategy, and two optional strategies include the collection of sputum and oropharyngeal swabs. In many patients with extreme dyspnea, collecting sputum samples is challenging; in contrast, obtaining oropharyngeal swabs is a non-invasive and easy process. Moreover, a comparative analysis of the changes that occur in sputum and oropharyngeal samples has not been previously performed. This observational study was conducted to explore oropharyngeal microbial community dynamics over time and to examine the relationship between oropharyngeal and sputum samples.

MATERIALS AND METHODS

Sample Collection

This study was carried out in accordance with the recommendations of the International Ethical Guidelines for Biomedical Research Involving Human Subjects and the ethics committee of Southern Medical University (Permit No. 2012-072), and all subjects provided written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of Southern Medical University. A total of 114 samples were collected at Nanfang Hospital, Southern Medical University (Guangzhou, China) between June 2012 and December 2012. The samples were collected from four patients with severe AECOPD during hospitalization stays of 14–17 days. All patients were male, and their age range was 68–83 years (**Table 1**). For sputum induction and processing, we used the guidelines suggested by the Task Force on Induced Sputum of the European Respiratory Society (Paggiaro et al., 2002; Vignola et al., 2002). The samples were immediately stored at −80°C for subsequent DNA extraction. Swabs were taken from the oropharyngeal wall.

Pulmonary Function (PF) Tests

Spirometry was performed using a Jaeger Masterscope spirometry system (Jaeger, Wuerzburg, Germany) according to

TABLE 1 | Clinical information for the four study subjects.

Subject	Age	Gender	FEV1%	Sputum culture results	Hospitalized days
A	83	Male	47.5	*Acinetobacter baumannii*	17
B	68	Male	23.3	None	15
C	80	Male	None	None	14
D	79	Male	31.2	NA	14

the American Thoracic Society (ATS) guidelines (Miller et al., 2005). The forced expiratory volume in 1 s as a percentage of predicted (FEV1%) is considered an important diagnostic measurement of COPD. Commonly, FEV1% is used to measure the grade of COPD as follows: mild: FEV1% $\geq$ 80; moderate: 50 $\leq$ FEV1% $<$ 80; severe: 30 $\leq$ FEV1% $<$ 50; and very severe: FEV1% $<$ 30. However, FEV1% alone may not adequately reflect a patient's overall health status (van der Molen and Cazzola, 2012).

Culture Method

Sputum culturing included homogenization with dithiothreitol and the plating of aliquots of serial dilutions on blood, chocolate, and MacConkey culture agars.

DNA Extraction and PCR

Genomic DNA was extracted from each sample using a Total Nucleic Acid Extraction Kit (Bioeasy Technology, Inc., China) according to the manufacturer's instructions. The 16S rRNA gene was amplified using barcoded V4 primers. The ITS gene was amplified using barcoded ITS1 primers and then purified and pooled as described in our earlier studies (He et al., 2013; Su et al., 2015). The 16S rRNA and ITS PCR products were sequenced at the Beijing Genomics Institute using paired-end sequencing on an Illumina HiSeq 2000 platform. For bacteria, PCR was performed with the bacterial-specific primers 514F-5′ GTGCCAGGMGCCGCGGTAA 3′ and 805R-5′ GGACTACHVGGGTWTCTAAT 3′. The reaction conditions were as follows: initialization at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s, and elongation at 72°C for 45 s, and a final elongation step at 72°C for 5 min. For fungi, PCR was performed with the fungal-specific primers ITSF 5′ CTTGGTCATTTAGAGGAAGTAA 3′ and ITSR 5′ GCTGCGTTCTTCATCGATGC 3′. The reaction conditions were as follows: initialization at 94°C for 15 min, followed by 5 cycles of denaturation at 95°C for 30 s, annealing at 50°C for 30 s, and elongation at 72°C for 60 s, and subsequently, 35 cycles of denaturation at 95°C for 30 s, annealing at 65°C for 30 s, and elongation at 72°C for 60 s, followed by a final elongation step at 72°C for 5 min.

Sequence Processing and Analysis

The sequences were demultiplexed and quality filtered using the QIIME (Quantitative Insights Into Microbial Ecology) platform (1.9.1) (Caporaso et al., 2010). Sequencing was conducted to generate 100-bp paired-end reads using an Illumina HiSeq 2000 sequencer according to the manufacturer's instructions. The Illumina sequencing quality report revealed that the sequence quality was relatively high for fragment sizes of up to 80 bp, with a sharp decrease in quality for larger fragments. Thus, we trimmed the raw sequences to 80 bp for each read pair. The sequences were then de-multiplexed, trimmed of barcodes and primers and filtered if they contained ambiguous bases or mismatches in the primer regions, according to the BIPES protocol (Zhou et al., 2011). The detailed protocol was as follows: first, we deleted the sequences with barcoded primers that contained ambiguous reads or mismatches in the primer region; then, we removed the primers and kept the remaining clean sequences of the 16S and ITS genes. Second, we removed any sequences with more than one mismatch within the 40–70 bp region at each end. Next, we used 30Ns to concatenate adjacent single-ended sequences for the subsequent sequence analyses, as our paired-end sequences did not extend to the V4 regions of the 16S rRNA gene (He et al., 2013). All the tools used in this study have been validated for use with gapped sequences. We screened for and removed chimeras using UCHIME in *de novo* mode (Edgar et al., 2011). The final high-quality sequence reads for the 16S and ITS genes were generated after the sequences were screened with UCHIME. The sequences were deposited in the European Nucleotide Archive (ENA), and the accession number was ERS1659093.

Subsequent analyses were implemented using the QIIME. The sequences were then clustered into operational taxonomic units (OTUs) using USEARCH with the default parameters and with the threshold distance set to 0.03 for 16S genes and 0 for the ITS genes. QIIME-derived reference-based alignments of representative sequences were performed using PyNAST with the Greengenes 13_8 database as the template file (Al-Hebshi et al., 2015). The Ribosome Database Project (RDP) algorithm was applied to classify representative 16S sequences into specific taxa using the default database (Wang et al., 2007). Representative ITS DNA gene sequences were classified using QIIME_ITS database as a reference (version information: sh_qiime_release_01.08.2015) (Caporaso et al., 2010). The biome data were filtered using the filter_otus_from_otu_table.py script with the parameter (-s 3) to remove low-abundance OTUs. Next, each 16S rRNA and ITS DNA gene sample was normalized to 1,000 sequences to avoid biases resulting from uneven sequencing depth among samples. All samples with less than 1,000 sequences and their paired (oropharyngeal swab/sputum) samples were not included in the later analysis.

To determine which sampling routine provided higher alpha diversity while controlling for patient and day effects, we used a simple up/down scoring system whereby on each subject-day, we recorded which sample had higher diversity. The overall results were then aggregated into a distribution and assessed using the Chi-square test. Adonis was used to estimate the dissimilarity of the microbial compositions between groups. Procrustes analysis was performed to determine whether the beta diversity results were similar between the oropharyngeal swabs and sputum samples. Weighted uniFrac distances based on the phylogenetic metric were used for these two analyses. Correlations of the core genera (representing $>$10% of the sequences in any sample) between the two sample types were assessed by Spearman correlation analysis. We identified differential features between the two groups using a linear discriminant analysis (LDA) effect size (LEfSe) method. The threshold cutoff values of the logarithmic LDA scores for identifying taxa that differed in abundance between the comparison groups were 3.5 for bacteria and 2.0 for fungi. To obtain OTU data for LEfSe analysis, we selected OTUs with a relative abundance of over 10% in any sample.

RESULTS

Associations between Sample Sites, Individuals and Microbial Communities

Alpha diversity measurements were used to assess variations in community structure among the microbiota collected in the oropharyngeal and sputum samples. Specifically, phylogenetic (PD_whole_tree), richness (Observed_OTUs), and evenness and richness metrics (Shannon) were employed to analyze community alpha diversity (**Figure 1**). We detected significant differences in the bacterial alpha diversity between the sputum and oropharyngeal samples for the PD_whole_tree and Shannon indices ($P < 0.01$). However, the comparisons did not reveal any significant differences in the fungal community compositions. These results indicated that the oropharyngeal samples had higher alpha diversity for bacterial communities.

Adonis analysis of the weighted_uniFrac distances performed using phylogenetic information revealed significant differences in the bacterial communities among individuals, indicating the presence of inter-individual differences in microbial composition (weighted_uniFrac distances, $P < 0.01$). In addition, the results revealed that the samples of the two specimen types had significantly different beta diversity results (weighted_uniFrac distances, $P < 0.01$), and greater inter-individual differences than inter-specimen differences were detected (**Figure 2**). In contrast, the results revealed non-significant differences in the fungal beta diversity results both between specimen types and among individuals. To visualize the microbial similarity and dissimilarity among the individuals and between the two sample types, PCoA was performed in the QIIME pipeline using the weighted_uniFrac distance. The results revealed broad overlap between the oropharyngeal and sputum samples, suggesting that both sample types shared similar bacterial/fungal communities. The samples from the four individuals were divided into four distinct clusters in terms of bacterial community composition.

Microbial Composition Dynamics in Sputum and Oropharyngeal Samples

Large inter-individual differences in bacterial community structure were observed under different dynamic conditions. In Subject A, Proteobacteria (68.59%) and Firmicutes (20.52) were the predominant phyla, with *Psychrobacter* (43.76%) and *Lactobacillus* (13.77%) as the most prevalent genera. With regard to the dynamics in Subject A, the abundance of *Haemophilus* was high on the first 2 days and decreased starting on the 3rd day. In contrast, *Psychrobacter* sharply increased starting on the 3rd day, subsequently decreased on the 14th day and then increased again starting on the 16th day. The abundance of *Enterobacteriaceae* peaked on days 14 and 15. In addition, the abundance of *Psychrobacter* decreased quickly after piperacillin and sulbactam were added (**Figure 3B**).

In Subject B, Firmicutes (34.40%), Proteobacteria (27.71%), Bacteroidetes (14.50%), and Actinobacteria (13.47%) were the main phyla, and *Psychrobacter* (12.25%), *Lactobacillus* (11.21%),

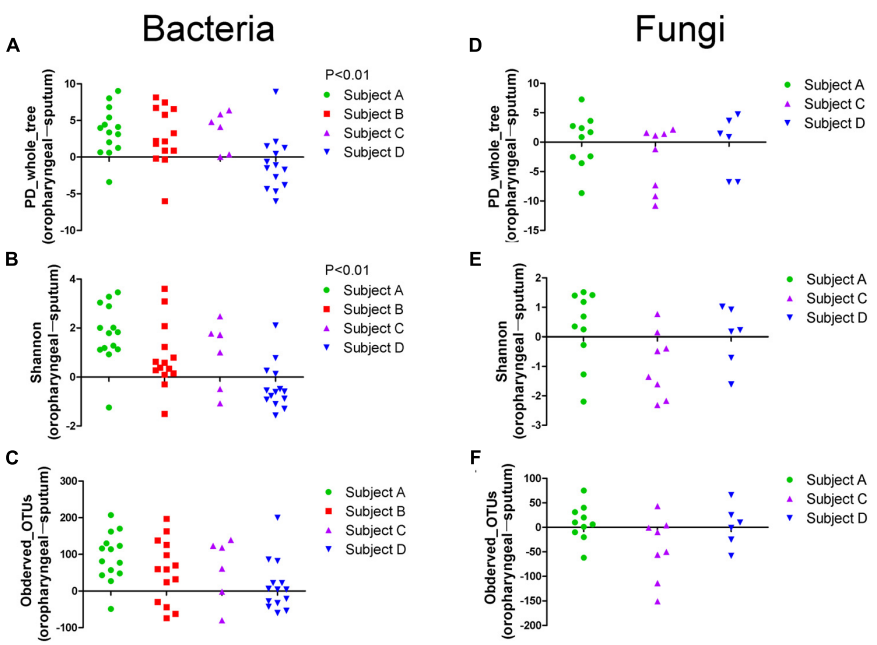


FIGURE 1 | Scatter plot showing the alpha diversity measurements of the microbiota collected from the oropharyngeal and sputum samples. **(A)** PD_whole_tree, **(B)** Shannon and **(C)** Observed_OTUs indices of bacterial community. **(D)** PD_whole_tree, **(E)** Shannon and **(F)** Observed_OTUs indices of fungal community. For each patient, swab samples are compared to sputum samples from different days. Each point represents a difference between the oropharyngeal swab and sputum sample, as determined using a simple up/down scoring system whereby on each subject-day, the sample with the higher diversity was recorded (obtained using the Chi-square test). PD_whole_tree, Observed_OTUs, and Shannon indices were used as phylogenetic, richness, and evenness and richness metrics, respectively, to determine the diversities of the bacterial and fungal communities.

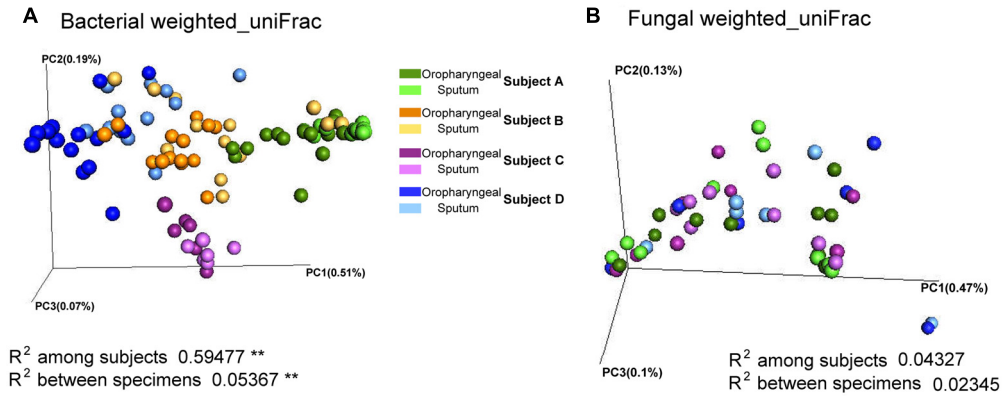


FIGURE 2 | Beta diversity of sputum and oropharyngeal samples. **(A)** PCoA based on weighted_uniFrac distances for bacterial sequences obtained from Subjects A (green), B (orange), C (purple), and D (blue). **(B)** PCoA based on weighted_uniFrac distances for fungal sequences obtained from Subjects A (green), C (purple), and D (blue). The different sample types are indicated by the different shades (Adonis test, $^{**}P < 0.01$).

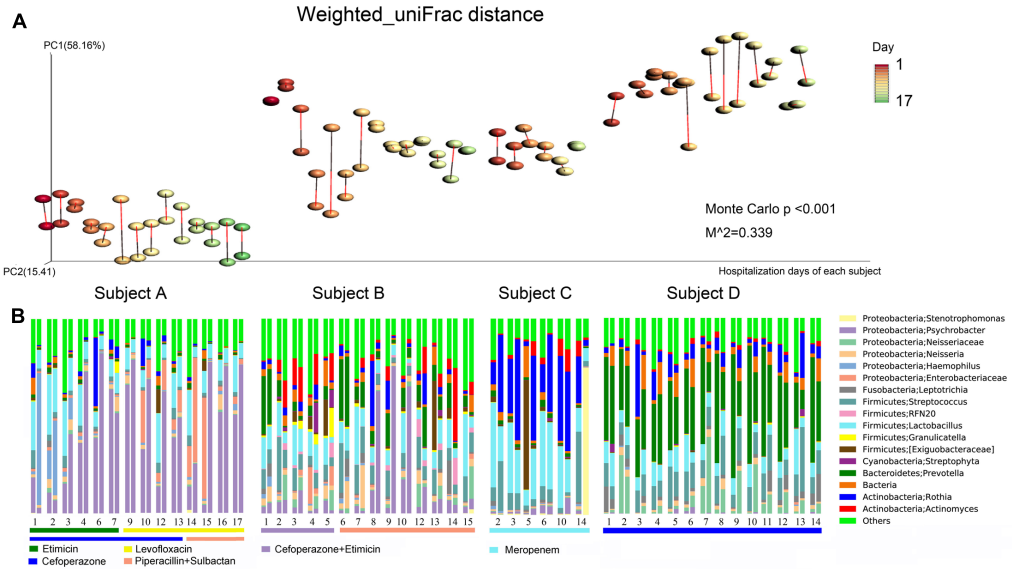


FIGURE 3 | Bacterial community composition dynamics of the daily samples. **(A)** Procrustes analysis based on weighted_uniFrac distances was performed, which revealed the same beta diversity results for the oropharyngeal swabs and sputum samples for each subject-day. The points linked with bars represent data obtained on the same day but from two different species, red represents the oropharyngeal samples, and black represents the sputum samples in the edge-color pairs. **(B)** Relative abundances of the core bacterial genera in the daily samples. For the four subjects, the single-day paired bars are separated, with the oropharyngeal samples on the left and the sputum samples on the right. The rectangular bars show which classes of antibiotic each patient received during their hospital stay.

and *Prevotella* (10.00%) were the most common genera. The abundance of *Prevotella* was high on the 1st day, after which it decreased quickly and subsequently increased again starting on the 6th day. *Actinomyces* increased starting on the 2nd day, decreased on the 6th day, and then increased again on day 12. The abundance of *Psychrobacter* peaked on days 9 and 10, and it was only found to have a high abundance in the sputum samples. The abundances of *Psychrobacter*, *Actinomyces*, *Enterobacteriaceae,* and *Streptophyta* were decreased following the addition of piperacillin and sulbactam (**Figure 3B**).

In Subject C, Firmicutes (42.46%), Actinobacteria (35.88%) and Proteobacteria (13.90%) were the most common phyla, and *Rothia* (28.67%) and *Lactobacillus* (24.63%) were the predominant genera. In addition, the abundance of *Streptococcus* decreased whereas that of *Actinomyces* increased starting on the 10th day. In Subject D, Firmicutes (25.53%) and Proteobacteria (14.84%) constituted the majority of sequences, and *Prevotella* (34.77%) and *Streptococcus* (12.18%) were the most common genera (**Figure 3B**). In addition, we observed that the sputum and oropharyngeal sample dynamics were similar.

Procrustes analysis based on the bacterial weighted_uniFrac distances of the time series data revealed that the beta diversity results were the same for both sample types (**Figure 3A**). These findings further demonstrated that the oropharyngeal swab and

sputum sample dynamics were similar, indicating that they had similar bacterial community structures.

In terms of fungal composition, the inter-individual and inter-specimen differences were small (**Figure 2B**). Meanwhile, Procrustes analysis based on the fungal weighted_uniFrac distances of the time series data did not revealed a same beta diversity for both sample types (**Figure 4A**). Ascomycota was the most predominant phylum among all the samples (72.59%), followed by Glomeromycota (19.89%) and Basidiomycota (9.66%). These three phyla represented over 99% of all the fungi sequences. At the genus level, *Eurotiales|unidentified* (27.91%), *Glomeraceae|unidentified* (18.33%), *Aspergillus* (10.25%), *Candida* (9.14%), and *Trichocomaceae|Other* (9.13%) were the most frequently identified fungi for the three individuals. In addition, the microbial compositions of the sputum and oropharyngeal samples collected on the same day and from the same subject were found to be very similar on most of the observation days (**Figure 4B**).

We further demonstrated that most sequences were shared between the two specimen types and that the unique taxonomic units and OTUs found in the sputum and oropharyngeal samples had relative abundances of less than 0.001. Among the 32 core genera (defined as the taxa representing more than 10% of the relative abundance in any sample), 16 were successfully assigned to a specific genus, including: *Psychrobacter, Stenotrophomonas, Haemophilus,* and *Neisseria* from Proteobacteria; *Streptococcus*, *Granulicatella,* and *Lactobacillus* from Firmicutes; *Actinomyces* and *Rothia* from Actinobacteria; *Leptotrichia* from Fusobacteria; *Prevotella* from Bacteroidetes; *Streptophyta* from Cyanobacteria; *Aspergillus*, *Acremonium,* and *Candida* from Ascomycota; and *Sterigmatomyces* from Basidiomycota. We further examined the correlations between the core genera in the paired samples. The results showed that 12 out of 17 of the core bacterial genera were significantly correlated between the two specimens (**Table 2**). Furthermore, we found no significant correlations between the sample types in terms of the core fungal genera.

Differences in Microbiota Composition between the Oropharyngeal and Sputum Samples

We performed LEfSe analysis for each individual and identified several similar discriminating features among the different individuals. For bacteria, the oropharyngeal samples were enriched with Bacteroidetes, i.e., *Prevotella* and *Dysgonomonas*; Firmicutes, i.e., *Lactobacillus*, *Coprococcus,*

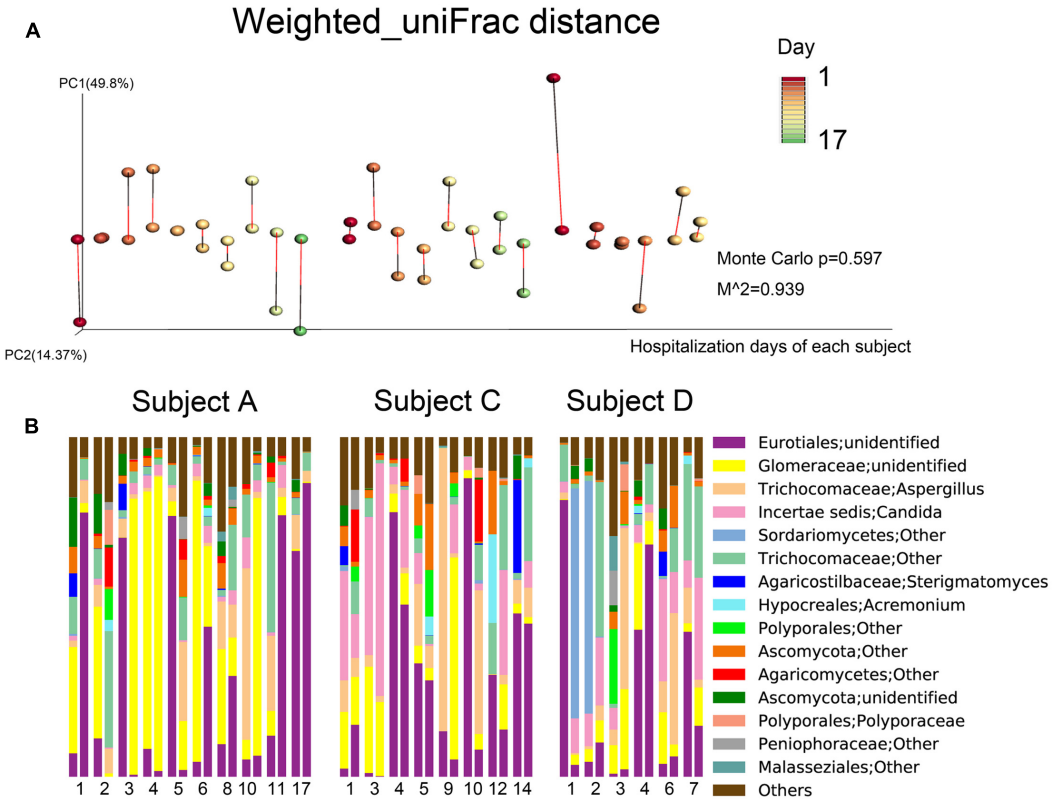


FIGURE 4 | Fungal community composition dynamics of the daily samples. **(A)** Procrustes analysis based on weighted_uniFrac distances was performed, which revealed the same beta diversity results for the oropharyngeal swabs and sputum samples for each subject-day. The points linked with bars represent data obtained on the same day but from two different species, red represents the oropharyngeal samples, and green represents the sputum samples in the edge-color pairs. **(B)** Relative abundances of the core fungal genera in the daily samples. For the three subjects, the single-day paired bars are separated, with the oropharyngeal samples on the left and the sputum samples on the right.

TABLE 2 | Correlations between the core genera in the paired samples (>10% sequences in any sample).

Genus	Spearman's rank correlation coefficient (rs)	Significance (*p*)
Proteobacteria| *Psychrobacter*	0.782	0.000
Proteobacteria| *Stenotrophomonas*	0.360	0.012
Proteobacteria| *Haemophilus*	0.536	0.000
Bacteroidetes| *Prevotella*	0.822	0.000
Proteobacteria| *Enterobacteriaceae;g_*	0.698	0.000
Firmicutes| *Exiguobacteraceae;g_*	0.321	0.026
Actinobacteria| *Rothia*	0.446	0.001
Proteobacteria| *Neisseriaceae;g_*	0.642	0.000
Actinobacteria| *Actinomyces*	0.663	0.000
Firmicutes| *Streptococcus*	0.662	0.000
Firmicutes| *Granulicatella*	0.363	0.011
Firmicutes| *RFN20*	0.816	0.000

and *Streptococcus*; and Fusobacteria, i.e., *Fusobacteriales*. In contrast, the sputum samples contained a high prevalence of Proteobacteria. In addition, several distinct features were only detected in a single individual. Genera showing increased abundances in the oropharyngeal samples included *Actinomyces*, *Brevibacillus*, *Peptostreptococcus*, *Anaerovibrio*, *Sutterella*, *Neisseria*, *Desulfovibrio*, *Atopobium*, *Fusobacterium,* and *Oscillospira*. In contrast, genera with elevated abundances in the sputum samples included *Haemophilus*, *Enterococcus*, *RFN20*, *Oleibacter*, *Catonella*, *Leptotrichia*, *Lautropia*, and *Akkermansia*. Shared distinct features were more frequently observed in the oropharyngeal swabs from the different individuals, whereas differences in the distinct features were more commonly detected in the sputum samples from the individuals (**Figure 5**). LEfSe analysis, performed using the OTUs for further examination, revealed that *Prevotella melaninogenica* and *Lactobacillus iners* were the most abundant species in the oropharyngeal swabs (**Figure 7**).

For fungi, a LEfSe comparison of the oropharyngeal and sputum samples showed that *Penicillium* existed at a high abundance in the oropharyngeal samples (**Figure 6**). The results of our study demonstrated that the microbial community structures of these two ecological niches were much more similar for fungi than for bacteria.

DISCUSSION

Acute exacerbation of chronic obstructive pulmonary disease is a chronic progressive disease characterized by shortness of breath, expectoration, and the gradual development of severe dyspnea. It is difficult for some patients with severe to very severe AECOPD to expel sputum, necessitating respiratory support (Kim et al., 2011). Sputum-based longitudinal airway microbiome studies have been performed for the deep exploration of therapeutic targets and the development of improved treatment options (Wang et al., 2016). The upper airway is considered the beginning of the microbiological community of the body, including bacteria and fungi (Delhaes et al., 2012). Recent studies have indicated that oropharynx communities vary in terms of relative abundances and that they resemble those in sputum samples, consistent with the results of this study (Botero et al., 2014; Zemanick et al., 2015). In this work, we first examined the daily microbiota dynamics present in oropharyngeal swabs and sputum samples collected from COPD patients. We discovered that both specimen types exhibited similar microbial compositions and that the dynamics of these compositions were largely consistent.

The initial comparison showed that the oropharyngeal samples had higher diversity than the sputum samples in AECOPD, in contrast with previous studies reporting no differences in diversity between these two specimen types (Zemanick et al., 2015). This discrepancy may reflect influences of antibiotic use. The observed beta diversity in the bacterial/fungal communities according to the day of hospitalization indicates that the differences between the individuals were far greater than the differences between the sputum and oropharyngeal samples. Furthermore, the bacterial communities were found to have characteristic structures in different individuals (Wang et al., 2016). The fungal community structure also showed that the differences between individuals were greater than the differences between sampling points. Previous research on microbial communities in oropharyngeal swab and sputum samples has demonstrated high abundances of *Haemophilus*, *Prevotella*, and *Streptococcus* (Huang and Boushey, 2015; Ogorodova et al., 2015). These common genera were also detected in our study, with high linear correlations in the paired samples. Further, in this study, the bacterial community compositions were found to vary identically in each paired sample analyzed. Moreover, each microbiome with an abundance of more than 0.001 was shared between the sputum and oropharyngeal samples.

It has previously been shown that Proteobacteria, Firmicutes, and Actinomycetes account for a large proportion of the species present in the lower respiratory tract of moderate to severe COPD patients (44, 16, and 13%, respectively), which is similar to our results regarding these three phyla (Garcia-Nunez et al., 2014). According to previous studies, the relative abundance of *Actinomycetes* was 2% in the lower respiratory tract of COPD patients and 10% in the lower respiratory tract of healthy individuals. Similarly, the relative abundances of *Prevotella* in COPD patients and healthy individuals were 4.2 and 13.42%, respectively (Park et al., 2014; Einarsson et al., 2016). Our results for *Actinomycetes* in the oropharyngeal and sputum samples were similar to those of previous studies. In addition, the microbial composition of the sputum and oropharyngeal samples was consistent between pairs of samples collected each day. In our study, *Psychrobacter*, *Lactobacillus*, *Rothia*, *Prevotella*, *Neisseria*, *Streptococcus*, *Haemophilus*, *Actinomyces*, *Leptotrichia*, and *Aspergillus* were observed to have high relative abundances in both the sputum and oropharyngeal samples in individuals with severe AECOPD. According to previous reports, oropharyngeal and sputum samples from COPD patients have high abundances of *Haemophilus*, *Streptococcus,* and *Prevotella,* whereas sputum samples from COPD patients have high abundances of *Psychrobacter* and *Neisseria* (Aguirre et al., 2015; Huang and Boushey, 2015; Ogorodova et al., 2015; Wang et al.,

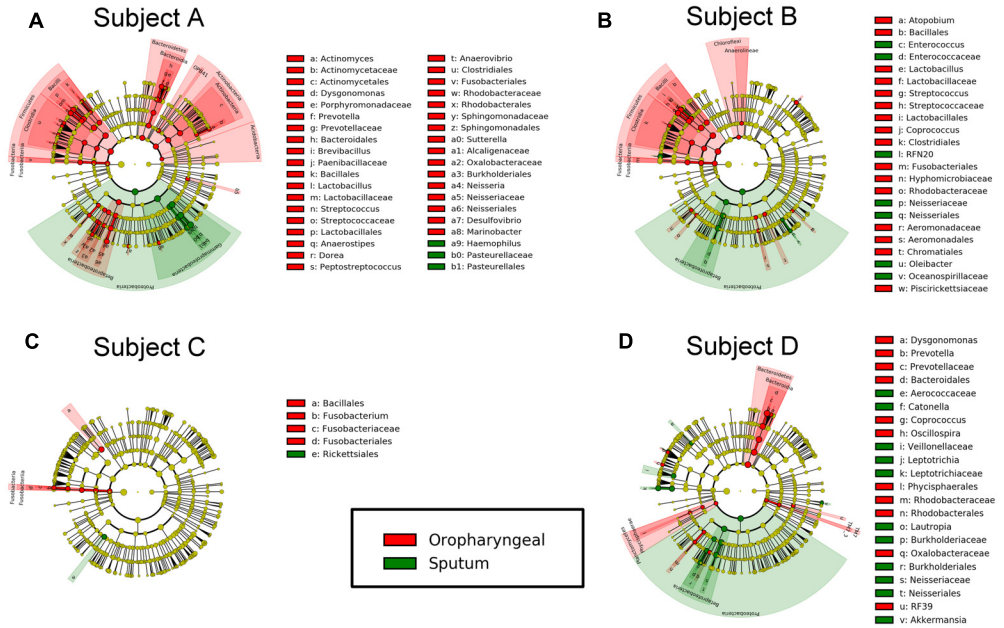


FIGURE 5 | Differential features in bacteria selected via LEfSe between the sputum and oropharyngeal samples from each individual. **(A)** For Subject A, **(B)** for Subject B, **(C)** for Subject C, **(D)** for Subject D. Red represents the oropharyngeal samples, and green represents the sputum samples.

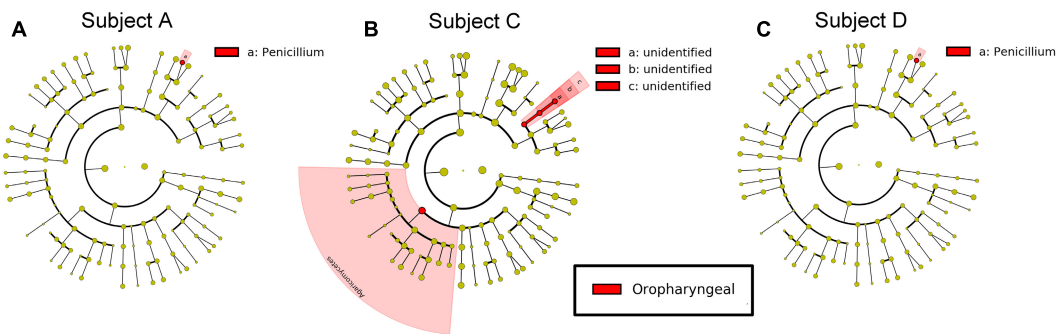


FIGURE 6 | Differential features in fungi selected via LEfSe between the sputum and oropharyngeal samples from each individual. **(A)** For Subject A, **(B)** for Subject C, **(C)** for Subject D. Red represents the oropharyngeal samples.

2016). In previous reports, *Aspergillus* has been found to be associated with severe AECOPD, and its presence can lead to decreased lung function in affected patients (Morris et al., 2008; Barberan and Mensa, 2014). We compared the correlation between oropharyngeal and sputum samples for taxa with relative abundances above 10% and found that the two sample types were highly correlated with regard to the structure of these major taxa. Indeed, 70.59% of the bacteria had a high statistical correlation. This further supports the finding that the two sample types are consistent in describing the microbial community of AECOPD patients.

For bacteria, *Prevotella*, *Dysgonomonas*, *Lactobacillus*, *Coprococcus*, *Streptococcus*, and *Fusobacteriales* from Bacteroidetes, Firmicutes, and Fusobacteria accounted for most of the increased taxonomic abundances in the oropharyngeal samples, which is consistent with the findings of previous studies. Among these, *Streptococcus*, *Lactobacillus*, and *Prevotella* are the most common genera in the oral cavity (Zaura et al., 2009). *Haemophilus* and *Lautropia* were significantly more abundant in the sputum samples than in the oropharyngeal samples in our study. In addition, healthy lower airways have been reported to possess a higher relative abundance of *Haemophilus* than upper airways (Morris et al., 2013). At the OTU level, *P. melaninogenica* and *L. iners* were significantly increased in the oropharyngeal samples. *P. melaninogenica* typically colonizes oral cavities and is transferred from maternal saliva to children shortly after birth (Kononen et al., 1994). Similarly, *L. iners* is a common oral cavity bacteria (Anderson et al., 2014). The presence of resident bacteria in the oral cavity may reduce the proportion of major pathogens in AECOPD to a certain extent.

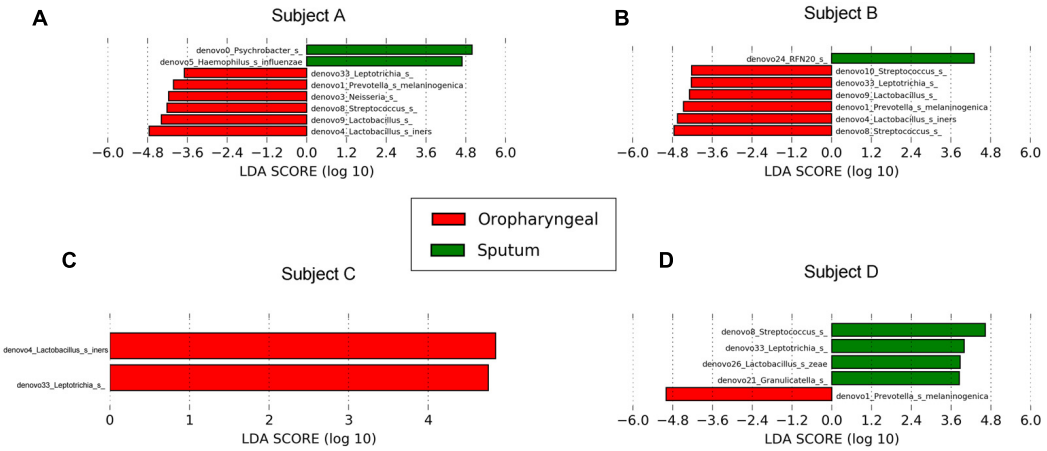


FIGURE 7 | Differential core OTUs with more than 10% relative abundance in bacteria selected via LEfSe between the sputum and oropharyngeal samples from each individual. **(A)** For Subject A, **(B)** for Subject B, **(C)** for Subject C, **(D)** for Subject D. Red represents the oropharyngeal samples, and green represents the sputum samples.

Subject A was treated with etimicin and cefoperazone, an aminoglycoside and β-lactam antibiotic, respectively. These antibiotics are sensitive to most Gram-positive and negative bacteria, especially *P. aeruginosa*, *H. influenza,* and *S. pneumoniae*. However, these antibiotics may fail to kill β-lactamase-producing bacteria; thus, they were replaced with piperacillin and sulbactam after 14 days for this subject. After the antibiotics were switched, we found a decreased abundance of *Psychrobacter*. This result is in agreement with previous reports showing that β-lactamase is secreted by *Psychrobacter* (Feller et al., 1995, 1997). The antibiotic treatment regimens were similar for Subjects B and A. We observed that the relative abundance of *Prevotella* was the lowest after the 1st day of treatment. In addition, the combined treatment with piperacillin and sulbactam inhibited *Psychrobacter*, *Actinomyces*, *Enterobacteriaceae,* and *Streptophyta*. These results showed that the distinctive spectrum of different antibiotics and antibiotic combination were relatively more efficient at killing the anti-pathogenic bacteria (Chandrasekaran et al., 2016). For the other subjects, meropenem or cefoperazone was administered based on their clinical conditions. Subject C displayed an increased abundance of *Stenotrophomonas,* which might have been caused by drug fast to the antibiotics (Zhao et al., 2017).

Previous studies have reported high levels of *Moraxella*, *Staphylococcus*, and *Pseudomonas* in COPD patients (Zakharkina et al., 2013), but this was not observed in our study. Most likely, this discrepancy occurred mainly because the included patients were being treated with antibiotics. It has also been shown that the V4 protocol used does not efficiently detect *Staphylococcus*, which might explain the low abundance of *Staphylococcus* observed in this study (Kong, 2016). The main microbial communities of the four included subjects were found to be quite unique. We examined four individuals to perform comprehensive day-to-day comparisons between sputum and oropharyngeal microbial communities for increased accuracy. In this study, we analyzed the ITS data with the classification of each individual sequence to generate a classification-based OTU table (with clustering at a threshold distance of 0), in contrast with a previous study with clustering at a threshold distance of 3 (Sokol et al., 2015). In this study, clustering at 0% difference increased the fungal signals and showed a more meaningful pattern in the fungal signals. Thus, we think clustering at 0% difference is suitable for analysis of our ITS data.

CONCLUSION

The airway microbial communities were similar in terms of the main phylum and genus compositions, and the oropharyngeal swab and sputum sample dynamics were similar, demonstrating similar bacterial community structures. For oral bacterial colonization, the oropharyngeal bacterial community diversity was higher than that observed in the sputum samples. These results suggest that the sputum microbiome is remarkably similar to the oropharyngeal microbiome; thus, oropharyngeal swabs can potentially be used instead of sputum samples for patients with exacerbation of COPD.

AUTHOR CONTRIBUTIONS

Substantial contributions to the conception or design of the work or the acquisition, analysis, or interpretation of data for the work: H-YL, S-YZ, W-YY, H-WZ, YH, X-FS and JS. Drafting the work or revising it critically for important intellectual content: H-WZ, YH, X-FS, JS, H-YL, S-YZ and W-YY. Final approval of the version to be published: H-WZ, JS, H-YL, YH, X-FS, S-YZ and W-YY. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of

any part of the work are appropriately investigated and resolved: W-YY, H-YL, S-YZ, H-WZ, YH, X-FS, and JS.

FUNDING

This work was supported by State's Key Project of Research and Development Plan (2017YFSF110078), the National Natural Science Foundation of China (NSFC31570497), the Open Project of the State Key Laboratory of Respiratory Disease (SKLRD2016OP014), and the Guangzhou Healthcare Collaborative Innovation Major Project (201604020012).

ACKNOWLEDGMENTS

We gratefully acknowledge Zhen-Yu Liang from the State Key Laboratory of Respiratory Disease, Guangzhou Institute of Respiratory Disease, First Affiliated Hospital of Guangzhou Medical University for guidance in interpreting the data described in this work.

REFERENCES

Aguirre, E., Galiana, A., Mira, A., Guardiola, R., Sanchez-Guillen, L., Garcia-Pachon, E., et al. (2015). Analysis of microbiota in stable patients with chronic obstructive pulmonary disease. *APMIS* 123, 427–432. doi: 10.1111/apm.12363

Al-Hebshi, N. N., Nasher, A. T., Idris, A. M., and Chen, T. (2015). Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples. *J. Oral Microbiol.* 7, 28934–28942. doi: 10.3402/jom.v7.28934

Anderson, A. C., Sanunu, M., Schneider, C., Clad, A., Karygianni, L., Hellwig, E., et al. (2014). Rapid species-level identification of vaginal and oral lactobacilli using MALDI-TOF MS analysis and 16S rDNA sequencing. *BMC Microbiol.* 14:312. doi: 10.1186/s12866-014-0312-5

Barberan, J., and Mensa, J. (2014). Invasive pulmonary aspergillosis in patients with chronic obstructive pulmonary disease. *Rev. Iberoam. Micol.* 31, 237–241. doi: 10.1016/j.riam.2014.07.004

Botero, L. E., Delgado-Serrano, L., Cepeda, M. L., Bustos, J. R., Anzola, J. M., Del Portillo, P., et al. (2014). Respiratory tract clinical sample selection for microbiota analysis in patients with pulmonary tuberculosis. *Microbiome* 2, 29–35. doi: 10.1186/2049-2618-2-29

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Chandrasekaran, S., Cokol-Cakmak, M., Sahin, N., Yilancioglu, K., Kazan, H., Collins, J. J., et al. (2016). Chemogenomics and orthology-based design of antibiotic combination therapies. *Mol. Syst. Biol.* 12, 872–893. doi: 10.15252/msb.20156777

Charlson, E. S., Bittinger, K., Haas, A. R., Fitzgerald, A. S., Frank, I., Yadav, A., et al. (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am. J. Respir. Crit. Care Med.* 184, 957–963. doi: 10.1164/rccm.201104-0655OC

Delhaes, L., Monchy, S., Frealle, E., Hubans, C., Salleron, J., Leroy, S., et al. (2012). The airway microbiota in cystic fibrosis: a complex fungal and bacterial community–implications for therapeutic management. *PLoS ONE* 7:e36313. doi: 10.1371/journal.pone.0036313

Dmitrieva, L. I. (2013). Diagnostic algorithm and quality indicator for the radiodiagnosis of chronic obstructive pulmonary disease on the principles of evidence-based medicine. *Vestn. Rentgenol. Radiol.* 196, 50–56.

Dy, R., and Sethi, S. (2016). The lung microbiome and exacerbations of COPD. *Curr. Opin. Pulm. Med.* 22, 196–202. doi: 10.1097/MCP.0000000000000268

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Einarsson, G. G., Comer, D. M., Mcilreavey, L., Parkhill, J., Ennis, M., Tunney, M. M., et al. (2016). Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax* 71, 795–803. doi: 10.1136/thoraxjnl-2015-207235

Feller, G., Sonnet, P., and Gerday, C. (1995). The beta-lactamase secreted by the antarctic psychrophile *Psychrobacter immobilis* A8. *Appl. Environ. Microbiol.* 61, 4474–4476.

Feller, G., Zekhnini, Z., Lamotte-Brasseur, J., and Gerday, C. (1997). Enzymes from cold-adapted microorganisms. The class C beta-lactamase from the antarctic psychrophile *Psychrobacter immobilis* A5. *Eur. J. Biochem.* 244, 186–191. doi: 10.1111/j.1432-1033.1997.00186.x

Garcia-Nunez, M., Millares, L., Pomares, X., Ferrari, R., Perez-Brocal, V., Gallego, M., et al. (2014). Severity-related changes of bronchial microbiome in chronic obstructive pulmonary disease. *J. Clin. Microbiol.* 52, 4217–4223. doi: 10.1128/JCM.01967-14

He, Y., Zhou, B. J., Deng, G. H., Jiang, X. T., Zhang, H., and Zhou, H. W. (2013). Comparison of microbial diversity determined with the same variable tag sequence extracted from two different PCR amplicons. *BMC Microbiol.* 13:208. doi: 10.1186/1471-2180-13-208

Hill, A. T., Campbell, E. J., Hill, S. L., Bayley, D. L., and Stockley, R. A. (2000). Association between airway bacterial load and markers of airway inflammation in patients with stable chronic bronchitis. *Am. J. Med.* 109, 288–295. doi: 10.1016/S0002-9343(00)00507-6

Huang, Y. J., and Boushey, H. A. (2015). The sputum microbiome in chronic obstructive pulmonary disease exacerbations. *Ann. Am. Thorac. Soc.* 12(Suppl. 2), S176–S180. doi: 10.1513/AnnalsATS.201506-319AW

Huang, Y. J., Sethi, S., Murphy, T., Nariya, S., Boushey, H. A., and Lynch, S. V. (2014). Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *J. Clin. Microbiol.* 52, 2813–2823. doi: 10.1128/JCM.00035-14

Huerta, A., Soler, N., Esperatti, M., Guerrero, M., Menendez, R., Gimeno, A., et al. (2014). Importance of *Aspergillus* spp. isolation in Acute exacerbations of severe COPD: prevalence, factors and follow-up: the FUNGI-COPD study. *Respir Res.* 15, 17–25. doi: 10.1186/1465-9921-15-17

Kim, V., Garfield, J. L., Grabianowski, C. L., Krahnke, J. S., Gaughan, J. P., Jacobs, M. R., et al. (2011). The effect of chronic sputum production on respiratory symptoms in severe COPD. *COPD* 8, 114–120. doi: 10.3109/15412555.2011.558546

Kong, H. H. (2016). Details matter: designing skin microbiome studies. *J. Invest. Dermatol.* 136, 900–902. doi: 10.1016/j.jid.2016.03.004

Kononen, E., Saarela, M., Karjalainen, J., Jousimies-Somer, H., Alaluusua, S., and Asikainen, S. (1994). Transmission of oral *Prevotella melaninogenica* between a mother and her young child. *Oral Microbiol. Immunol.* 9, 310–314. doi: 10.1111/j.1399-302X.1994.tb00077.x

Mammen, M. J., and Sethi, S. (2016). COPD and the microbiome. *Respirology* 21, 590–599. doi: 10.1111/resp.12732

Miller, M. R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., et al. (2005). Standardisation of spirometry. *Eur. Respir. J.* 26, 319–338. doi: 10.1183/09031936.05.00034805

Montes de Oca, M., and Perez-Padilla, R. (2017). Global initiative for chronic obstructive lung disease (GOLD)-2017: the ALAT perspective. *Arch. Bronconeumol.* 53, 87–88. doi: 10.1016/j.arbr.2017.01.016

Morris, A., Beck, J. M., Schloss, P. D., Campbell, T. B., Crothers, K., Curtis, J. L., et al. (2013). Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am. J. Respir. Crit. Care Med.* 187, 1067–1075. doi: 10.1164/rccm.201210-1913OC

Morris, A., Sciurba, F. C., and Norris, K. A. (2008). Pneumocystis: a novel pathogen in chronic obstructive pulmonary disease? *COPD* 5, 43–51. doi: 10.1080/15412550701817656

Ogorodova, L. M., Fedosenko, S. V., Popenko, A. S., Petrov, V. A., Tyakht, A. V., Saltykova, I. V., et al. (2015). Comparison study of oropharyngeal microbiota in

case of bronchial asthma and chronic obstructive pulmonary disease in different severity levels. *Vestn. Ross. Akad. Med. Nauk* 70, 669–678.

Paggiaro, P. L., Chanez, P., Holz, O., Ind, P. W., Djukanovic, R., Maestrelli, P., et al. (2002). Sputum induction. *Eur. Respir. J. Suppl.* 37, 3s–8s.

Park, H., Shin, J. W., Park, S. G., and Kim, W. (2014). Microbial communities in the upper respiratory tract of patients with asthma and chronic obstructive pulmonary disease. *PLoS ONE* 9:e109710. doi: 10.1371/journal.pone.0109710

Sethi, S., and Murphy, T. F. (2008). Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N. Engl. J. Med.* 359, 2355–2365. doi: 10.1056/NEJMra0800353

Sokol, H., Leducq, V., Aschard, H., Pham, H. P., Jegou, S., Landman, C., et al. (2015). Fungal microbiota dysbiosis in IBD. *Gut* 66, 1039–1048. doi: 10.1136/gutjnl-2015-310746

Su, J., Liu, H. Y., Tan, X. L., Ji, Y., Jiang, Y. X., Prabhakar, M., et al. (2015). Sputum bacterial and fungal dynamics during exacerbations of severe COPD. *PLoS ONE* 10:e0130736. doi: 10.1371/journal.pone.0130736

van der Molen, T., and Cazzola, M. (2012). Beyond lung function in COPD management: effectiveness of LABA/LAMA combination therapy on patient-centred outcomes. *Prim. Care Respir. J.* 21, 101–108. doi: 10.4104/pcrj.2011.00102

Vignola, A. M., Rennar, S. I., Hargreave, F. E., Fah, J. V., Bonsignore, M. R., Djukanovic, R., et al. (2002). Standardised methodology of sputum induction and processing. Future directions. *Eur. Respir. J. Suppl.* 37, 51s–55s.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Wang, Z., Bafadhel, M., Haldar, K., Spivak, A., Mayhew, D., Miller, B. E., et al. (2016). Lung microbiome dynamics in COPD exacerbations. *Eur. Respir. J.* 47, 1082–1092. doi: 10.1183/13993003.01406-2015

Zakharkina, T., Heinzel, E., Koczulla, R. A., Greulich, T., Rentz, K., Pauling, J. K., et al. (2013). Analysis of the airway microbiota of healthy individuals and patients with chronic obstructive pulmonary disease by T-RFLP and clone sequencing. *PLoS ONE* 8:e68302. doi: 10.1371/journal.pone.0068302

Zaura, E., Keijser, B. J., Huse, S. M., and Crielaard, W. (2009). Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol.* 9:259. doi: 10.1186/1471-2180-9-259

Zemanick, E. T., Wagner, B. D., Robertson, C. E., Stevens, M. J., Szefler, S. J., Accurso, F. J., et al. (2015). Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods. *Ann. Am. Thorac. Soc.* 12, 221–229. doi: 10.1513/AnnalsATS.201407-310OC

Zhao, S., Yang, L., Liu, H., and Gao, F. (2017). *Stenotrophomonas maltophilia* in a university hospital of traditional Chinese medicine: molecular epidemiology and antimicrobial resistance. *J. Hosp. Infect.* doi: 10.1016/j.jhin.2017.04.001

Zhou, H. W., Li, D. F., Tam, N. F., Jiang, X. T., Zhang, H., Sheng, H. F., et al. (2011). BIPES, a cost-effective high-throughput method for assessing microbial diversity. *ISME J.* 5, 741–749. doi: 10.1038/ismej.2010.160

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Liu, Zhang, Yang, Su, He, Zhou and Su. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 April 2018
doi: 10.3389/fmicb.2018.00723

Oral Administration of Recombinant *Saccharomyces boulardii* Expressing Ovalbumin-CPE Fusion Protein Induces Antibody Response in Mice

*Ghasem Bagherpour[1], Hosnie Ghasemi[1], Bahare Zand[1], Najmeh Zarei[1], Farzin Roohvand[2], Esmat M. Ardakani[3], Mohammad Azizi[1] and Vahid Khalaj[1]**

[1] Department of Medical Biotechnology, Pasteur Institute of Iran, Tehran, Iran, [2] Department of Virology, Pasteur Institute of Iran, Tehran, Iran, [3] Department of Molecular Medicine, Pasteur Institute of Iran, Tehran, Iran

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific and Technical Research Council (CONICET), Argentina

Reviewed by:
Julio Villena, Centro de Referencia para Lactobacilos, Argentina
Pamela Del Carmen Mancha-Agresti, Universidade Federal de Minas Gerais, Brazil

***Correspondence:**
Vahid Khalaj
v_khalaj@yahoo.com

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: 14 January 2018
Accepted: 27 March 2018
Published: 13 April 2018

Citation:
Bagherpour G, Ghasemi H, Zand B, Zarei N, Roohvand F, Ardakani EM, Azizi M and Khalaj V (2018) Oral Administration of Recombinant Saccharomyces boulardii Expressing Ovalbumin-CPE Fusion Protein Induces Antibody Response in Mice. Front. Microbiol. 9:723. doi: 10.3389/fmicb.2018.00723

Saccharomyces boulardii, a subspecies of *Saccharomyces cerevisiae*, is a well-known eukaryotic probiotic with many benefits for human health. In the present study, a recombinant strain of *S. boulardii* was prepared to use as a potential oral vaccine delivery vehicle. In this sense, a ura3 auxotroph strain of *S. boulardii* CNCM I-745 (known as *S. cerevisiae* HANSEN CBS 5926, Yomogi®) was generated using CRISPR/Cas9 methodology. Then a gene construct encoding a highly immunogenic protein, ovalbumin (OVA), was prepared and transformed into the *ura3*$^{-}$ *S. boulardii*. To facilitate the transport of the recombinant immunogen across the intestinal barrier, a claudin-targeting sequence from *Clostridium perfringens* enterotoxin (CPE) was added to the C-terminus of the expression cassette. The recombinant *S. boulardii* strain expressing the OVA-CPE fusion protein was then administered orally to a group of mice, and serum IgG and fecal IgA levels were evaluated by ELISA. Our results demonstrated that anti-OVA IgG in serum significantly increased in test group ($P < 0.001$) compared to control groups (receiving wild type *S. boulardii* or PBS), and the fecal IgA titer was significantly higher in test group ($P < 0.05$) than control groups. In parallel, a recombinant *S. boulardii* strain expressing the similar construct lacking C-terminal CPE was also administered orally. The result showed an increased level of serum IgG in group receiving yeasts expressing the CPE negative construct compared to control groups; however, the fecal IgA levels did not increase significantly. In conclusion, our findings indicated that the yeast *S. boulardii*, as a delivery vehicle with possible immunomodulatory effects, and c-CPE, as a targeting tag, synergistically assist to stimulate systemic and local immunity. This proposed recombinant *S. boulardii* system might be useful in the expression of other antigenic peptides, making it as a promising tool for oral delivery of vaccines or therapeutic proteins.

Keywords: *Saccharomyces boulardii*, antigen delivery, recombinant ovalbumin, *Clostridium perfringens* enterotoxin, mucosal immunity

INTRODUCTION

The yeast *Saccharomyces boulardii* is a known GRAS (generally regarded as safe) microorganism with the probiotic activity against a wide range of microbial pathogens in intestinal lumen (Czerucka et al., 2007). This probiotic yeast is often marketed in lyophilized form, called as *S. boulardii lyo*. The whole genome of *S. boulardii* has been sequenced, and the comparative genome analysis has been carried out (Khatri et al., 2017).

From a clinical point of view, the results of several randomized controlled trials in adult patients have frequently confirmed the significant positive effect of this probiotic in the treatment of acute and chronic diseases of intestine (Berni et al., 2011).

Based on several published data, the administration of *S. boulardii* in mice demonstrated immunomodulatory effects and resulted in increased levels of secretory IgA and serum IgG (Rodrigues et al., 2000; Qamar et al., 2001) as well as serum IgM (Stier and Bischoff, 2016). These immunological effects along with other probiotic features of *S. boulardii*, such as bile resistance, acid resistance, and the optimum growth temperature of 37°C, make this host as a potential vehicle for oral delivery of vaccines and other therapeutics into the lumen of intestine (Czerucka et al., 2007; Edwards-Ingram et al., 2007; Kelesidis and Pothoulakis, 2012). However, in a recent report, it has been shown that the immunoactivity of the probiotic yeast is limited in a healthy intestine as the majority of *S. boulardii* do not contact the gastrointestinal epithelium, and their uptake by Peyer's patches (PPs) is infrequent (Hudson et al., 2016a). Hence, it is necessary to make arrangements for efficient expression and delivery of antigens or therapeutic proteins to intestinal epithelium by *S. boulardii*.

Chicken ovalbumin (OVA) has two antigenic epitopes (OVA 257–264 and OVA 323–339) that bind to MHC class I and class II (Johnsen and Elsayed, 1990; Rotzschke et al., 1991). For efficient delivery of the recombinant antigen to immune cells underlying intestinal epithelium, an epithelium targeting ligand, *Clostridium perfringens* enterotoxin (CPE), was chosen to target claudin. Claudins are a family of integral membrane proteins in tight junctions (TJs) with at least 24 members (Lu et al., 2013). Among them, claudin-3 and -4 can be targeted by CPE, which has been shown to act as a TJs modulator and to increase the adsorption of fused proteins by losing TJs in intestinal epithelium (Suzuki et al., 2012). Detailed studies, however, have demonstrated that the carboxy-terminal region of CPE (c-CPE, amino acids 194–319) binds claudins, while the NH_2-terminal part of the molecule forms pores in the plasma membrane and induces cell death (Van Itallie et al., 2008). Here, the c-CPE sequence was fused C-terminally to OVA sequence, and the whole gene construct was expressed in *S. boulardii* CNCM I-745 as a probiotic yeast (Moré and Swidsinski, 2015; Stier and Bischoff, 2016; Kabbani et al., 2017). The recombinant *S. boulardii* was orally administered to C57BL/6 mice, and immunological assays were performed to evaluate possible immune responses.

MATERIALS AND METHODS

Strains and Plasmids

Saccharomyces boulardii CNCM I-745 (Yomogi®) was used for the construction of ura3 auxotroph mutant, the expression of OVA fusion protein, and animal studies. *Saccharomyces cerevisiae* BY4742 was employed as a control strain in transformation and expression experiments. *Escherichia coli* strain Top 10F′ was used as a host for plasmid preparations. Yeasts strains were grown and kept in the YPD medium (1% yeast extract, 2% polypeptone, and 2% dextrose). Yeast nitrogen base medium (YNB; Sigma-Aldrich, United States) was prepared at a concentration of 0.67% and was supplemented with 2% glucose. For the selection of *ura3*$^-$ strains, YNB medium was supplemented with 0.1% 5-fluoroorotic acid (5-FOA, Sigma-Aldrich), and 10 mM uracil (YNB-FOA-U). The pGEM®-T Easy cloning vector (Promega, United States) was used as intermediate vector for cloning of various DNA fragments. The episomal yeast plasmid, pYES2 (Invitrogen, United States), containing *S. cerevisiae ura3* gene and the inducible Gal1 promoter was utilized as a control plasmid in transformation experiments as well as in preparation of expression cassettes. Plasmids carrying Cas9 (pTEF1p-Cas9-CYCt1_kanMX; Stovicek et al., 2015) and *ura3* gRNA (pTAJAK-98; Jakočiūnas et al., 2015) constructs were used in the preparation of *ura3* auxotroph yeast strain and were kindly provided by Irina Borodina (Technical University of Denmark).

Restriction endonucleases and T4 DNA ligase were obtained from Fermentas (Waltham, United States). Primers were synthesized by SinaClon BioScience (Tehran, Iran). All chemicals and reagents used were purchased from standard commercial sources.

Preparation of ura3 Auxotroph Mutant of *S. boulardii*

A targeted gene inactivation method using CRISPR/Cas9 system was used to prepare the *ura3*$^-$ strain of *S. boulardii*. Plasmids pTEF1p-Cas9-CYCt1_kanMX and pTAJAK-98 were co-transformed into the wild-type *S. boulardii* using a standard electroporation method (Benatuil et al., 2010; Hudson et al., 2014). The electroporated cells were plated on YNB-FOA-U agar medium, and *ura3*$^-$ strains were recovered after 3 days at 30°C.

Preparation of Different Ovalbumin Expression Constructs

The OVA amino acid sequence was retrieved from UniProt (accession number: P01012). An N-terminal secretion signal from *S. cerevisiae* alpha-mating factor (MF, pre-pro-region) was then added to OVA sequence. A 6×His tag sequence was also added to the N-terminus of OVA sequence immediately after MF sequence. For the targeted attachment of the expressed protein to TJs of intestinal epithelium, the c-terminal region of CPE (amino acids 194–319) was joined to OVA C-terminus using a (G4S)3 linker. The complete amino acid sequence of fusion protein (MF-6×His-OVA-G4S-CPE) was back-translated to DNA sequence, and the resulting sequence was codon optimized according to *S. cerevisiae* codon usage. To facilitate

the cloning procedures, the sequence was further modified to contain *Sal*I and *Spe*I restriction sites at 5′ and 3′ ends, respectively. The sequence also carried two *Hin*dIII restriction sites at 5′ and 3′ ends of c-CPE for removing this fragment when required. The final fusion gene (FG) fragment (1667 bp) was synthesized commercially (Generay Biotech, Shanghai). To construct episomal expression cassettes, the synthetic FG was cloned into *Sal*I/*Spe*I site of pYES2 vector containing Gal1 promoter. Gal1 is one of the widely used inducible promoters in yeast expression vectors such as pYES and pESC series introduced by Invitrogen and Agilent, respectively. The resulting plasmid was called pYES2-Gal1-FG plasmid and used for subsequent construction of other episomal plasmids. For preparation of two new divergent expression cassettes, the Gal1 promoter in pYES2-FG was replaced by either Tef1 or Pgk1 promoters. Tef1 and Pgk1 are two strong constitutive promoters with a constant activity at various concentration of glucose (Partow et al., 2010). To prepare these vectors, both Tef1 and Pgk1 promoters were PCR amplified by High Fidelity Taq polymerase (TransGen Biotech Co., Ltd., United Kingdom). The Tef1 promoter (402 bp) was amplified from pPICZα plasmid (Thermo Fisher Scientific, United States) using specific primers Tef1F/Tef1R (**Table 1**). Likewise, the Pgk1 promoter (992 bp) was amplified from *S. cerevisiae* genome (GenBank accession number: CP020125.1, nucleotides 142078–143034) using Pgk1F/PgkR primers. These primer sets were designed in a way to add *Sac*I and *Sal*I restriction sites at 5′ and 3′ ends of PCR products, respectively. Each amplified promoter was separately cloned into the *Sac*I/*Sal*I site of pYES2-Gal1-FG, and the resulting plasmids were called as pYES2-Tef1-FG and pYES2-Pgk1-FG (**Figure 1**). For preparation of episomal construct lacking c-CPE fragment, the pYES2-Tef1-FG plasmid was digested by *Hin*dIII to cut out the c-CPE sequence and the resulting plasmid was religated to end up with pYES2-Tef1-FG-w/o-cpe.

To prepare the integrative expression plasmid, the whole sequence of Tef1-MF-Ova-CPE was cut out from pYES2-Tef-FG vector using *Sac*I/*Spe*I digestion. This fragment was subsequently cloned into the respective sites of pGEM-T easy vector. To add the *ura3* selection marker to this integrative construct, the complete sequence of *ura3* gene was amplified from pYES2 vector using Ura3F/Ura3R primers (**Table 1**) and cloned into the *Sph*I site of the latter vector, yielding pIP-FG1. To make CPE-less expression cassette, the whole sequence of c-CPE fragment was released from pIP-FG1 by *Hin*dIII digestion, and the remaining vector was religated using T4 DNA ligase. This CPE-negative construct was called pIP-FG2 (**Figure 2**).

TABLE 1 | Primers used in this study.

Name	Primers
ova-F	AACTGCTGCTGACCAGGC
ova-R	TGAGAAATCTTCAAAGACTCGGCAG
Cpe-F	AGAGATTGAACTTGACCGACGC
cpe-R	CGTAATGATCTTTGACTCCGTCTCCC
Pgk-1F	atGAGCTCGGAAGTACCTTCAAAGAATG
Pgk-1R	tatGTCGACTTGTTTTATATTTGTTGTAAAAAG
Tef-1F	tatGAGCTCCACACACCATAGCTTCAAAATGTTTCTAC
Tef-1R	tatGTCGACCTTAGATTAGATTGCTATGCTTTC
ura3-F	TATGCATGCTCTAGTATAGTCTATAGTCCGTGG
ura3-R	TATGCATGCTTAGTTTTGCTGGCCGCATC

The underlined sequences represent the restriction enzyme site in each primer.

Yeast Transformation

All yeast strains were transformed using a standard electroporation protocol as described before (Benatuil et al., 2010; Hudson et al., 2014). Briefly, an overnight culture of yeast was refreshed and grown in YPD broth medium to reach an OD600 of 1.6. Cells were washed with ice cold water and a buffer containing 1 M sorbitol and 1 mM $CaCl_2$. Before electroporation, the cells were resuspended in 100 mM LiOAc/10 mM DTT solution and incubated at 30°C for 30 min. The cell suspension was then centrifuged, and the resulting pellet was resuspended in electroporation buffer (1 M sorbitol/1 mM $CaCl_2$). Electroporation was performed in a volume of 400 μl mixture of cells and plasmid DNA (1 μg) using Gene Pulser Xcell^{TM} electroporation system (Bio-Rad Laboratories, United States). The transformation mixture was then spread on the selective medium, and positive transformants were recovered after 3 days incubation at 30°C.

Analysis of Protein Expression

For the expression analysis of recombinant OVA (with or without CPE), yeast transformants were cultivated in 50 ml synthetic defined-casamino acid medium (SD-CAA: 5 g/l casamino acids, 20 g/l dextrose, 1.7 g/l yeast nitrogen base without ammonium sulfate/without amino acid, 10.2 g/l $NaHPO_4$-$7H_2O$, and 8.6 g/l NaH_2PO_4-H_2O) and incubated at 37°C with shaking for 72 h. Afterward, the culture medium was collected by centrifugation and concentrated 50× using Amicon Ultra-15 Centrifugal Filtration system (Merck Millipore, Germany). The concentrated media (50 μl) were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the expression of recombinant OVA was confirmed by Western blotting using a 1:500 dilution of unlabeled rabbit anti-OVA antibody (Millipore, Chemicon). A horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (Razi Biotech Co., Iran) was used as the secondary antibody (1:2000 diluted) to detect the immunoreactive bands, followed by visualization with a chemiluminescence reagent (Amersham, United States). For the quantification of secreted recombinant protein, a His-Tag Protein assay Kit (Cell Biolabs, Inc., United States) based on competitive ELISA was used according to the protocol provided by the manufacturer.

Animal Study

Female C57BL/6 mice (6–8 weeks old) were provided by the animal facility of Pasteur Institute of Iran. The mice were housed at 23 ± 1°C with a 12-h light/dark cycle and were allowed free access to standard rodent chow and water. After their arrival, the mice were permitted to adapt to their environment for at least 1 week before the experiments. The study was approved by the Ethics Committee of Pasteur Institute of Iran

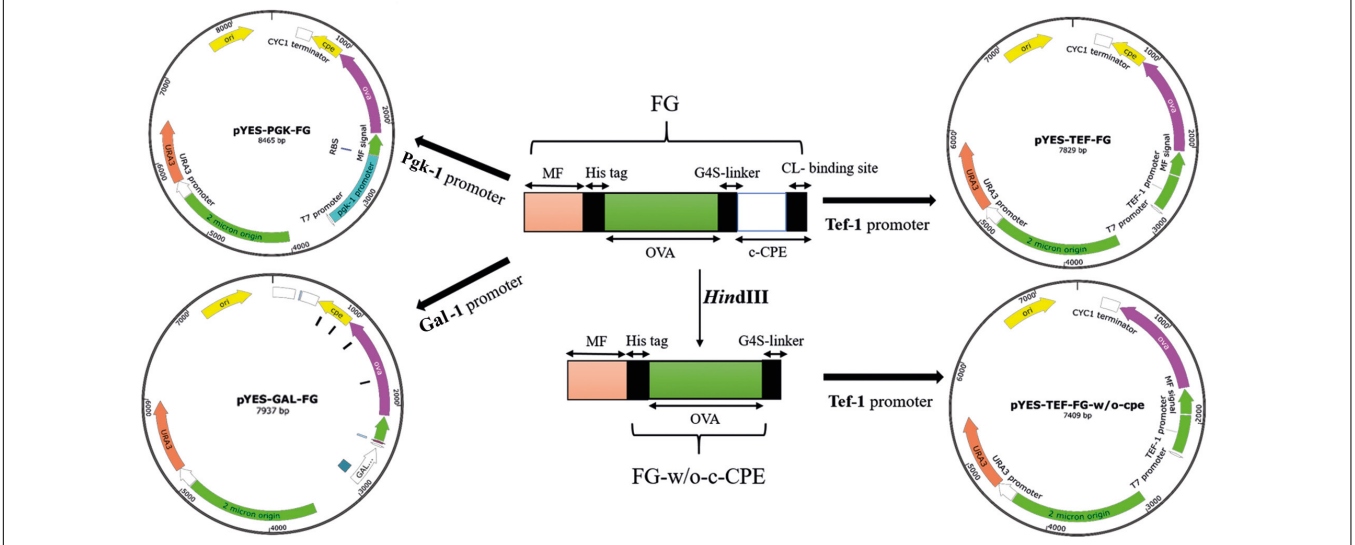


FIGURE 1 | The schematic representation of constructed vectors. Three promoters, Tef-1, Pgk-1, and Gal-1, were used to drive the expression of the fusion gene (FG) by pYES-based episomal plasmids. Tef-1 promoter was selected as a strong promoter in *S. boulardii* (Yomogi®) and used for preparation of the CPE-less construct (pYES-TEF-FG-w/o-cpe).

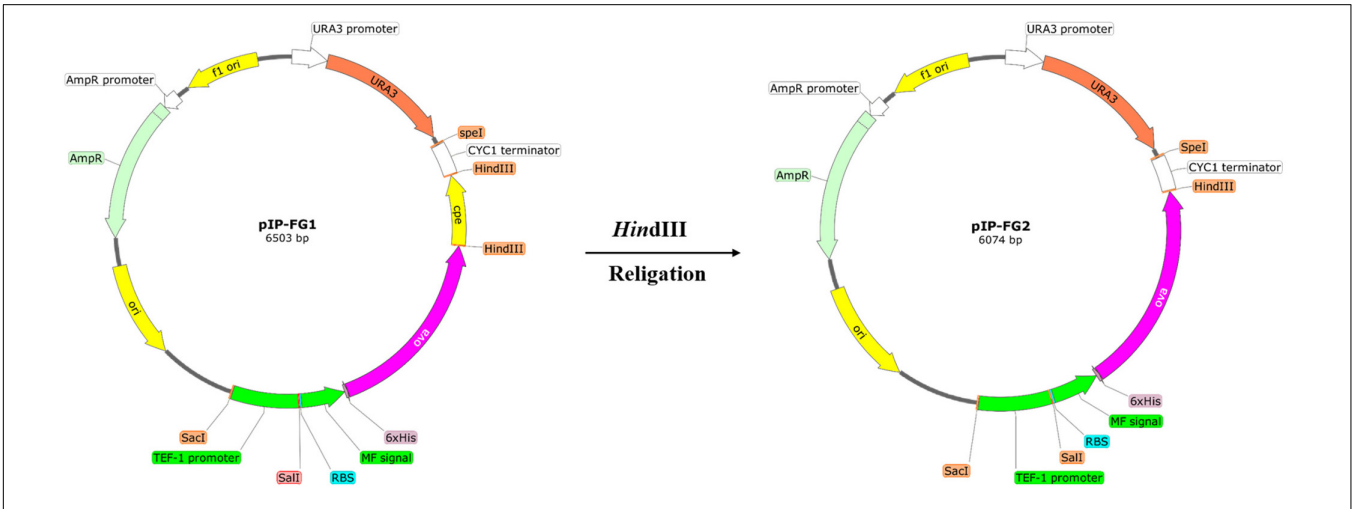


FIGURE 2 | The schematic representation of the constructed Integrative plasmids. pIP-FG1 was contained Tef1 promoter, OVA-CPE gene sequence and ura3 selection marker. pIP-FG2 was prepared by *Hin*dIII digestion and subsequent relegation of pIP-FG1.

and conforms to the European Communities Council Directive of 24 November 1986 (86/609/EEC). Oral immunizations were performed after an overnight fasting of the mice (water was provided *ad libitum*). An oral dose of 2×10^9 yeast cell/mouse was administered using gavage for four successive days, followed by a boost immunization with the same amount of yeasts four times with a 10-day interval (**Figure 3**; Shin et al., 2007). Previous studies have shown that *S. boulardii* reaches to a steady state concentration in colon after 3 days of oral administration and accordingly our feeding was continued for 4 days in each oral immunization step (Blehaut et al., 1989). In this oral vaccination study, the experimental groups of three mice were arranged as follows: group A (PBS), group B (yeast only), group C (yeast expressing CPE negative OVA), and group D (yeast expressing OVA-CPE).

Sample Preparation and Immunological Assays

Sample preparation was based on the previous works (Dion et al., 2004; Kakutani et al., 2010). Serum and fecal samples were collected 10 days after the last immunization. Fecal pellets (100 mg) were suspended in 1 ml of PBS buffer (pH 7.6) and in complete protease inhibitor cocktail tablets from Roche (Switzerland; one mini-tablet, EDTA-free, per 10 ml buffer) by vortexing for 10 min. Then samples were centrifuged at 3000 $\times$ *g* for 10 min, and the final supernatants were used

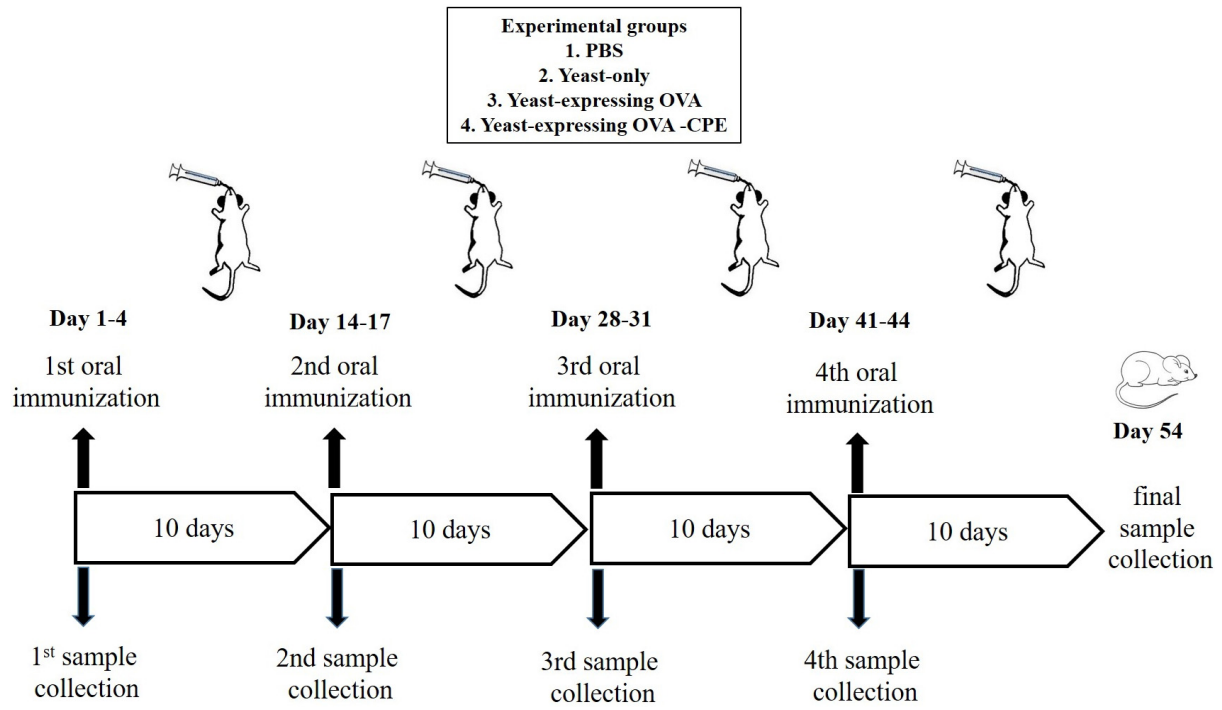


FIGURE 3 | Schematic representation of oral immunization protocol using recombinant *S. boulardii* (Yomogi®).

as fecal extracts. The titer of OVA-specific antibody in serum and extracts were determined by ELISA. For ELISA, a Maxisorb multiwall plate (Nunc, Roskilde, Denmark) was coated with chicken OVA (100 ng/well). The plates were washed with PBS containing 0.05% Tween 20 (PBS-T) and blocked with 200 µl blocking buffer (PBS-T containing 2% skim milk) at 37°C for 1 h. Subsequently, 10-fold serial dilutions of these samples were added to the immunoplate, followed by the addition of HRP-conjugated anti-mouse IgG and IgA. The OVA-specific antibodies were detected using TMB peroxide substrate. Absorbance of reactions was read at 450 nm. All ELISA tests were performed in duplicates.

Statistical Analysis

Statistical analysis of the total IgG and IgA levels were performed with Prism version 6.07 (Graph Pad Software, La Jolla, CA, United States). Experimental and control groups differences were analyzed by one-way analysis of variance (ANOVA), and the *P*-values of <0.05 were considered as statistically significant.

RESULTS

Generation of *S. boulardii* (Yomogi®) ura3 Auxotroph Strain

To generate a stable uracil auxotroph strain, the deletion of *ura3* gene was mediated by CRISPR/Cas9 system. Co-transforming of pTEF1p-Cas9-CYCt1_kanMX and pTAJAK-98 plasmids into *S. boulardii* resulted in several transformants on YNB-FOA-U agar medium. To confirm the inactivation of *ura3* gene, these transformants were grown on YNB and YNB-U media (**Figure 4**); the ura3 auxotrophs are not able to grow in YNB lacking uracil but grow normally in YNB-U. After primary confirmation, one colony was selected and analyzed further. This selected colony was transformed with pYES2 vector carrying *ura3* gene, and the resulting transformants were recovered on YNB medium. The complementation of *ura3*$^{-}$ mutant of *S. boulardii* by pYES2 confirmed the successful deactivation of *ura3* gene in this strain by CRISPR/Cas9 method. The growth phenotype of this ura3 auxotroph strain was compared with wild type, and the results did not detect any significant difference between the wild-type and the mutant strain (data not shown).

Expression Analysis of OVA-CPE Fusion Protein in *S. cerevisiae* and *S. boulardii* CNCM I-745

To validate our constructed vectors (pYES2-Gal1-FG, pYES2-Tef-FG, and pYES2-Pgk-FG) and select the most suitable promoter, all the pYES2-based vectors were individually transformed into *S. cerevisiae* BY4742, and the expression pattern was analyzed as described earlier. SD-CAA was used as expression medium for all constructs except for pYES2-Gal1-FG containing Gal1 promoter in which all transformants were cultivated in the presence of galactose rather than glucose. As shown in **Figure 5A**, in *S. cerevisiae* BY4742, all the episomal plasmids successfully expressed

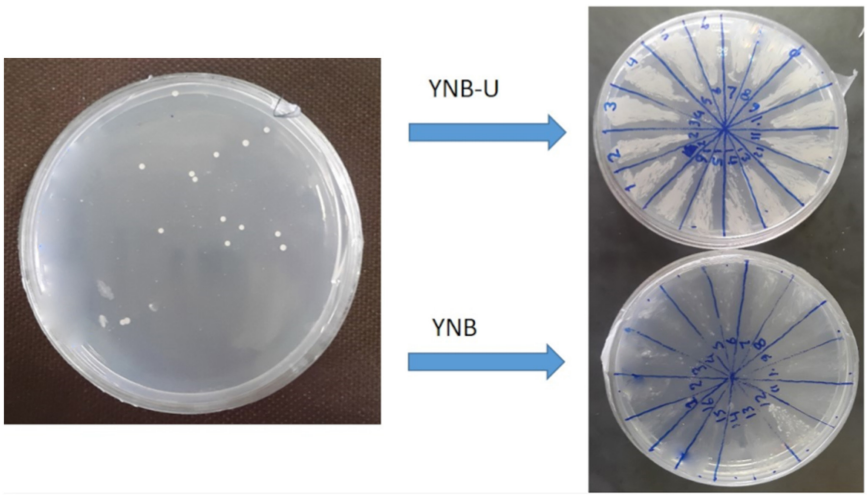


FIGURE 4 | Isolation of ura3 auxotroph strains of *S. boulardii*. The auxotroph mutants were able to grow in YNB-U medium but not in YNB.

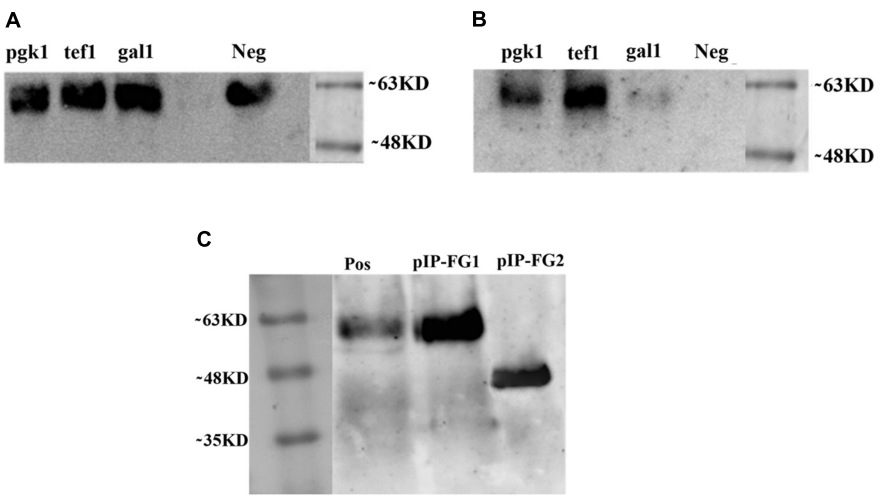


FIGURE 5 | Western blot analysis on culture supernatant of different recombinant yeasts. Constitutive promoters, Pgk-1 and Tef-1, and inducible promoter, Gal-1 were used in preparation of constructs. **(A)** Expression of OVA-CPE (∼60 kD) by *S cerevisiae* BY 4742 strain using episomal constitutive and inducible promoters. **(B)** Expression of same constructs in *S. boulardii* (Yomogi®) strain. **(C)** Expression of integrative constructs, pIP-FG1 and pIP-FG2 in *S. boulardii.* The smaller size of CPE-less construct (∼45 kD) is shown.

OVA-CPE fusion protein (∼60 kD). However, we could not detect any expression by pYES2-Gal1-FG construct in *S. boulardii* (**Figure 5B**). Based on quantitative analysis of His-Tag OVA, the yield of expressed protein by pYES2-Tef-FG construct was 4.7 μg/4.3 × 10^9cells/ml. The amount of secreted protein by pYES2-Pgk-FG construct in similar cultivation mode was 2.1/4.3 × 10^9cells/ml, indicating more efficiency of Tef1 promoter. The Western blot analysis results also showed a stronger band of the target protein when the Tef1 promoter was used. Hence, the Tef1 promoter was chosen for preparation of integrative construct, pIP-FG1. As it is indicated in **Figure 5C**, the Western blot analysis of culture supernatant from pIP-FG1 and pIP-FG2 transformants of *S. boulardii* confirmed the successful expression and the secretion of OVA-CPE and CPE negative OVA (∼45 kD) proteins into the culture medium.

Immunological Assay

To find whether the oral administration of the recombinant *S. boulardii* is able to stimulate any anti-OVA-specific IgG and IgA responses, both serum anti-OVA IgG and fecal anti-OVA IgA were measured by ELISA. The sera and feces of every mouse were collected 10 days after each oral intubation and analyzed. However, the level of antibodies did not rise, until 10 days after the last intubation. Those C57BL/6 mice that were fed either with yeast expressing OVA-CPE (pIP-FG1 construct transformant, $P < 0.001$) or with OVA alone (pIP-FG2 construct transformant, $P < 0.05$) showed a significant IgG response

compared to the control groups (wild-type *S. boulardii*-fed and PBS-fed groups). The measurement of fecal anti-OVA IgA antibody revealed that in the OVA-CPE treated group ($P < 0.05$), the IgA titers were significantly higher than that of the control group, however, there was no meaningful difference between IgA titers in mice that were fed yeast expressing OVA and controls (**Figure 6**).

DISCUSSION

Recently, attention has been focused on *S. boulardii* as a probiotic with the potential of engineering to produce a wide range of recombinant therapeutics (Hamedi et al., 2013; Hudson et al., 2014). Previous studies have highlighted the role of this yeast in the improvement of intestinal inflammatory or infective conditions. Three different mechanisms of action have been proposed for *S. boulardii*, including luminal action, trophic action, and mucosal-anti-inflammatory signaling effects (McFarland, 2010).

A recent report by Hudson et al. (2016a) has suggested that the therapeutic effect of *S. boulardii* is mainly attributed to the local activities of this yeast inside the lumen, including interfering with pathogen attachment or releasing antimicrobial toxins, but are not immune-mediated. Furthermore, in the same report, it has been found that *S. boulardii* does not induce any antibody response against its own antigens. The latter finding is an advantage in using this yeast for the repeated administration of various target antigens.

Mucosal vaccination is a simple and painless vaccination method that activates both the systemic and mucosal immune responses (Holmgren and Svennerholm, 2012). The oral administration route is particularly attractive as it is non-invasive and shows a good patient compliance (Zhu and Berzofsky, 2013). However, a major challenge for antigen proteins is to maintain their integrity while passing through the harsh environment of stomach and intestine. As a probiotic yeast, *S. boulardii* is able to survive in the acidic condition of stomach and to tolerate bile acids, and it is also capable of maintaining heterologous protein expression inside the intestine (Hudson et al., 2014). Hence, it can serve as a suitable carrier for the safe delivery of antigenic peptides to the intestinal lumen. Despite the above facts, there are some limiting factors such as the presence of intestinal protective mucus layer that hampers the efficient access of *S. boulardii* to immune system effectors (Edwards-Ingram et al., 2007). The gut-associated lymphoid tissue (GALT) is involved in antigen-specific mucosal immune responses in the gastrointestinal tract (Iijima et al., 2001). The initial uptake of antigens occurs in PPs of GALT, and few studies have demonstrated the infrequent presence of live *S. boulardii* in PPs after oral administration (Elmore, 2006; Hudson et al., 2014, 2016b). Moreover, the apical and basal TJs between intestinal epithelial cells act as a barrier, preventing macromolecules to access the paracellular spaces (Mayer, 2003).

Here, we hypothesized that the immune system can be stimulated if a target antigen released into the intestinal lumen by an engineered *S. boulardii*. Our target protein (OVA) was armed with c-CPE sequence to allow more penetration into the epithelium and more access to immune cells.

The first step in the construction of engineered *S. boulardii* was the generation of a ura3 auxotroph strain to use as a host for the transformation of various expression cassettes. Various methods, including classical UV mutagenesis or molecular tools such as Cre-loxP system, have been used for disruption of *ura3* in *S. boulardii* (Hamedi et al., 2013; Hudson et al., 2014; Wang et al., 2015). Here, we applied the CRISPR/Cas9 gene deletion method

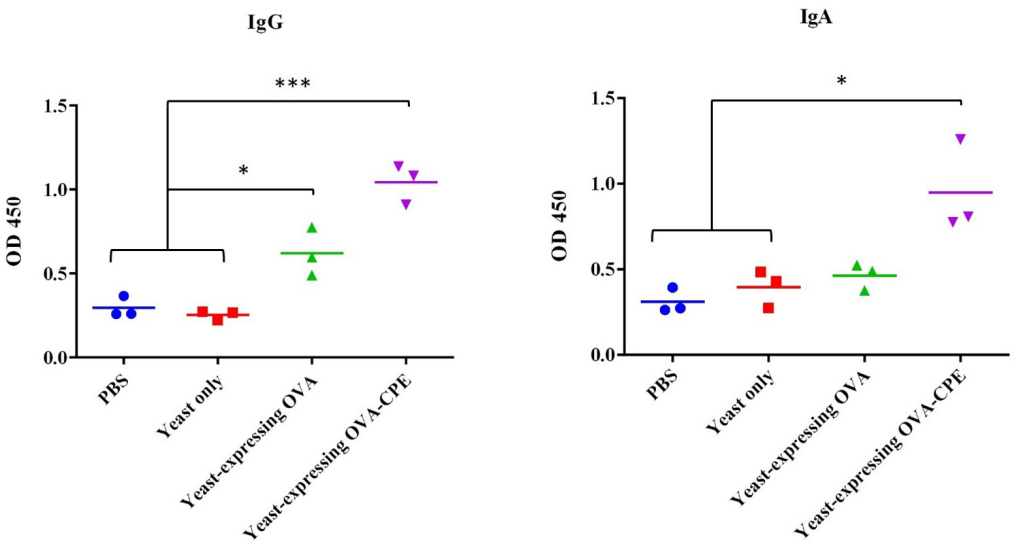


FIGURE 6 | Evaluation of antibody response in mice. Systemic IgG and local IgA immune assay was done for serum and fecal samples, respectively. Serum IgG in both yeast expressing OVA-CPE (pIP-FG1 construct) and yeast expressing OVA (pIP-FG2 construct) groups was significantly higher than control groups (PBS and yeast only). Fecal IgA was higher in pIP-FG1 group in comparison to pIP-FG2 and the control groups (*$P < 0.05$, ***$P < 0.001$).

to knock out the *ura3* gene activity. CRISPR/cas9 methodology is a simple, reliable, and cheap method of genome editing. It promotes the sequence-specific dsDNA breaks, followed by repair via non-homologous end joining or homologous recombination. This method is particularly efficient when the host contains more than one copy of the target gene (Wilkinson and Wiedenheft, 2014). Zhang et al. (2014) have constructed a quadruple auxotrophic mutant of *S. cerevisiae* using RNA-guided Cas9 nuclease. Recently, Liu et al. (2016) have used CRISPR/Cas9 gene editing system to create a triple auxotroph of *S. boulardii* used in metabolic engineering.

To drive the transcription of expression cassettes, three different promoters, GAL-1, transcriptional elongation factor EF-1α (Tef-1), and phosphoglycerate kinase (Pgk-1) were used. Our expression experiments showed that *S. cerevisiae* (BY4742) but not *S. boulardii* CNCM I-745 (Yomogi®) is able to express OVA when Gal-1 promoter is used. This finding is in agreement with other studies that demonstrated the lack of galactose utilization by *S. boulardii* (Mitterdorfer et al., 2001). Using two other promoters, Tef-1 and Pgk-1, our target protein was successfully expressed and secreted into the culture medium. These promoters have been examined before, and it has been shown that Tef-1 promoter had stronger activity in both glucose-consuming and glucose-exhausted phases compared to Pgk-1 promoter (Partow et al., 2010). Similarly, in our experiments, Tef-1 promoter resulted in a stronger expression pattern similar to Partow et al.'s (2010) report.

In our experiments, the genetically engineered *S. boulardii* was orally administered, and immunological assays confirmed both IgA and IgG responses to OVA. Although both OVA (group C) and OVA-CPE (group D) expressing yeasts could induce IgG response in mice, the highest antibody response was observed in group D ($P < 0.001$). An explanation could be the presence of CPE partner that targets CL-4 at TJs. CL-4 is highly expressed in follicle-associated epithelium (FAE) of PPs in GALTs (Tamagawa et al., 2003), and its targeting by CPE may facilitate the access of CPE-fused antigen to initiators of immune response. Targeting of claudins, especially CL-4, by CPE-fused OVA and its efficiency in mice immunization have been already proved by Suzuki et al. (2012). Regarding the local immune response, IgA was detected at higher levels in group D with a significant difference ($P < 0.05$) compared to other groups. In group C although IgA was induced against OVA, the difference was not significant compared to the control group. The latter observation was in agreement with a previous report on weak IgA response induced by CPE negative OVA in mucosal immunization (Suzuki et al., 2012).

Taken together, the results of the present study indicate the use of *S. boulardii* as a probiotic with the ability of delivering immunogenic or therapeutic proteins to intestinal lumen *via* oral administration. Similar studies using different engineered yeasts have also shown promising results, leading to the description of "Whole Yeast Vaccine" as a new platform for vaccine development (Shin et al., 2007; Wang et al., 2016; Roohvand et al., 2017). Although we did not investigate the presence of intact antigens in the intestinal lumen, the induction of both serum IgG and fecal IgA responses indicate the release and accessibility of the recombinant antigen to the immune effector cells. In this study, we have used c-CPE for tight junction targeting and augmentation of intestinal absorption of foreign antigens. However, this process can be further improved by addition of other ligands such as transcytotic peptides (TP) (Kang et al., 2008; Lee et al., 2015), which have been recommended for intestinal absorption of large molecules such as therapeutic peptides.

AUTHOR CONTRIBUTIONS

GB performed the experiments and wrote the first draft of the manuscript. HG helped in protein expression experiments. BZ helped in gene construct preparations. NZ helped in yeast transformation and the related set ups. FR helped in animal study design and immunological assays. EA supervised animal handling and immunization steps. MA helped in gene constructs' design. VK designed and coordinated the study, supervised all experimental steps, and revised and finalized the manuscript.

FUNDING

This work has been a part of Ph.D. Project of GB and financially supported by Pasteur Institute of Iran.

ACKNOWLEDGMENTS

We thank Dr. Irina Borodina from Technical University of Denmark, Kongens Lyngby, Novo Nordisk Foundation Center for Biosustainability for gifting CRISPR/Cas9 and gRNA vectors.

REFERENCES

Benatuil, L., Perez, J. M., Belk, J., and Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* 23, 155–159. doi: 10.1093/protein/gzq002

Berni, C. R., Cucchiara, S., Cuomo, R., Pace, F., and Papale, F. (2011). *Saccharomyces boulardii*: a summary of the evidence for gastroenterology clinical practice in adults and children. *Eur. Rev. Med. Pharmacol. Sci.* 15, 809–822.

Blehaut, H., Massot, J., Elmer, G., and Levy, R. (1989). Disposition kinetics of *Saccharomyces boulardii* in man and rat. *Biopharm. Drug Dispos.* 10, 353–364. doi: 10.1002/bdd.2510100403

Czerucka, D., Piche, T., and Rampal, P. (2007). Review article: yeast as probiotics–*Saccharomyces boulardii*. *Aliment. Pharmacol. Ther.* 26, 767–778. doi: 10.1111/j.1365-2036.2007.03442.x

Dion, C., Montagne, P., Bene, M. C., and Faure, G. (2004). Measurement of faecal immunoglobulin A levels in young children. *J. Clin. Lab. Anal.* 18, 195–199. doi: 10.1002/jcla.20022

Edwards-Ingram, L., Gitsham, P., Burton, N., Warhurst, G., Clarke, I., Hoyle, D., et al. (2007). Genotypic and physiological characterization of *Saccharomyces boulardii*, the probiotic strain of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 73, 2458–2467. doi: 10.1128/AEM.02201-06

Elmore, S. A. (2006). Enhanced histopathology of mucosa-associated lymphoid tissue. *Toxicol. Pathol.* 34, 687–696. doi: 10.1080/01926230600939989

Hamedi, H., Misaghi, A., Modarressi, M. H., Zahraei Salehi, T., Khorasanizadeh, D., and Khalaj, V. (2013). Generation of a uracil auxotroph strain of the probiotic yeast *Saccharomyces boulardii* as a host for the recombinant protein production. *Avicenna J. Med. Biotechnol.* 5, 29–34.

Holmgren, J., and Svennerholm, A.-M. (2012). Vaccines against mucosal infections. *Curr. Opin. Immunol.* 24, 343–353. doi: 10.1016/j.coi.2012.03.014

Hudson, L. E., Fasken, M. B., McDermott, C. D., McBride, S. M., Kuiper, E. G., Guiliano, D. B., et al. (2014). Functional heterologous protein expression by genetically engineered probiotic yeast *Saccharomyces boulardii*. *PLoS One* 9:e112660. doi: 10.1371/journal.pone.0112660

Hudson, L. E., McDermott, C. D., Stewart, T. P., Hudson, W. H., Rios, D., Fasken, M. B., et al. (2016a). Characterization of the probiotic yeast *Saccharomyces boulardii* in the healthy mucosal immune system. *PLoS One* 11:e0153351. doi: 10.1371/journal.pone.0153351

Hudson, L. E., Stewart, T. P., Fasken, M. B., Corbett, A. H., and Lamb, T. J. (2016b). Transformation of probiotic yeast and their recovery from gastrointestinal immune tissues following oral gavage in mice. *J. Vis. Exp.* 108:e53453. doi: 10.3791/53453

Iijima, H., Takahashi, I., and Kiyono, H. (2001). Mucosal immune network in the gut for the control of infectious diseases. *Rev. Med. Virol.* 11, 117–133. doi: 10.1002/rmv.307

Jakočiūnas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., et al. (2015). CasEMBLR: Cas9-facilitated multiloci genomic integration of in vivo assembled DNA parts in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 4, 1226–1234. doi: 10.1021/acssynbio.5b00007

Johnsen, G., and Elsayed, S. (1990). Antigenic and allergenic determinants of ovalbumin–III. MHC Ia-binding peptide (OA 323-339) interacts with human and rabbit specific antibodies. *Mol. Immunol.* 27, 821–827. doi: 10.1016/0161-5890(90)90147-R

Kabbani, T. A., Pallav, K., Dowd, S. E., Villafuerte-Galvez, J., Vanga, R. R., Castillo, N. E., et al. (2017). Prospective randomized controlled study on the effects of *Saccharomyces boulardii* CNCM I-745 and amoxicillin-clavulanate or the combination on the gut microbiota of healthy volunteers. *Gut Microbes* 8, 17–32. doi: 10.1080/19490976.2016.1267890

Kakutani, H., Kondoh, M., Fukasaka, M., Suzuki, H., Hamakubo, T., and Yagi, K. (2010). Mucosal vaccination using claudin-4-targeting. *Biomaterials* 31, 5463–5471. doi: 10.1016/j.biomaterials.2010.03.047

Kang, S. K., Woo, J. H., Kim, M. K., Woo, S. S., Choi, J. H., Lee, H. G., et al. (2008). Identification of a peptide sequence that improves transport of macromolecules across the intestinal mucosal barrier targeting goblet cells. *J. Biotechnol.* 135, 210–216. doi: 10.1016/j.jbiotec.2008.01.021

Kelesidis, T., and Pothoulakis, C. (2012). Efficacy and safety of the probiotic *Saccharomyces boulardii* for the prevention and therapy of gastrointestinal disorders. *Ther. Adv. Gastroenterol.* 5, 111–125. doi: 10.1177/1756283X11428502

Khatri, I., Tomar, R., Ganesan, K., Prasad, G., and Subramanian, S. (2017). Complete genome sequence and comparative genomics of the probiotic yeast *Saccharomyces boulardii*. *Sci. Rep.* 7:371. doi: 10.1038/s41598-017-00414-2

Lee, J.-Y., Kang, S.-K., Li, H.-S., Choi, C.-Y., Park, T.-E., Bok, J.-D., et al. (2015). Production of recombinant human growth hormone conjugated with a transcytotic peptide in *Pichia pastoris* for effective oral protein delivery. *Mol. Biotechnol.* 57, 430–438. doi: 10.1007/s12033-014-9835-0

Liu, J.-J., Kong, I. I., Zhang, G.-C., Jayakody, L. N., Kim, H., Xia, P.-F., et al. (2016). Metabolic engineering of probiotic *Saccharomyces boulardii*. *Appl. Environ. Microbiol.* 82, 2280–2287. doi: 10.1128/AEM.00057-16

Lu, Z., Ding, L., Lu, Q., and Chen, Y.-H. (2013). Claudins in intestines: distribution and functional significance in health and diseases. *Tissue Barriers* 1:e24978. doi: 10.4161/tisb.24978

Mayer, L. (2003). Mucosal immunity. *Pediatrics* 111(Suppl. 3), 1595–1600.

McFarland, L. V. (2010). Systematic review and meta-analysis of *Saccharomyces boulardii* in adult patients. *World J. Gastroenterol.* 16, 2202–2222. doi: 10.3748/wjg.v16.i18.2202

Mitterdorfer, G., Kneifel, W., and Viernstein, H. (2001). Utilization of prebiotic carbohydrates by yeasts of therapeutic relevance. *Lett. Appl. Microbiol.* 33, 251–255. doi: 10.1046/j.1472-765X.2001.00991.x

Moré, M. I., and Swidsinski, A. (2015). *Saccharomyces boulardii* CNCM I-745 supports regeneration of the intestinal microbiota after diarrheic dysbiosis–a review. *Clin. Exp. Gastroenterol.* 8, 237–255. doi: 10.2147/CEG.S85574

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., and Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* 27, 955–964. doi: 10.1002/yea.1806

Qamar, A., Aboudola, S., Warny, M., Michetti, P., Pothoulakis, C., LaMont, J. T., et al. (2001). *Saccharomyces boulardii* stimulates intestinal immunoglobulin a immune response to *Clostridium difficile* toxin a in mice. *Infect. Immun.* 69, 2762–2765. doi: 10.1128/IAI.69.4.2762-2765.2001

Rodrigues, A., Cara, D., Fretez, S., Cunha, F., Vieira, E., Nicoli, J., et al. (2000). *Saccharomyces boulardii* stimulates sIgA production and the phagocytic system of gnotobiotic mice. *J. Appl. Microbiol.* 89, 404–414. doi: 10.1046/j.1365-2672.2000.01128.x

Roohvand, F., Shokri, M., Abdollahpour-Alitappeh, M., and Ehsani, P. (2017). Biomedical applications of yeast- a patent view, part one: yeasts as workhorses for the production of therapeutics and vaccines. *Expert Opin. Ther. Pat.* 27, 929–951. doi: 10.1080/13543776.2017.1339789

Rotzschke, O., Falk, K., Stevanovic, S., Jung, G., Walden, P., and Rammensee, H. G. (1991). Exact prediction of a natural T cell epitope. *Eur. J. Immunol.* 21, 2891–2894. doi: 10.1002/eji.1830211136

Shin, S. J., Shin, S. W., Kang, M. L., Lee, D. Y., Yang, M. S., Jang, Y. S., et al. (2007). Enhancement of protective immune responses by oral vaccination with *Saccharomyces cerevisiae* expressing recombinant *Actinobacillus pleuropneumoniae* ApxIA or ApxIIA in mice. *J. Vet. Sci.* 8, 383–392. doi: 10.4142/jvs.2007.8.4.383

Stier, H., and Bischoff, S. C. (2016). Influence of *Saccharomyces boulardii* CNCM I-745on the gut-associated immune system. *Clin. Exp. Gastroenterol.* 9, 269–279. doi: 10.2147/CEG.S111003

Stovicek, V., Borodina, I., and Forster, J. (2015). CRISPR–Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. *Metab. Eng. Commun.* 2(Suppl. C), 13–22. doi: 10.1016/j.meteno.2015.03.001

Suzuki, H., Kondoh, M., Takahashi, A., and Yagi, K. (2012). Proof of concept for claudin-targeted drug development. *Ann. N. Y. Acad. Sci.* 1258, 65–70. doi: 10.1111/j.1749-6632.2012.06503.x

Tamagawa, H., Takahashi, I., Furuse, M., Yoshitake-Kitano, Y., Tsukita, S., Ito, T., et al. (2003). Characteristics of claudin expression in follicle-associated epithelium of Peyer's patches: preferential localization of claudin-4 at the apex of the dome region. *Lab. Invest.* 83, 1045–1053. doi: 10.1097/01.LAB.0000078741.55670.6E

Van Itallie, C. M., Betts, L., Smedley, J. G., McClane, B. A., and Anderson, J. M. (2008). Structure of the claudin-binding domain of *Clostridium perfringens* enterotoxin. *J. Biol. Chem.* 283, 268–274. doi: 10.1074/jbc.M708066200

Wang, L. J., Sun, H., Zhang, J., Liu, Q., Wang, T. T., Chen, P. P., et al. (2015). Establishment and application of target gene disruption system in *Saccharomyces boulardii*. *Biotechnol. Bioprocess Eng.* 20, 26–36. doi: 10.1007/s12257-014-0197-z

Wang, X., Wang, Z., Xu, H., Xiang, B., Dang, R., and Yang, Z. (2016). Orally administrated whole yeast vaccine against porcine epidemic diarrhea virus induced high levels of IgA response in mice and piglets. *Viral Immunol.* 29, 526–531. doi: 10.1089/vim.2016.0067

Wilkinson, R., and Wiedenheft, B. (2014). A CRISPR method for genome engineering. *F1000Prime Rep.* 6:3. doi: 10.12703/P6-3

Zhang, G.-C., Kong, I. I., Kim, H., Liu, J.-J., Cate, J. H., and Jin, Y.-S. (2014). Construction of a quadruple auxotrophic mutant of an industrial polyploid *Saccharomyces cerevisiae* strain by using RNA-guided Cas9 nuclease. *Appl. Environ. Microbiol.* 80, 7694–7701. doi: 10.1128/AEM.02310-14

Zhu, Q., and Berzofsky, J. A. (2013). Oral vaccines: directed safe passage to the front line of defense. *Gut Microbes* 4, 246–252. doi: 10.4161/gmic.24197

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bagherpour, Ghasemi, Zand, Zarei, Roohvand, Ardakani, Azizi and Khalaj. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 11 January 2018
doi: 10.3389/fbioe.2017.00084

Meta-Analysis of *Aedes aegypti* Expression Datasets: Comparing Virus Infection and Blood-Fed Transcriptomes to Identify Markers of Virus Presence

Kiyoshi Ferreira Fukutani[1†§], José Irahe Kasprzykowski[1,2§], Alexandre Rossi Paschoal[3], Matheus de Souza Gomes[4], Aldina Barral[1,5], Camila I. de Oliveira[1,5], Pablo Ivan Pereira Ramos[1‡] and Artur Trancoso Lopo de Queiroz[1,2,6*‡]*

[1] Instituto Gonçalo Moniz, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil, [2] Post-Graduation Program in Biotechnology in Health and Investigative Medicine, Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil, [3] Federal University of Technology—Paraná, UTFPR, Campus Cornélio Procópio, Cornélio Procópio, Brazil, [4] Federal University of Uberlândia, Patos de Minas, Brazil, [5] Post-Graduation Program in Health Sciences, School of Medicine, Federal University of Bahia, Salvador, Brazil, [6] Post-Graduation Program in Applied Computation, Universida de Estadual de Feira de Santana, Feira de Santana, Brazil

The mosquito *Aedes aegypti* (L.) is vector of several arboviruses including dengue, yellow fever, chikungunya, and more recently zika. Previous transcriptomic studies have been performed to elucidate altered pathways in response to viral infection. However, the intrinsic coupling between alimentation and infection were unappreciated in these studies. Feeding is required for the initial mosquito contact with the virus and these events are highly dependent. Addressing this relationship, we reinterrogated datasets of virus-infected mosquitoes with two different diet schemes (fed and unfed mosquitoes), evaluating the metabolic cross-talk during both processes. We constructed coexpression networks with the differentially expressed genes of these comparison: virus-infected versus blood-fed mosquitoes and virus-infected versus unfed mosquitoes. Our analysis identified one module with 110 genes that correlated with infection status (representing ~0.7% of the *A. aegypti* genome). Furthermore, we performed a machine-learning approach and summarized the infection status using only four genes (AAEL012128, AAEL014210, AAEL002477, and AAEL005350). While three of the four genes were annotated as hypothetical proteins, AAEL012128 gene is a membrane amino acid transporter correlated with viral envelope binding. This gene alone is able to discriminate all infected samples and thus should have a key role to discriminate viral infection in the *A. aegypti* mosquito. Moreover, validation using external datasets found this gene as differentially expressed in four transcriptomic experiments. Therefore, these genes may serve as a proxy of viral infection in the mosquito and the others 106 identified genes provides a framework to future studies.

Keywords: *Aedes aegypti*, alimentation, blood-feeding, meta-analysis, transcriptomics, vector-borne diseases, virus infection

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Xianwen Ren, Peking University, China
Anna Honko, Data Sciences International, United States

***Correspondence:**
Kiyoshi Ferreira Fukutani ferreirafk@gmail.com;
Artur Trancoso Lopo de Queiroz artur.queiroz@bahia.fiocruz.br

†Present address:
Kiyoshi Ferreira Fukutani, Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

‡These authors have contributed equally to the work.

§These authors are co-first author's.

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *23 October 2017*
Accepted: *15 December 2017*
Published: *11 January 2018*

Citation:
Fukutani KF, Kasprzykowski JI, Paschoal AR, Gomes MdS, Barral A, de Oliveira CI, Ramos PIP and Queiroz ATL (2018) Meta-Analysis of Aedes aegypti Expression Datasets: Comparing Virus Infection and Blood-Fed Transcriptomes to Identify Markers of Virus Presence. Front. Bioeng. Biotechnol. 5:84. doi: 10.3389/fbioe.2017.00084

INTRODUCTION

The mosquito *Aedes aegypti* (L.) is the main vector of dengue virus (DENV), West Nile virus (WNV), and Yellow fever virus (YFV), present worldwide (Mackenzie et al., 2004; Lorenzo et al., 2014); more than 2.5 billion people in over 100 countries are at risk of contracting dengue alone (World Health Organization, 2015), while yellow fever remains endemic in tropical regions of Africa and South America (Bae et al., 2005). West Nile fever, despite causing occasional small outbreaks, shows an extremely high mortality rate (Pradier et al., 2012). Beyond established pathogens, other viruses are also on the rise as public health problems: chikungunya virus, formerly restricted to parts of Africa, is now globally spread (Cauchemez et al., 2014). Zika virus has recently become a global concern, after initial outbreaks in the Pacific region in 2007, followed by a larger spread in the Americas (Roth et al., 2014; Zanluca et al., 2015), including Brazil (Morens and Fauci, 2014; Petersen et al., 2016; Slavov et al., 2016). Despite extensive vector control measures to curb transmission, including source reduction, pesticides, public education and biological control, these efforts were largely unsuccessful (Medlock et al., 2012), highlighting the need for enhanced control methods and better knowledge about mosquito biology (Seixas et al., 2013; Porse et al., 2015).

In 2007, the *A. aegypti* complete genome was released (Nene et al., 2007) and vector-specific databases were developed, such as Vector Base (Giraldo-Calderon et al., 2015). This allowed expression assays addressing viral infection (Colpitts et al., 2011), enabling new insights about the *A. aegypti* gene regulation and transcriptional processes (Dissanayake et al., 2010; Colpitts et al., 2011). Further studies focused on the gene expression profile related to mosquitoes blood feeding (females) compared to nonblood-fed (N-BF, males), suggesting that sex- and stage-specific genes play an important role on the feeding response (Dissanayake et al., 2010). During feeding, female mosquitoes acquire blood that is necessary for egg development and may subsequently become infected with pathogens. Both infection and blood feeding processes induce changes in gene expression levels. The intrinsic physiological crosstalk between these process are linked and a joint analysis is required to assess patterns of infection possibly unappreciated in previous studies, due to the difficulty of separating the gene expression patterns that arise from feeding on blood from that resultant of the infection process due to host–pathogen interactions.

To address this issue, we performed an integrated gene expression analysis of currently available data sets. One dataset is derived from mosquitoes infected with DENV, WNV, YFV, or uninfected and another from blood-fed (BF) or sugar-fed *A. aegypti*. We identified 110 genes specifically associated with an infection expression profile. Following data mining, we propose three main candidate genes (*AAEL014210*, *AAEL002477*, and *AAEL005350*) that relate to the infection caused by each virus and one gene, *AAEL0012128*, able to summarize the infection profile.

MATERIALS AND METHODS

Description of *A. aegypti* Discovery Dataset

We jointly analyzed two previously published microarray datasets, available from the GEO under accession n. GSE28208 and GSE22339. The Colpitts et al. (2011) dataset (GSE28208) reports *A. aegypti* mosquitoes (Rockefeller strain) artificially infected through intrathoracic inoculation with DENV (type 2), YFV, or WNV and uninfected controls sugar-fed with raisins. The Dissanayake et al. (2010) dataset (GSE22339) reports *A. aegypti* mosquitoes [Liverpool (LVP) strain] that were strictly sugar-fed with raisins or that, besides having access to sugar feeding with raisins, were also BF on anesthetized mice. That study investigated the differential gene expression during the feeding process. From the total of 61 gene expression samples reported by both studies, we used a subset of 58 samples in our combined analysis, as discovery dataset. **Table 1** summarizes the main characteristics of these samples.

To facilitate reading, in what follows we refer to the group of uninfected mosquitoes that did not receive blood as N-BF and to the group of uninfected mosquitoes that had access to a blood meal as BF. These are samples from GSE22339 dataset. The groups of mosquitoes infected by any of the three included viruses (DENV, WNV, or YFV) are referred to as "infected," independent of their feeding regimen and the uninfected mosquitoes were referred as uninfected. These are samples from GSE28208 dataset. In summary, our discovery dataset was composed by 28 virus-infected (DENV, YFV, and WNV) samples, 27 fed (18 with blood and 9 uninfected) and 3 N-BF uninfected samples.

Data Collection, Preprocessing and Correction

Raw expression data from the 58 samples in both studies were downloaded from the GEO database.[1] Quantile normalization was applied using the *preprocessCore*R package (R 3.2.2, R Foundation, Vienna, Austria). Only probes mapping to genes common to both datasets were kept. Since we performed a joint analysis of two different datasets of interest, the expression data was submitted to a correction procedure using an empirical Bayes framework implemented in the COMBAT tool (Johnson et al., 2007). COMBAT corrects for experimental variation, commonly known as batch effects. Combining microarray data sets makes it possible to increase statistical power when detecting biological phenomena from diverse experiments. The present samples were classified into two batches according to the origin of each dataset

[1]http://www.ncbi.nlm.nih.gov/geo/.

TABLE 1 | Description of *A. aegypti* studies used in the integrative analysis.

Dataset	Characteristics (infection status, sex)	Group classification	Number of samples
GSE28208 (Colpitts et al.)	DENV infected, nr	Infected	9
	YFV infected, nr	Infected	10
	WNV infected, nr	Infected	9
	Uninfected, nr	uninfected	9
GSE22339 (Dissanayake et al.)	Uninfected, ♀	BF	18
	Uninfected, ♀	N-BF	3

♀, *female; nr, sex not reported; DENV, dengue virus; WNV, West Nile virus; YFV, yellow fever virus; BF, blood-fed; N-BF, nonblood-fed.*

(either GSE22339 or GSE28208), resulting in a merged dataset with corrected expression values. After merging and batch effect correction, the expression table was log2-transformed and differentially expressed genes (DEGs) were identified using an absolute log2-fold-change threshold of ≥1.0, and *t*-tests comparisons were performed with the Benjamini–Hochberg false discovery rate adjustment for multiple testing set at 5%.

Coexpression Network Analysis

Weighted gene correlation network analysis (WGCNA) methodology was applied in the DEGs to construct a gene coexpression network with weighted interactions (Langfelder and Horvath, 2008). The biweight mid correlation algorithm implemented in the *bicor* function in WGCNA was used as correlation metric to compare gene expression values, being similar to Pearson's statistic but is more robust to outliers (Langfelder and Horvath, 2012). In the WGCNA framework, the correlation matrix is transformed into a weighted adjacency matrix by applying a power transformation, $f(x) = x^{\beta}$, where β is chosen such that the topology of the obtained adjacency matrix is approximately scale-free. Herein, the appropriate β parameter was set to 7. The scale-free model fitting index using this parameter was $R^2 = 0.85$. Next, a topological overlap matrix (TOM) was derived from the adjacency matrix, taking into account gene expression connectivity. 1-TOM was used as a dissimilarity metric for hierarchical clustering and the detection of coexpression modules. The dynamic tree cut algorithm within WGCNA, set at default parameters, was used for module assignment. Module eigengenes (MEs), the first principal component of all gene expression values in a module, which summarizes expression values in a given module, were tested with respect to associations with the traits of interest (infected by each virus, BF and N-BF), and those found to be significantly correlated (absolute Pearson's $r \geq 0.6$; p-value < 0.05) were further studied by means of functional analysis using Gene Ontology (GO) terms.

Functional Analysis Using GO Terms

Functional analysis of the significant genes were assessed by mapping GO terms extracted from Vector Base[2] (Giraldo-Calderon et al., 2015). GO provides a controlled vocabulary of molecular functions, biological processes and cellular localization. Revigo was used to summarize the GO terms, employing *Drosophila melanogaster* as a reference (Supek et al., 2011).

Data Mining

Decision trees were employed to identify a minimal set of gene expression measurements allowing separation between the infected from uninfected groups. This method analyzes all the phenotypic attributes (gene expression measurements) and selects the most relevant attributes that allow group classification (Sathler-Avelar et al., 2016). As input for tree construction, we used the 110 genes (and their expression values) that we identified as most related to infection independent of feeding background (available in Table S1 in Supplementary Material). The J48 algorithm implemented in the WEKA program (Waikato Environment for Knowledge Analysis, version 3.6.11, University of Waikato, New Zealand) was used to build a decision tree using default parameters (Espíndola et al., 2015). To estimate the classification accuracy of the decision tree models, we employed a 10-fold cross validation methodology. This methodology splits the dataset in a training set and testing set. The partition procedure is applied to avoid bias in sampling of training/test sets. Thus, the training set was used to tune the parameters, learning and building a model. The validation set was used to test the performance of the classifier in an unbiased way. The sensibility and specificity were measured from the confusion matrix and the receiver–operating characteristic curve (ROC).

Hierarchical Clustering and Principal Component Analysis (PCA)

Hierarchical clustering was performed with genes with significant expression differences using Euclidean distance as a measure of dissimilarity and average linkage for between-cluster separation (*hclust* function in the *stats* package in R 3.2.2). Heatmap was generated in R *via* the *heatmap.2* function in the *gplots* package, using the "scale = 'row'" switch to Z-score standardize the rows. The Z-score standardization measures the expression level of a gene in terms of number of SDs from the mean expression of the gene in all compared samples. PCA was performed in R 3.2.2 (function *prcomp*) in order to compare and visualize the grouping between infected, BF and N-BF samples using the gene expression data as input. The graphing package *ggplot2* (Wickham, 2016) was used to plot these results.

Description of Validation Dataset

Expression data from the studies of Behura et al. (2011), Sim and Dimopoulos (2010), Bonizzoni et al. (2011), and Bonizzoni et al. (2012) available in GEO (accession nos. GSE16563, GSE33274, GSE24872, and GSE32074, respectively) were used to evaluate whether the same expression pattern observed in gene *AAEL012128* in our results could be verified in independent datasets, since they were not used in the Discovery step. GEO2R was used to access this gene expression in each validation dataset. In summary, the total samples used in validation were 16 samples infected with DENV and 12 fed samples.

RESULTS

Blood-Feeding Triggers Differential Gene Expression Compared to Strictly Sugar-fed *A. aegypti*

We analyzed gene expression data from 58 samples stemming from two different studies, the Colpitts et al. (2011) and the Dissanayake et al. (2010) datasets. We merged both datasets, correcting for batch effects, and reanalyzed them with a focus toward the investigation of DEGs in: (i) BF vs. uninfected mosquitoes and (ii) BF vs. N-BF mosquitoes (**Figure 1**). BF and uninfected mosquitoes showed a similar gene expression pattern, without any differentially modulated genes (**Figure 1A**). However, comparison of BF vs. N-BF mosquitoes enabled the

[2]https://www.vectorbase.org.

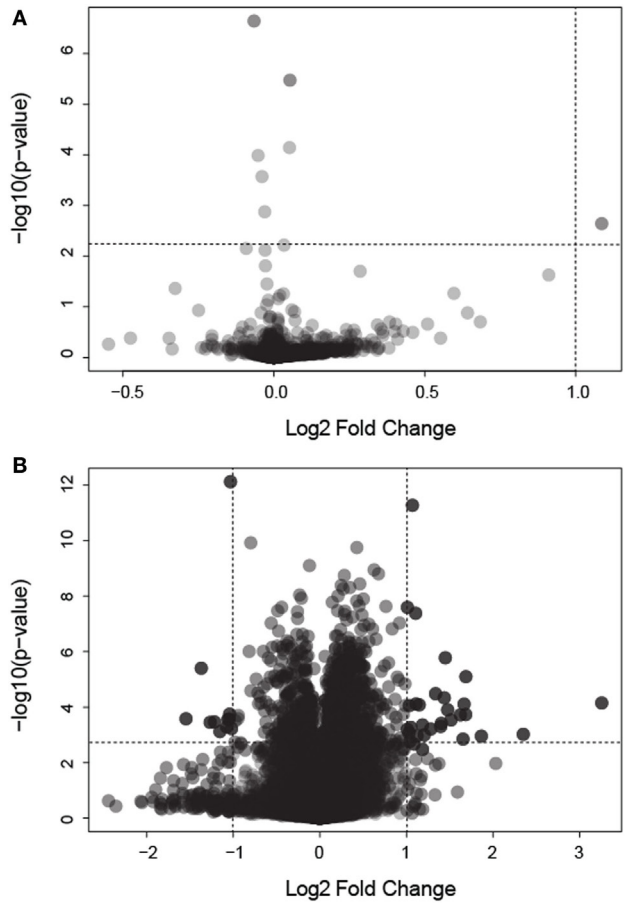


FIGURE 1 | Volcano plot depicting differentially expressed genes (DEGs) in mosquitoes submitted to different feeding regimens. **(A)** Comparison of blood-fed (BF) vs. uninfected mosquitoes. **(B)** Comparison of BF vs. non-BF mosquitoes. All data were log2 transformed. Dotted lines represent the significant *p*-value and fold change cutoffs used to identify DEGs (see Materials and Methods).

identification of 42 DEGs (32 upregulated and 10 downregulated genes) (**Figure 1B**).

The Global Perturbation of Virus Infection Is Influenced by Feeding Status

The DEGs identified in BF vs. N-BF mosquitoes is indicative that the feeding process modulates gene expression. Next, we evaluated DEGs during infection caused by DENV, YFV, and WNV compared to BF and N-BF mosquitoes (**Figure 2**). In the first part, we evaluated whether infection could drive differences in gene expression. For this, we compared infected mosquitoes separately vs. BF and 725 genes were found to be downregulated; YFV ($n = 418$ genes), WNV ($n = 26$), and DENV ($n = 281$). The intersection between all infection conditions showed a greater proportion of downregulated genes ($n = 50$), compared to upregulated ($n = 43$) (**Figure 1A**). In the comparison of infected (DENV, YFV, and WNV) vs. N-BF mosquitoes, we observed a different pattern: more upregulated genes ($n = 626$); YFV ($n = 263$), WNV ($n = 29$), and DNV ($n = 334$) in comparison with downregulated. The intersection between all conditions shows a higher number of upregulated genes ($n = 131$) compared to repressed genes ($n = 54$) (**Figure 2B**). The most altered genes were *AAEL014672* and *AAEL000870* in YFV-infected samples, *AAEL000611* and *AAEL011460* in DENV-infected samples and *AAEL011669*, *AAEL000611*, *AAEL003012* in WNV-infected samples (**Figures 2A,B**).

DEGs Obtained from Infected Mosquitoes Can Discriminate Virus Infection in N-BF and BF Mosquitoes

After we identified the global differences between the expression profiles in infected groups (DENV, YFV, and WNV) in comparison with feeding status (BF and N-BF), we performed PCA to verify whether these significant DEGs could discriminate infected, uninfected and N-BF mosquitoes (**Figure 3**). First, we used the 42 DEGs observed in the comparison of BF vs. N-BF (**Figure 1B**); however, these genes did not allow differentiation between BF, N-BF and infected mosquitoes (**Figure 3A**). Next, the 1,328 DEGs from the comparison of mosquitoes infected with DENV, YFV or WNV vs. N-BF allowed separation of infected and N-BF mosquitoes (**Figure 3B**). Finally, the 1,410 DEGs from the comparison of infected vs. BF mosquitoes could discriminate BF from infected mosquitoes with less variance, in respect to the DEGs from the infected vs. N-BF comparison (**Figure 3C**).

Modules of Coexpressed Genes Related to Infection in BF and N-BF Compose Infection-Specific Expression Patterns Independent of the Dietary Background

We constructed two weighted gene coexpression networks with WGCNA using as input the DEGs identified in the comparisons of infected (WNV, YFV, or DENV) vs. BF (1,410 genes) and infected vs. N-BF (1,328 genes) mosquitoes (**Figures 2A,B**). MEs, which represent the first principal component of each module, effectively summarizing its expression, were correlated with traits of interest (BF, N-BF, all infected mosquitoes, and infected by each individual virus). The first network was constructed using the DEGs from the comparison of infected vs. BF mosquitoes, resulting in five coexpression modules (color-labeled *black, green, purple, blue*, and *red*) (**Figure 4A**). The *black* ME showed a strong, positive correlation with viral infection ($r = 0.94$; p-value < 0.0001) as well as with infection by each individual virus, as expected (**Figure 4A**). The second coexpression network was constructed from the DEGs arising from the comparison of infected vs. N-BF mosquitoes, allowing the identification of two modules (labeled *turquoise* and *brown*). Module *turquoise* eigengene also had a strong, positive correlation with viral infection ($r = 0.97$; p-value < 0.0001) (**Figure 4B**).

Both *black* and *turquoise* were the most significant modules correlated with viral infection in each comparison. Next, we analyzed the 329 genes present in *black* and 201 genes grouped into the *turquoise* modules, identifying 110 genes in common (**Figure 5A**) (available in Table S1 in Supplementary Material).

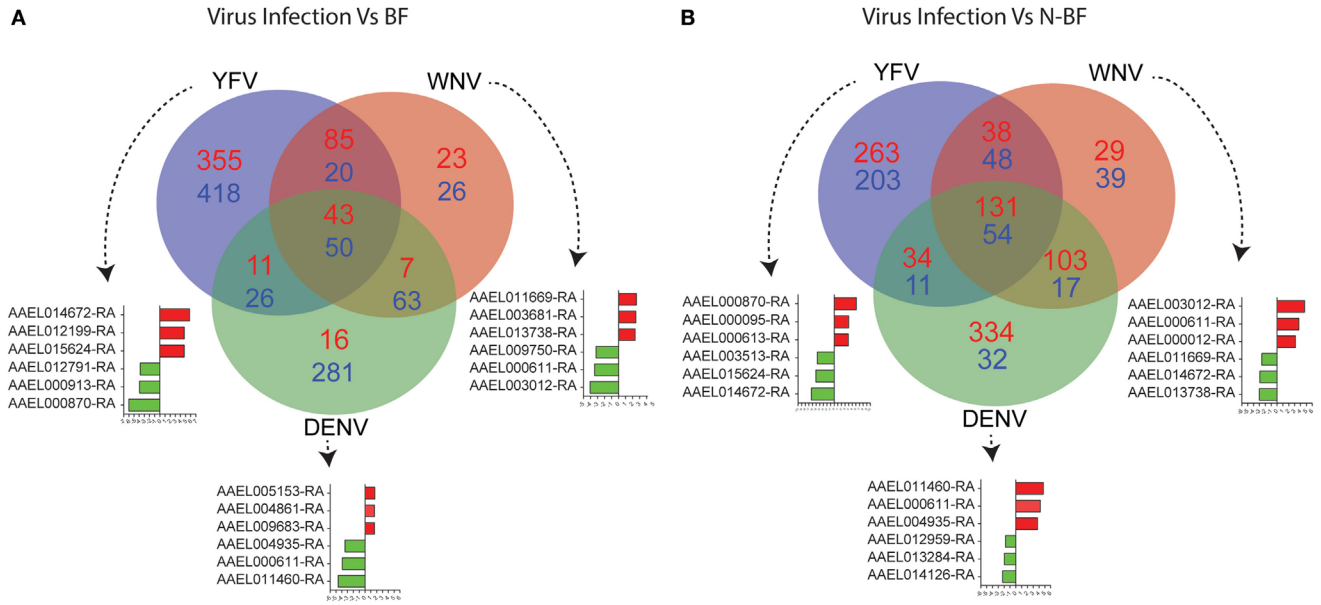


FIGURE 2 | Critical signature differences in the distinct feeding schedules. **(A)** Differentially expressed genes (DEGs) from the comparison of genes from virus infected mosquitoes [yellow fever virus (YFV), dengue virus (DENV), and West Nile virus (WNV) vs. blood-fed (BF)]. **(B)** DEGs from the comparison of genes from virus infected mosquitoes (YFV, DENV, and WNV vs. non-BF). The blue numbers represent downregulated genes and the red numbers represent upregulated genes by absolute $\log_2$-transformed fold-change ≥1 criterium. Red and green bars correspond to up- and downregulated genes, respectively.

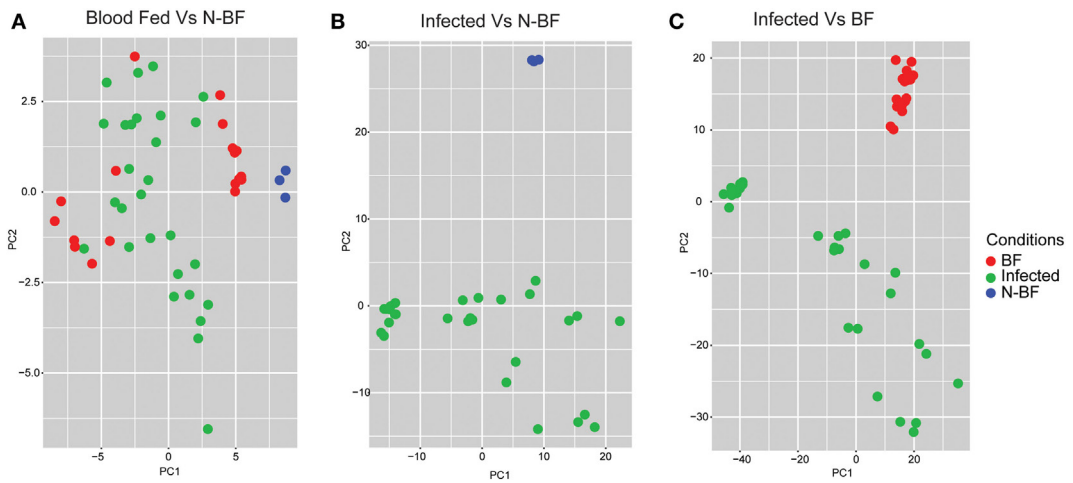


FIGURE 3 | Principal component analysis of differentially expressed genes (DEGs). **(A)** DEGs from the comparison between blood-fed (BF) vs. nonblood-fed (N-BF) did not separate the BF (red), N-BF (blue), and infected (green) groups. **(B)** DEGs from the comparison of the infected (WNV, YFV, and DENV) vs. N-BF separates the N-BF (red) from the infected (blue) groups. **(C)** DEGs from the comparison of the infected (WNV, YFV, and DENV) vs. BF mosquitoes separates the BF groups (blue) and infected (red).

The 110 genes in the intersection of both comparisons were regarded as forming a viral pattern of infection independent of the dietary background. To verify this, we performed hierarchical clustering analysis followed by heatmap expression visualization of these 110 genes (**Figure 5**). This unsupervised analysis resulted in the forming of two main groups separating the infected from the uninfected samples, suggesting an important role of these genes.

Functional Analysis of Genes in the Coexpression Modules Reveals Distinct Processes Modulated by Infection and Feeding

We performed functional annotation of the 110 identified genes using biological processes GO terms extracted from the Vector Base repository. The main functional role of each module was

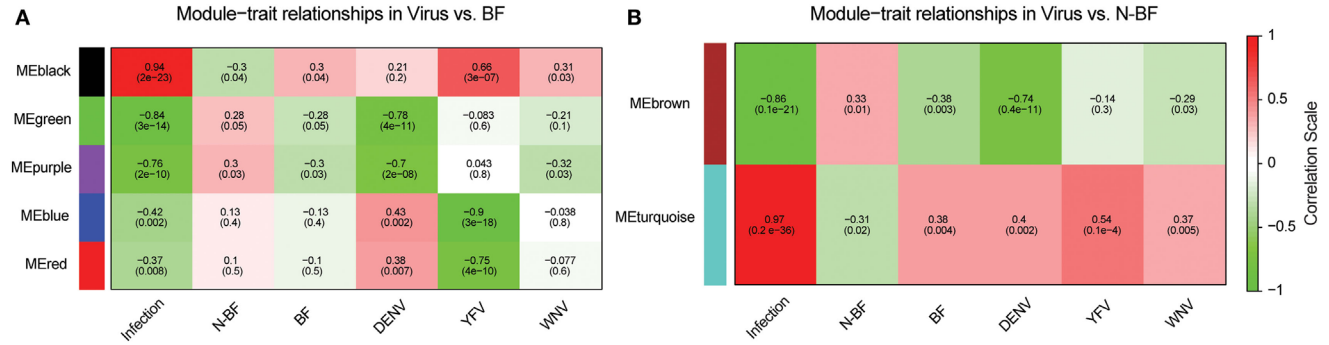


FIGURE 4 | Identification of coexpression modules in infected vs. blood-fed (BF) and infected vs. nonblood-fed (N-BF) comparison. Pearson's correlation of module eigengenes with traits (infection by each virus, BF, N-BF) are presented for each module (labeled *black*, *green*, *purple*, *blue, red, brown*, and *turquoise*) and corresponding *p*-value of the correlation in cells colored by the strength of the correlation. **(A)** Comparison of differentially expressed genes (DEGs) of infected vs. BF mosquitoes. **(B)** DEGs of infected vs. N-BF mosquitoes.

summarized considering the top frequent GO term for the most variable (up- or downregulated) genes. Upregulated genes present in *black* module are mainly related to embryo development, sensory perception of chemical stimulus and conjugation with cellular fusion, while genes related to transport, cytoskeleton organization and protein localization were found downregulated. Upregulated genes in the *turquoise* module are related to "protein localization," "growth," and "lipid metabolic process," while "regulation of transcription, DNA template," "cell morphogenesis," and "cell cycle processes" genes appeared downregulated. Finally, genes in the intersection were observed to play a role in the activation of functions related to "cell differentiation," "transport," and "signal transduction," while "proteolysis," "oxidative-reduction process," and "lipid metabolic processes" were downregulated (**Figure 6**).

Machine Learning Analysis Reduces the Infection-Specific Expression Pattern to a Small Set of Genes

We applied data mining techniques to further elucidate the importance of the genes identified in the previous analyses (**Figures 5** and **6**) and investigate potentially hidden connections within the gene expression datasets. A decision tree based in the J48 algorithm was constructed to identify a minimal set of genes that could explain the infection status. Four genes were able to separate infected from uninfected mosquitoes (*AAEL012128*, *AAEL014210*, *AAEL002477*, and *AAEL005350*) (**Figure 7**). The first gene, *AAEL012128,* classifies the mosquitoes into groups of infected (characterized by decreased expression of *AAEL012128*) and uninfected (characterized by increased expression of the gene) (**Figure 7A**). The remainder three genes allow further stratification of the groups according to their infection by each of the three studied viruses. The true positive rate using these genes in this classifier summed 85.71% with 7 (14.28%) incorrectly classified instances out of a total of 49, with kappa statistic (a measure of classification accuracy) of 0.8. The confusion matrix shows the number of classification errors (**Figure 7B**). The area under the ROC in the classification of mosquitoes infected by YFV, WNV, and DENV was, respectively, of 0.96, 0.87, and 0.83 (**Figure 7B**). On the other hand, area under the ROC for the classification of BF and N-BF mosquitoes was, respectively, of 0.94 and 0.82 (**Figure 7B**). These results pinpoint that probing the expression of a reduced number of genes in the *A. aegypti* mosquito allows the identification of its infected status.

Independent Validation of the Role of *AAEL012128* during Infection

In order to independently validate the role of *AAEL012128*, gene identified as summarizing the infection status in our combined study, we identified two datasets related to DENV infection and verified the expression of this gene. The first dataset, reported by Sim and Dimopoulos, (2010), compared the expression of DENV-infected (live or heat-inactivated) and naïve *A. aegypti* cells (Aag2 cell line). These results are presented in Figure S1A in Supplementary Material. In line with our findings, expression of this gene was decreased in the cells exposed to the live pathogen, while samples from heat-inactivated virus had increased expression of *AAEL012128*. The second dataset, reported by Behura et al. (2011), consists of a time-course experiment comparing four DENV infected samples at 3 h and 18 h post infection (p.i.) and a control sample of RNA isolated following an uninfected blood meal. Figure S1B in Supplementary Material presents the expression of *AAEL012128* at both time-points, and a significant decrease of expression is observed at 18 h p.i. compared to the control (Mann–Whitney *U*-test; p-value < 0.05), as well as between the 3 and 18 h p.i. samples (Mann–Whitney *U*-test; p-value < 0.05), but not between 3 h p.i. and the control samples. These results show that virus presence downregulate the *AAEL012128* gene. In the other hand, to exclude the alimentation background we assessed two other datasets reported by Bonizzoni et al. (2011) and Bonizzoni et al. (2012). The first dataset used RNA-seq analyses of BF and sugar-fed mosquitoes (LVP strain) to investigate the differential gene expression in *A. aegypti* females. The second dataset used three *A. aegypti* strains, Chetumal (CTM), Rexville D-Puerto Rico, and LVP and also compared gene expression in these three strains between sugar-fed and BF alimentation regimens. The gene

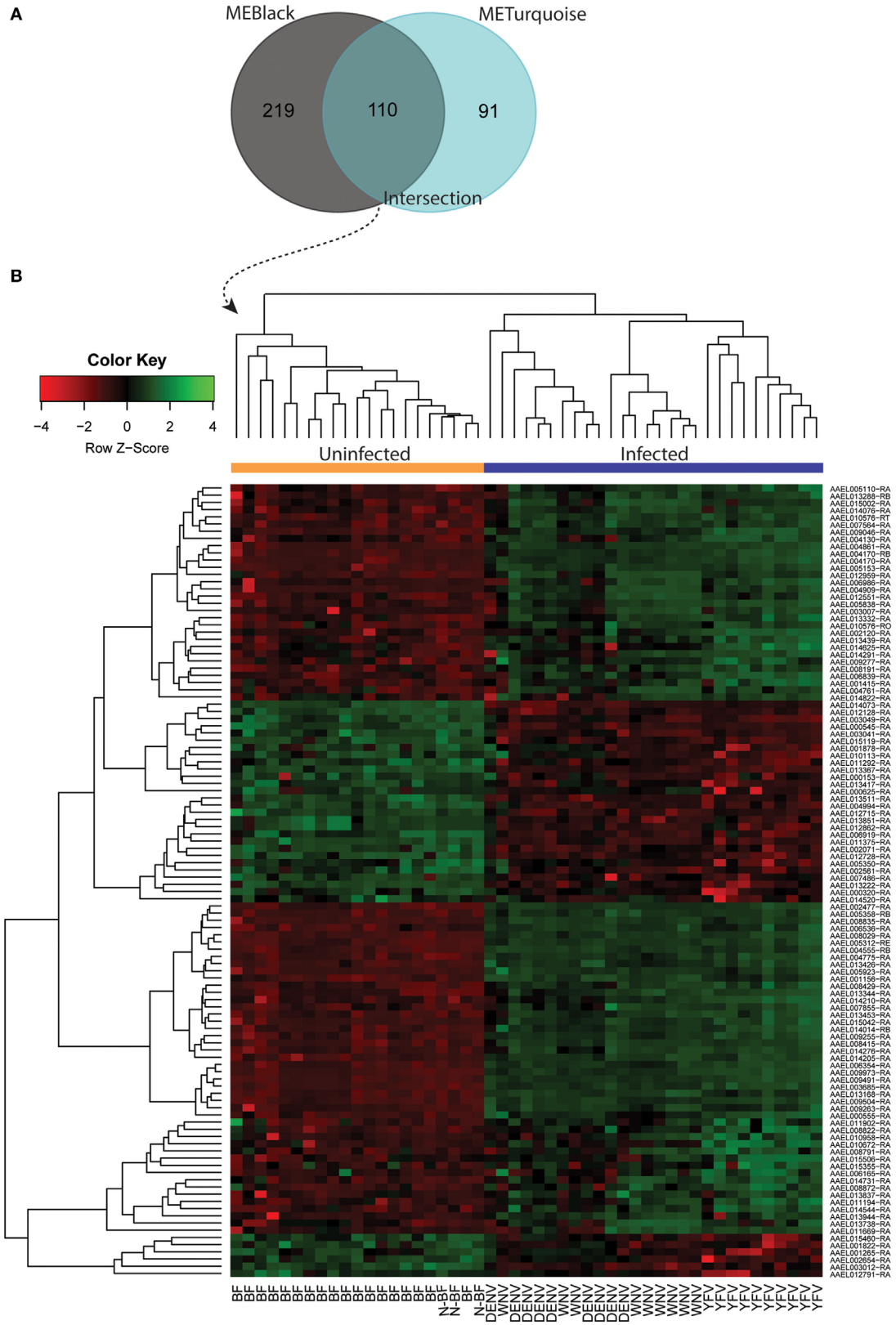


FIGURE 5 | Infection-specific genes identification by removing the blood-feeding noise. **(A)** Venn diagram with genes present within the *black* and *turquoise* modules. **(B)** Hierarchical clustering and heatmap of 110 genes present in the intersection.

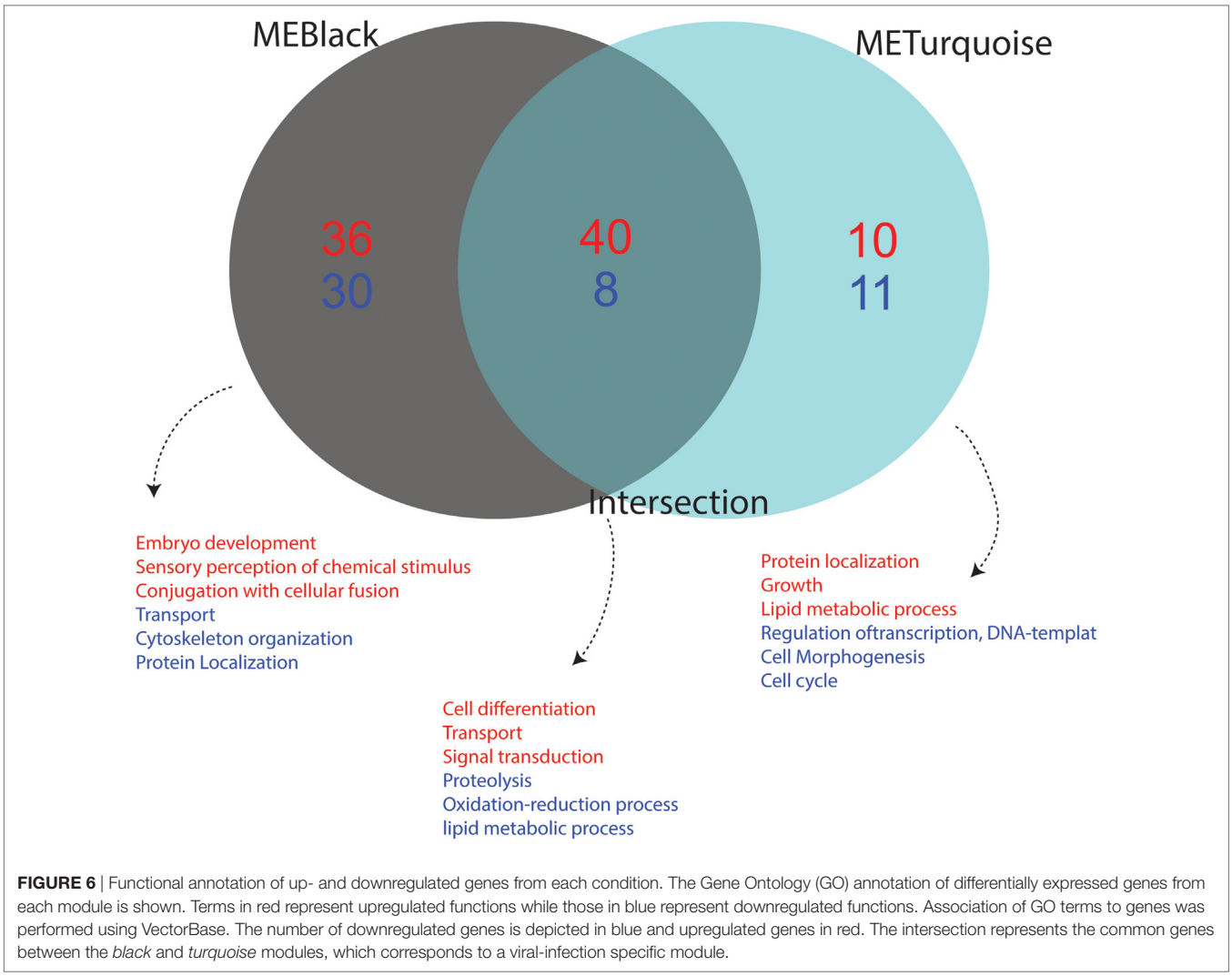


FIGURE 6 | Functional annotation of up- and downregulated genes from each condition. The Gene Ontology (GO) annotation of differentially expressed genes from each module is shown. Terms in red represent upregulated functions while those in blue represent downregulated functions. Association of GO terms to genes was performed using VectorBase. The number of downregulated genes is depicted in blue and upregulated genes in red. The intersection represents the common genes between the *black* and *turquoise* modules, which corresponds to a viral-infection specific module.

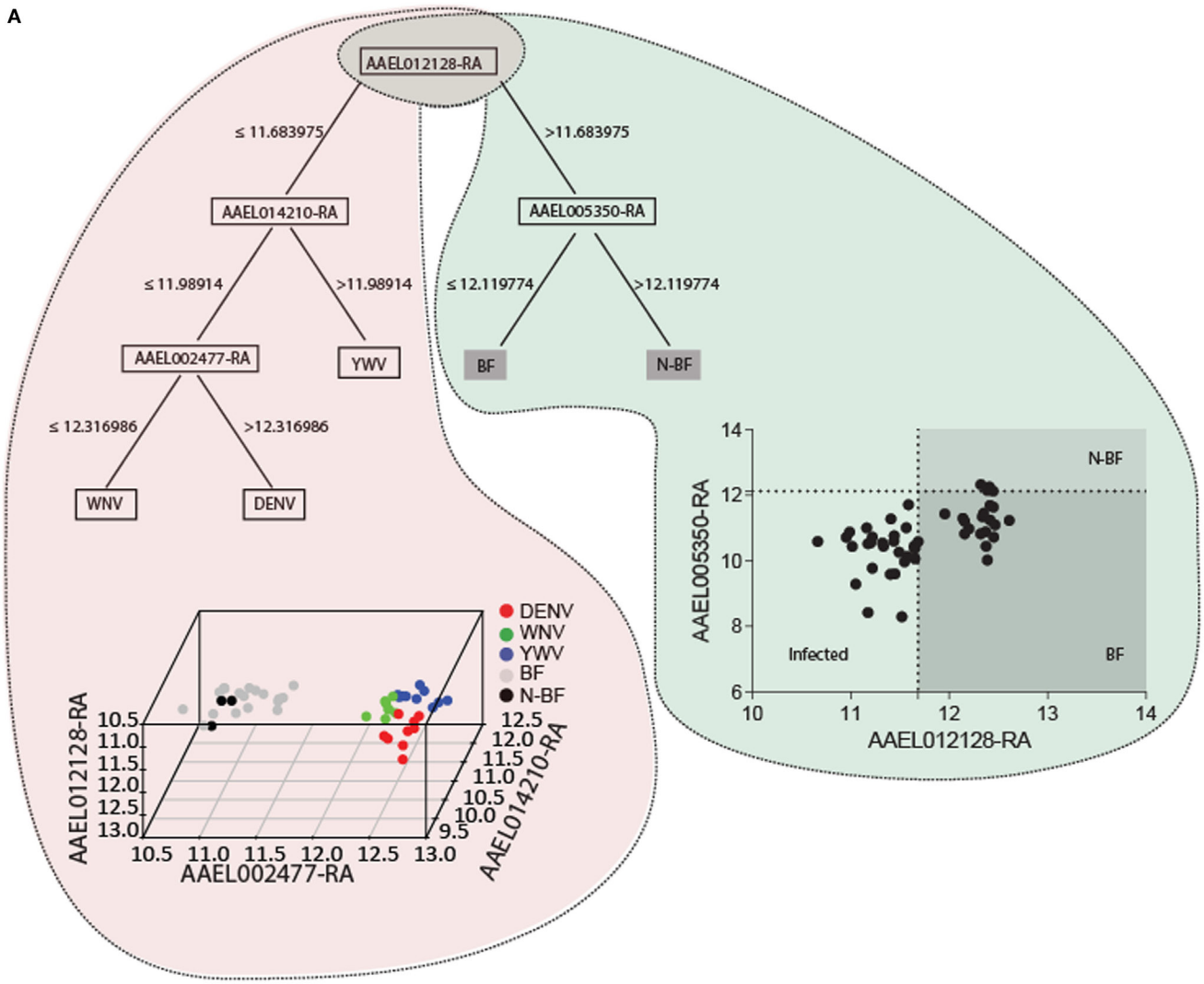
expression behavior of *AAEL012128* was assessed in both studies, and in line with our results this gene did not appear differentially expressed in most samples, except in the CTM strains from the Bonizzoni et al. (2012) dataset where a slight (log2FC = −0.7), but statistically significant expression decrease in BF mosquitoes was observed.

DISCUSSION

The establishment that the infection in a vector is influenced by multiple factors, including alimentary behavior, seasonal effects and pathogen cooccurrence is recent (Ricklefs et al., 2016). Transcriptomic analyses provide the information to better address the intermediate steps between genes and their biological roles (Wang et al., 2009). Given the myriad of diseases that can be transmitted by these vectors, identification of viral-infection markers is crucial to propose new surveillance and control strategies. Well-known pathways related to viral infection have been identified in many vectors including *A. aegypti* and *Culex quinquefasciatus* and involves activation of Toll, Imd, JAK-STAT, and RNAi pathways, which serve as defense system for controlling the infection (Gruber et al., 2008; Souza-Neto et al., 2009; Kerpedjiev et al., 2015).

However, this identification is not completely unbiased, since the infection process is closely associated with feeding behavior and consequent simultaneous modulation of genes related to both processes. In this context, we performed a combined analysis of expression data from studies that compared infected (by DENV, WNV, or YFV) and mock-infected mosquitoes (Colpitts et al., 2011) and BF against N-BF mosquitoes (Dissanayake et al., 2010). The present analysis comprises a finer-level understanding of transcriptional mechanisms associated with infection independent of the transcriptional effect due to blood feeding.

First, we investigated the similarity between the uninfected mosquitoes from the Colpitts et al. study (GSE28208), which had access to raisins as dietary source, and the BF mosquitoes from the Dissanayake et al. report (GSE22339), with both diets driving a similar expression profile (**Figure 1A**). However, when different diet schemes are used, we showed the existence of a high number of DEGs in the comparisons of infected vs. BF or



B

Confusion Matrix

	DENV	YFV	WNV	BF	N-BF
DENV	6	2	1	0	0
YFV	0	10	0	0	0
WNV	0	1	7	1	0
BF	0	0	0	17	1
N-BF	0	0	0	1	2

FIGURE 7 | The key genes for trait classification resulting from data mining of differentially expressed genes in the modules intersection. **(A)** The J48 algorithm was used to build a decision tree using default parameters summarizes 110 genes of the intersection into four informative genes. The scatter-plot depicts by a cutoff the samples distribution discriminating each group. In the 3D plot samples are discriminated by colors according to the expression of the four genes. **(B)** The confusion matrix of the classification of the four genes estimates the classification accuracy of the decision tree. A 10-fold cross validation was performed.

N-BF mosquitoes (**Figure 2**) suggestive of influences due to an alimentation noise. PCA was applied to the DEGs arising from these comparisons, allowing the separation of infected, BF and N-BF samples, albeit with smaller variance in the comparison with the latter (**Figure 3**).

Next, we constructed gene expression correlation networks using DEGs from these comparisons as input. These networks have widespread use in the detection of genes that take part in common biological processes or that are regulated by an overlapping set of transcriptional factors (Kogelman et al., 2014). By constructing two separate coexpression networks (one with DEGs from infected vs. BF and another with DEGs from the infected vs. N-BF comparison) and calculating the gene content intersection in modules that related to infection (*black* and *turquoise* modules)

we were able to single out 110 genes which behave infection-specific without noise interference (**Figure 5A**). The hierarchical clustering of these genes allowed unambiguous classification of infected and uninfected samples (**Figure 5B**). Interestingly, we identified in this set infection-specific expression patterns (either activation or repression) of genes playing important roles in immunity, stress and chemosensory reception. For instance, two members of the cytochrome P450 family which have been previously related to infection (Bartholomay et al., 2010; Skalsky and Cullen, 2010; Colpitts et al., 2011; Pan et al., 2012) were found among these 110 genes that discriminate infected from uninfected mosquitoes, although they were found less expressed in the infected group (*AAEL000320* $\log_2$FC = −1.33 and *AAEL002071* $\log_2$FC = −1.50, respectively). Also, genes related to immunity were identified such as *AAEL014544,* which codes for a prophenoloxidase, an insect type-3 copper enzyme involved in melanization against invading pathogens and blockade of infection (Chen et al., 2012). This gene appeared more activated in the virus-infected samples ($\log_2$FC = 1.72).

Also immune related was a fibrinogen-like sequence (*AAEL013417*) that was less expressed in infected samples ($\log_2$FC = −1.43) but which may provide a level of pathogen recognition given the lack of antibody-mediated immunity in these organisms (Dong and Dimopoulos, 2009), as well as the AMP cecropin (coded in *AAEL000625*). These peptides have been proposed as exerting both antibacterial and antiviral activities (Luplertlop et al., 2011), although expression of this gene was lower in the infected group ($\log_2$FC = −1.26). On the other hand, an odorant receptor coded in the gene *AAEL015506* presented the second largest expression variation relative to the uninfected group ($\log_2$FC = 1.95), although this change was more pronounced in the YFV-infected group. The finding that viral infection also modulates behavioral changes in the mosquito thus affecting vectorial capacity was previously reported in the context of DENV infection (Sim et al., 2012).

As a secondary result of our coexpression-based approach, we provide evidence that genes with yet unknown roles (hypothetical proteins) may actually play important parts on the infection process given their infection-specific activation patterns. One such sequence is coded in the *AAEL011669* gene, which does not have an associated function but analysis of its predicted peptide sequence reveals the presence of kinase domains (Pfam accession no. PF00069) that may be associated with the regulation of important cellular programs. Expression of this gene appeared most increased in the infected group ($\log_2$FC = 2.59). Another case is *AAEL013738*, also annotated as hypothetical protein. Expression of this gene was also higher in the infected group ($\log_2$FC = 1.82). Thus, the biological roles of these yet uncharacterized genes during infection in the *Aedes* mosquito requires further studies.

The association of biological process GO terms to the *black* coexpression module resulting from the comparison of infected versus uninfected, BF mosquitoes allowed the implication of pathways that may relate to infection and development. For instance, "embryo development" and "chemosensory perception" were both activated in the analyzed dataset. Indeed, oviposition in clean water sources starts immediately after female-mosquito feeding, and is preceded by egg maturation and site seeking (Davis et al., 2016). On the other hand, we identified decreased expression of genes that grouped to "transport," "cytoskeleton organization" and "protein localization" terms, suggestive of host cellular components reorganization during feeding. The *turquoise* module resulting from the comparison of infected vs. N-BF grouped activated functional terms related to "protein localization," "growth," and "lipid metabolic processes." These processes were not unexpected, considering the possibility of use of alternative energy sources such as fatty acids *via* β-oxidation (Arrese and Soulages, 2010).

Decreased expression of genes related to "cell cycle," "cell morphogenesis," and "transcriptional regulation" was observed, indicating that metabolic arrest may occur in these conditions. Finally, the intersection of both gene sets in the *black* and *turquoise* coexpression modules form the *core* infection-specific genes. Activated functional terms of these genes included "cell differentiation," "transport," and "signal transduction." This reinforces what is known about the infection process that involves viral uptake and the use of host protein machinery for self-replication and viral particle assembly (Mosso et al., 2008).

In order to further reduce the set of genes that explain virus infection, we applied a machine learning technique based in decision trees using as input the 110 virus-specific expressed genes. The proposed model includes four genes (**Figure 7**) that classify infected from healthy mosquitoes, with *AAEL012128* being the most informative. In *A. aegypti* this gene is predicted to code for a 12-pass transmembrane protein with cationic amino acid transporter function. Expression of this gene showed a one-fold decrease in infected mosquitoes ($\log_2$FC = −1.03). The finding that amino acid transporters may be targeted by viral particles, as a receptor, has been reported in other insects including the silkworm *Bombyx mori*. In this organism the deletion of a gene coding for an analogous amino acid transporter generates resistance against densovirus type 2 infection, a parvo-like virus (Ito et al., 2008). Similarly, mammalian amino acid transporters spanning 12–14 transmembranes were previously reported as retroviruses receptor (Wang et al., 1991). Although a number of *A. aegypti* amino acid transporters have been experimentally characterized to date (Umesh et al., 2003; Evans et al., 2009; Hansen et al., 2011; Boudko et al., 2015) their possible functioning as viral receptors was not evaluated, and also lacking is the study of *AAEL012128* in this context.

Given the gathered evidence, we put forward the hypothesis that this gene may play an important role in the context of virus infection in the *A. aegypti* mosquito, possibly acting as a receptor. We evaluated the expression of this gene in independent datasets (not used during our Discovery analyses) related to DENV infection, and the expression trends of *AAEL012128* are in line with our findings, being decreased in infected samples (Figure S1 in Supplementary Material) or remaining unaltered when no pathogen exposure is performed. Interestingly, in one of the evaluated datasets, a time-course experiment comparing infection at 3 h and 18 h, decreased expression of *AAEL012128* can only be perceived at 18 h. Considering that one round of DENV replication occurs at approximately 30 h (Helt and Harris, 2005), this indicates that the change in expression of this cationic transporter occurs still during the initial establishment of the infection. Additionally, this gene was not differentially expressed in two independent datasets

related to feeding schedule. This confirms the role of *AAEL012128* gene as an infection mark independent of blood feeding, at least in the probed mosquito strains.

The other three genes identified through data mining, *AAEL014210*, *AAEL002477*, and *AAEL005350* correspond to uncharacterized proteins. The first two may play regulatory roles due to the prediction of a zinc finger, DNA-binding domain (InterPro accession no. IPR013087) in *AAEL014210* and of a basic-leucine zipper domain (InterPro accession no. IPR004827) in *AAEL002477*, while *AAEL005350* harbors retinaldehyde-binding and alpha-tocopherol transport domains (InterPro accession nos. IPR001071, IPR001251).

Our study has some limitations: although the total number of samples analyzed was high ($n = 58$), regarding the infection condition we only included an equal number of mosquitoes samples infected with DENV, WNV, and YFV from the Colpitts et al. (2011) dataset. This strictly limits the generalization of our results to genes related to infection by these viruses. While there exists others expression sets related to *A. aegypti* infection, it is not surprising that most of them focus on DENV infection, such as the works of Behura et al. (2011) and Sim and Dimopoulos (2010). This occurs given the DENV relevance in most tropical and subtropical areas worldwide. These samples were not included in the first analysis in order to avoid overrepresentation of a DENV-specific transcriptional response and were used in external validation. This overrepresentation could lead to biases in our gene expression correlation-based approach.

Other publicly available expression datasets related to the *A. aegypti* mosquito address more specific questions such as insecticide resistance (Kasai et al., 2014), sex differences (GEO accession no. GSE7813), developmental aspects (GEO accession nos. GSE23039, GSE7811, GSE71221, and GSE90515), and circadian mechanisms (Ptitsyn et al., 2011; Leming et al., 2014; Bottino-Rojas et al., 2015; Jupatanakul et al., 2017). For this reason these studies were not considered in our analyses. Furthermore, our dataset contains only the LVP and Rockefeller mosquito strains in the discovery dataset. During the validation step, we were able to confirm our findings in a DENV infection study that used Moyo mosquitoes as well as in LVP and D-Puerto Rico mosquitoes strains from an alimentation dataset, but not in CTM samples from this same dataset. In an Aag2 cell line infected with DENV the expression response of *AAEL012128* was also in line with our findings. Thus, there exist strain-specific differences forming part of the *A. aegypti* response to infection and to feeding in blood that limits extrapolation of gene expression findings.

CONCLUSION

In this study, we performed an integrated analysis of *A. aegypti* expression datasets totaling 58 samples and validate in four different datasets. We aimed the identification of virus infection-specific gene sets independent of feeding behavior. Using a correlation-based analysis, we determined a set of 110 genes that are specific of the vector response to the viral infection. Further reduction of this dataset allowed the identification of four genes with high information gain on discriminating infected mosquitoes, and these were validated using independent datasets. Our derived, integrated dataset of *A. aegypti* transcripts could orientate experimental confirmation of the role of the identified genes during viral infection. Increased knowledge on the transcriptomic aspects specific to the infected mosquito could be a means to the design of novel vector control strategies and better understanding of vector biology during infection.

AVAILABILITY OF DATA AND MATERIALS

All data analyzed during this study are already publicly available or included in this article as Additional files. The original data are available in the Gene Expression Omnibus[3] under accession nos. GSE28208 and GSE22339 and associated manuscripts. Data used during validation are also deposited in GEO under accession nos. GSE16563, GSE33274, or within the supporting files of the respective manuscripts (http://www.g3journal.org/content/2/1/103.supplemental; and https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3042412/bin/1471-2164-12-82-S2.XLSX). The derived results obtained during our analyses (intersection of coexpression modules) supporting the conclusions of this manuscript are included within the text and in the associated supplementary material.

AUTHOR CONTRIBUTIONS

AQ, PR, and KF conceived of the study. KF, JK, PR, AP, and AQ performed data analysis. AQ, KF, PR, AB, CO, and MG drafted the manuscript with input from the other authors. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Dr. Bruno Bezerril Andrade and Dr. Cintia Figueiredo de Araujo for their comments providing insights that greatly improved the manuscript, the assistance of Mr. Andris Walter in english proofing, Dr. Eurico Arruda, Dr. Luciano Kalabric, Dr. Benedito Fonseca, the technicians of virology research center of FMRP-USP, Dr. Cleyson Barros and Dr. João Santana da Silva from Universidade de São Paulo for their scientific assistance.

FUNDING

AQ acknowledges financial support from Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB process no. JCB0004/2013). KF was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP no. 2017/03491-6). AP acknowledges financial support from CNPq—Grant Edital Universal MCTI/CNPQ/Universal14/2014 (Process No.: 454505/2014-0).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/articles/10.3389/fbioe.2017.00084/full#supplementary-material.

[3] http://www.ncbi.nlm.nih.gov/geo.

REFERENCES

Arrese, E. L., and Soulages, J. L. (2010). Insect fat body: energy, metabolism, and regulation. *Annu. Rev. Entomol.* 55, 207–225. doi:10.1146/annurev-ento-112408-085356

Bae, H.-G., Drosten, C., Emmerich, P., Colebunders, R., Hantson, P., Pest, S., et al. (2005). Analysis of two imported cases of yellow fever infection from Ivory Coast and the Gambia to Germany and Belgium. *J. Clin. Virol.* 33, 274–280. doi:10.1016/j.jcv.2004.12.001

Bartholomay, L. C., Waterhouse, R. M., Mayhew, G. F., Campbell, C. L., Michel, K., Zou, Z., et al. (2010). Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* 330, 88–90. doi:10.1126/science.1193162

Behura, S. K., Gomez-Machorro, C., Harker, B. W., deBruyn, B., Lovin, D. D., Hemme, R. R., et al. (2011). Global cross-talk of genes of the mosquito *Aedes aegypti* in response to dengue virus infection. *PLoS Negl. Trop. Dis.* 5:e1385. doi:10.1371/journal.pntd.0001385

Bonizzoni, M., Dunn, W. A., Campbell, C. L., Olson, K. E., Dimon, M. T., Marinotti, O., et al. (2011). RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti*. *BMC Genomics* 12:82. doi:10.1186/1471-2164-12-82

Bonizzoni, M., Dunn, W. A., Campbell, C. L., Olson, K. E., Marinotti, O., and James, A. A. (2012). Strain variation in the transcriptome of the dengue fever vector, *Aedes aegypti*. *G3 Genes Genomes Genet.* 2. doi:10.1534/g3.111.001107

Bottino-Rojas, V., Talyuli, O. A. C., Jupatanakul, N., Sim, S., Dimopoulos, G., Venancio, T. M., et al. (2015). Heme signaling impacts global gene expression, immunity and dengue virus infectivity in *Aedes aegypti*. *PLoS ONE* 10:e0135985. doi:10.1371/journal.pone.0135985

Boudko, D. Y., Tsujimoto, H., Rodriguez, S. D., Meleshkevitch, E. A., Price, D. P., Drake, L. L., et al. (2015). Substrate specificity and transport mechanism of amino-acid transceptor Slimfast from *Aedes aegypti*. *Nat. Commun.* 6, 8546. doi:10.1038/ncomms9546

Cauchemez, S., Ledrans, M., Poletto, C., Quenel, P., de Valk, H., Colizza, V., et al. (2014). Local and regional spread of chikungunya fever in the Americas. *Euro Surveill.* 19, 20854. doi:10.2807/1560-7917.ES2014.19.28.20854

Chen, Y., Liu, F., Yang, B., Lu, A., Wang, S., Wang, J., et al. (2012). Specific amino acids affecting *Drosophila melanogaster* prophenoloxidase activity in vitro. *Dev. Comp. Immunol.* 38, 88–97. doi:10.1016/j.dci.2012.04.007

Colpitts, T. M., Cox, J., Vanlandingham, D. L., Feitosa, F. M., Cheng, G., Kurscheid, S., et al. (2011). Alterations in the *Aedes aegypti* transcriptome during infection with West Nile, dengue and yellow fever viruses. *PLoS Pathog.* 7:e1002189. doi:10.1371/journal.ppat.1002189

Davis, T. J., Kline, D. L., and Kaufman, P. E. (2016). *Aedes albopictus* (Diptera: Culicidae) oviposition preference as influenced by container size and *Buddleja davidii* plants. *J. Med. Entomol.* 53, 273–278. doi:10.1093/jme/tjv201

Dissanayake, S. N., Ribeiro, J. M., Wang, M.-H., Dunn, W. A., Yan, G., James, A. A., et al. (2010). aeGEPUCI: a database of gene expression in the dengue vector mosquito, *Aedes aegypti*. *BMC Res. Notes* 3:248. doi:10.1186/1756-0500-3-248

Dong, Y., and Dimopoulos, G. (2009). *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *J. Biol. Chem.* 284, 9835–9844. doi:10.1074/jbc.M807084200

Espíndola, M. S., Lima, L. J. G., Soares, L. S., Cacemiro, M. C., Zambuzi, F. A., de Souza Gomes, M., et al. (2015). Dysregulated immune activation in second-line HAART HIV+ patients is similar to that of untreated patients. *PLoS ONE* 10:e0145261. doi:10.1371/journal.pone.0145261

Evans, A. M., Aimanova, K. G., and Gill, S. S. (2009). Characterization of a blood-meal-responsive proton-dependent amino acid transporter in the disease vector, *Aedes aegypti*. *J. Exp. Biol.* 212, 3263–3271. doi:10.1242/jeb.029553

Giraldo-Calderon, G. I., Emrich, S. J., MacCallum, R. M., Maslen, G., Dialynas, E., Topalis, P., et al. (2015). VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* 43, D707–D713. doi:10.1093/nar/gku1117

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* 36, W70–W74. doi:10.1093/nar/gkn188

Hansen, I. A., Boudko, D. Y., Shiao, S.-H., Voronov, D. A., Meleshkevitch, E. A., Drake, L. L., et al. (2011). AaCAT1 of the yellow fever mosquito, *Aedes aegypti*. *J. Biol. Chem.* 286, 10803–10813. doi:10.1074/jbc.M110.179739

Helt, A.-M., and Harris, E. (2005). S-phase-dependent enhancement of dengue virus 2 replication in mosquito cells, but not in human cells. *J. Virol.* 79, 13218–13230. doi:10.1128/JVI.79.21.13218-13230.2005

Ito, K., Kidokoro, K., Sezutsu, H., Nohata, J., Yamamoto, K., Kobayashi, I., et al. (2008). Deletion of a gene encoding an amino acid transporter in the midgut membrane causes resistance to a *Bombyx* parvo-like virus. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7523–7527. doi:10.1073/pnas.0711841105

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi:10.1093/biostatistics/kxj037

Jupatanakul, N., Sim, S., Angleró-Rodríguez, Y. I., Souza-Neto, J., Das, S., Poti, K. E., et al. (2017). Engineered *Aedes aegypti* JAK/STAT pathway-mediated immunity to dengue virus. *PLoS Negl. Trop. Dis.* 11:e0005187. doi:10.1371/journal.pntd.0005187

Kasai, S., Komagata, O., Itokawa, K., Shono, T., Ng, L. C., Kobayashi, M., et al. (2014). Mechanisms of pyrethroid resistance in the dengue mosquito vector, *Aedes aegypti*: target site insensitivity, penetration, and metabolism. *PLoS Negl. Trop. Dis.* 8:e2948. doi:10.1371/journal.pntd.0002948

Kerpedjiev, P., Hammer, S., and Hofacker, I. L. (2015). Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics* 31, 3377–3379. doi:10.1093/bioinformatics/btv372

Kogelman, L. J. A., Cirera, S., Zhernakova, D. V., Fredholm, M., Franke, L., and Kadarmideen, H. N. (2014). Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Med. Genomics* 7:57. doi:10.1186/1755-8794-7-57

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi:10.1186/1471-2105-9-559

Langfelder, P., and Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* 46, i11. doi:10.18637/jss.v046.i11

Leming, M. T., Rund, S. S., Behura, S. K., Duffield, G. E., and O'Tousa, J. E. (2014). A database of circadian and diel rhythmic gene expression in the yellow fever mosquito *Aedes aegypti*. *BMC Genomics* 15:1128. doi:10.1186/1471-2164-15-1128

Lorenzo, M. G., Vidal, D. M., and Zarbin, P. H. G. (2014). Control of neglected disease insect vectors: future prospects for the use of tools based on behavior manipulation-interference. *J. Braz. Chem. Soc.* 25, 1799–1809. doi:10.5935/0103-5053.20140202

Luplertlop, N., Surasombatpattana, P., Patramool, S., Dumas, E., Wasinpiyamongkol, L., Saune, L., et al. (2011). Induction of a peptide with activity against a broad spectrum of pathogens in the *Aedes aegypti* salivary gland, following infection with dengue virus. *PLoS Pathog.* 7:e1001252. doi:10.1371/journal.ppat.1001252

Mackenzie, J. S., Gubler, D. J., and Petersen, L. R. (2004). Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nat. Med.* 10, S98–S109. doi:10.1038/nm1144

Medlock, J. M., Hansford, K. M., Schaffner, F., Versteirt, V., Hendrickx, G., Zeller, H., et al. (2012). A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector Borne Zoonotic Dis.* 12, 435–447. doi:10.1089/vbz.2011.0814

Morens, D. M., and Fauci, A. S. (2014). Chikungunya at the door – Déjà Vu all over again? *N. Engl. J. Med.* 371, 885–887. doi:10.1056/NEJMp1408509

Mosso, C., Galván-Mendoza, I. J., Ludert, J. E., and del Angel, R. M. (2008). Endocytic pathway followed by dengue virus to infect the mosquito cell line C6/36 HT. *Virology* 378, 193–199. doi:10.1016/j.virol.2008.05.012

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z., et al. (2007). Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 316, 1718–1723. doi:10.1126/science.1138878

Pan, Y., Tiong, K. H., Abd-Rashid, B. A., Ismail, Z., Ismail, R., Mak, J. W., et al. (2012). Inhibitory effects of cytochrome P450 enzymes CYP2C8, CYP2C9, CYP2C19 and CYP3A4 by *Labisia pumila* extracts. *J. Ethnopharmacol.* 143, 586–591. doi:10.1016/j.jep.2012.07.024

Petersen, E., Wilson, M. E., Touch, S., McCloskey, B., Mwaba, P., Bates, M., et al. (2016). Rapid spread of Zika virus in the Americas – implications for public health preparedness for mass gatherings at the 2016 Brazil Olympic Games. *Int. J. Infect. Dis.* 44, 11–15. doi:10.1016/j.ijid.2016.02.001

Porse, C. C., Kramer, V., Yoshimizu, M. H., Metzger, M., Hu, R., Padgett, K., et al. (2015). Public health response to *Aedes aegypti* and *A. albopictus*

mosquitoes invading California, USA. *Emerg. Infect. Dis.* 21, 1827–1829. doi:10.3201/3210.150494

Pradier, S., Lecollinet, S., and Leblond, A. (2012). West Nile virus epidemiology and factors triggering change in its distribution in Europe. *Rev. Sci. Tech.* 31, 829–844. doi:10.20506/rst.31.3.2167

Ptitsyn, A. A., Reyes-Solis, G., Saavedra-Rodriguez, K., Betz, J., Suchman, E. L., Carlson, J. O., et al. (2011). Rhythms and synchronization patterns in gene expression in the *Aedes aegypti* mosquito. *BMC Genomics* 12:153. doi:10.1186/1471-2164-12-153

Ricklefs, R. E., Goldberg, T. L., Ruiz, M. O., Hamer, G. L., Medeiros, M. C. I., and Brawn, J. D. (2016). Overlap in the seasonal infection patterns of avian malaria parasites and West Nile virus in vectors and hosts. *Am. J. Trop. Med. Hyg.* 95, 1121–1129. doi:10.4269/ajtmh.16-0236

Roth, A., Mercier, A., Lepers, C., Hoy, D., Duituturaga, S., Benyon, E., et al. (2014). Concurrent outbreaks of dengue, chikungunya and Zika virus infections – an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012–2014. *Eurosurveillance* 19, 20929. doi:10.2807/1560-7917.ES2014.19.41.20929

Sathler-Avelar, R., Vitelli-Avelar, D. M., Mattoso-Barbosa, A. M., Perdigão-de-Oliveira, M., Costa, R. P., Elói-Santos, S. M., et al. (2016). Phenotypic features of circulating leukocytes from non-human primates naturally infected with *Trypanosoma cruzi* resemble the major immunological findings observed in human Chagas disease. *PLoS Negl. Trop. Dis.* 10:e0004302. doi:10.1371/journal.pntd.0004302

Seixas, G., Salgueiro, P., Silva, A. C., Campos, M., Spenassatto, C., Reyes-Lugo, M., et al. (2013). *Aedes aegypti* on Madeira Island (Portugal): genetic variation of a recently introduced dengue vector. *Mem. Inst. Oswaldo Cruz* 108, 3–10. doi:10.1590/0074-0276130386

Sim, S., and Dimopoulos, G. (2010). Dengue virus inhibits immune responses in *Aedes aegypti* cells. *PLoS ONE* 5:e10678. doi:10.1371/journal.pone.0010678

Sim, S., Ramirez, J. L., and Dimopoulos, G. (2012). Dengue virus infection of the *Aedes aegypti* salivary gland and chemosensory apparatus induces genes that modulate infection and blood-feeding behavior. *PLoS Pathog.* 8:e1002631. doi:10.1371/journal.ppat.1002631

Skalsky, R. L., and Cullen, B. R. (2010). Viruses, microRNAs, and host interactions. *Annu. Rev. Microbiol.* 64, 123–141. doi:10.1146/annurev.micro.112408.134243

Slavov, S. N., Otaguiri, K. K., Kashima, S., and Covas, D. T. (2016). Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. *Braz. J. Med. Biol. Res.* 49, e5420. doi:10.1590/1414-431X20165420

Souza-Neto, J. A., Sim, S., and Dimopoulos, G. (2009). An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17841–17846. doi:10.1073/pnas.0905006106

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE* 6(7), e21800. doi:10.1371/journal.pone.0021800

Umesh, A., Cohen, B. N., Ross, L. S., and Gill, S. S. (2003). Functional characterization of a glutamate/aspartate transporter from the mosquito *Aedes aegypti*. *J. Exp. Biol.* 206, 2241–2255. doi:10.1242/jeb.00430

Wang, H., Kavanaugh, M. P., North, R. A., and Kabat, D. (1991). Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. *Nature* 352, 729–731. doi:10.1038/352729a0

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10(1), 57–63. doi:10.1038/nrg2484

Wickham, H. (2016). *Elegant Graphics for Data Analysis*. Springer. Available at: http://moderngraphics11.pbworks.com/f/ggplot2-Book09hWickham.pdf

World Health Organization. (2015). *Global Strategy for Dengue Prevention and Control, 2012–2020*. WHO. Available at: http://www.who.int/denguecontrol/9789241504034/en/

Zanluca, C., Melo, V. C., Mosimann, A. L., Santos, G. I., Santos, C. N., and Luz, K. (2015). First report of autochthonous transmission of Zika virus in Brazil. *Mem. Inst. Oswaldo Cruz* 110, 569–572. doi:10.1590/0074-02760150192

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Fukutani, Kasprzykowski, Paschoal, Gomes, Barral, de Oliveira, Ramos and Queiroz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 April 2018
doi: 10.3389/fmicb.2018.00802

Microbial Community Analyses of the Deteriorated Storeroom Objects in the Tianjin Museum Using Culture-Independent and Culture-Dependent Approaches

Zijun Liu[1†], Yanhong Zhang[2†], Fengyu Zhang[1], Cuiting Hu[1], Genliang Liu[2] and Jiao Pan[1]*

[1] Key Laboratory of Molecular Microbiology and Technology of the Ministry of Education, Department of Microbiology, College of Life Sciences, Nankai University, Tianjin, China, [2] Tianjin Museum, Tianjin, China

In the storeroom C7 of the Tianjin Museum, one wooden desk and two leather luggages dated back to Qing dynasty (1644-1912 AD) presented viable microbial contamination. The aim of the present study was to investigate microbial communities responsible for the biodeterioration of storeroom objects using a combination of culture-independent and culture-dependent methods as well microscopic techniques. Scanning electron microscopy (SEM) revealed that the microflora on three storeroom objects were characterized by a marked presence of *Eurotium halophilicum*. Real-time quantitative polymerase chain reaction (qPCR) analysis proved that fungi were the main causative agents behind the biodeterioration in this case. Fungal internal transcribed spacer (ITS) amplicon sequencing documented the presence of two main fungi — *Eurotium halophilicum* and *Aspergillus penicillioides*. Molecular identification of fungal strains isolated from the surfaces and the air of the storeroom were most closely related to *Chaetomium*, *Aspergillus*, *Penicillium*, and *Fusarium*, showing discrepancies in fungal taxa compared to ITS amplicon sequencing. The most isolated bacterial phylum was *Firmicutes*, mostly *Bacillus* members. In addition, four biocide products — Preventol® D 7, P 91, 20 N and Euxyl® K 100 were selected to test their capability against fungal strains isolated from the surfaces. According to the susceptibility assay, Preventol® D 7 based on isothiazolinones was the most effective against fungal isolates. Findings from this study provided a knowledge about storeroom fungi, and exemplify a type of preliminary test that may be conducted before planning any biocide treatment, which may be useful to mitigate the fungal deterioration for further conservation of the museum.

Keywords: Tianjin Museum, fungi, biodeterioration, *Eurotium halophilicum*, biocides

OPEN ACCESS

Edited by:
Florence Abram, National University of Ireland Galway, Ireland

Reviewed by:
Mariusz Cycon, Medical University of Silesia, Poland
Ramón Alberto Batista-García, Universidad Autónoma del Estado de Morelos, Mexico

***Correspondence:**
Jiao Pan
panjiaonk@nankai.edu.cn

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *22 January 2018*
Accepted: *10 April 2018*
Published: *30 April 2018*

Citation:
Liu Z, Zhang Y, Zhang F, Hu C, Liu G and Pan J (2018) Microbial Community Analyses of the Deteriorated Storeroom Objects in the Tianjin Museum Using Culture-Independent and Culture-Dependent Approaches. Front. Microbiol. 9:802. doi: 10.3389/fmicb.2018.00802

INTRODUCTION

Museums are institutions which collect and preserve a wide variety of historical objects, such as paintings, parchment, wood, paper, and rubber. All of these objects represent organic substrates that can well support fungal and bacterial growth if the requirements for growth and the environmental conditions turn suitable, which can cause aesthetical changes on the surfaces of

objects, such as discoloration or biofilms, and can weaken the structure of materials until complete destruction occurs (Cappitelli et al., 2010; Sterflinger and Pinzari, 2012).

In museums, fungi play the most important role in biodeterioration, since, in comparison with bacteria, they can grow in environments with lower temperature and relative humidity. Typical fungal infections colonizing kinds of organic artifacts are often caused by species of slow-growing *Ascomycetes* such as genera *Aspergillus*, *Penicillium*, *Cladosporium*, *Alternaria*, *Chaetomium*, *Eurotium* etc. (Sterflinger and Piñar, 2013). Bacteria rarely exist on museum objects and their number increases significantly only when museum or library are damp, flooded or when the drying process of this type of material is too slow. Nevertheless, bacterial genera *Bacillus*, *Staphylococcus*, *Pseudomonas*, *Virgibacillus,* and *Micromonospora* have been still isolated from deteriorated parchments conserved in the Slovak National Library (Kraková et al., 2012). Microbial contaminations in environments depends not only on species-related properties but also on climatic conditions, such as temperature, humidity and ventilation (Sterflinger, 2010).

The detection and identification of microorganisms associated with biodeterioration are the first necessary step for understanding the effects of microorganisms on cultural heritages objects. Traditionally, the methodology was the application of cultivation methods or microscopy, which was useful for knowing the physiological characteristics of pure isolated strains and for the development of metabolic studies. However, these classical methods are known to have many disadvantages (e.g., only a small proportion microorganism could be isolated) that lead to an underestimation of the composition of the colonizing microflora (Ward et al., 1990). In the last decades, culture-independent methods such as denaturing gradient gel electrophoresis (DGGE) and clone library have been developed and widely applied to study microbial communities on biodeteriorated cultural materials (Hugenholtz and Pace, 1996; Laiz et al., 2003; Pangallo et al., 2015; Piñar et al., 2015b). Notwithstanding these methods based on molecular techniques have offered a deeper insight and understanding of microbial communities over traditional cultivation methods, the fact that its expensiveness, time consuming and heavy workload should not be overlooked. In recent years, next-generation sequencing (NGS) techniques have been developed to characterize microbial community structure in many fields, and also begin to becomes available in the field of conservation and restoration to study microorganisms involved in the biodeterioration of cultural heritage (Shokralla et al., 2012; Gutarowska et al., 2015; Adamiak et al., 2017; Liu et al., 2017). Another method that is broad-coverage, sensitive and specific is real-time quantitative polymerase chain reaction (qPCR), which have been widely used for microbial quantification in environmental sciences (Zhang and Fang, 2006; Kim et al., 2013). Nevertheless, only very few researchers applied it in the cultural assets studies.

In this study, an impressive example of microbial deterioration in the Tianjin Museum is documented and analyzed. In September 2016, visible signs of biodeterioration were observed on the surfaces of the storeroom objects. As a basis for the further preservation of these artifacts, it was necessary to analyze the microflora colonizing these storeroom objects. To this end, samples from biodeteriorated objects were analyzed by SEM and qPCR to reveal the nature of the microflora, and then microbial communities were assessed using the amplicon sequencing techniques. In addition, culture-dependent approaches were conducted in order to complement the data obtained by amplicon sequencing. Finally, four biocidal products were selected to test their effectiveness against fungal strains isolated from the storeroom objects.

MATERIALS AND METHODS

Description of the Studied Site

The Tianjin Museum can be traced back to a predecessor of the same name founded in 1918, making it one of the oldest museums in China. Today the Tianjin Museum's diverse collection includes over 200,000 objects and a 200,000-volume library, making the museum an institution that melds culture and history. In 2008 it was recognized as a first-grade Chinese museum. Work on the new Tianjin Museum building began in 2008, designed by the Architectural Design Department of South China University of Technology. It was completed and opened to the public in 2012. This modern building has five split-level floors above ground and one basement floor, giving an expansive sense of space.

All storerooms of Tianjin Museum are in the basement. Out of 49 storerooms, storeroom C7 is one of the largest storerooms, with 238.9 m^2. The temperature (T) is 22±2°C and relative humidity (RH) is 58.1±5% in this storeroom. Three artifacts—one wooded desk (number: 2010-kou-21-6) and two leather luggages (number: 2010-kou-23-1, 2010-kou-23-2) from Qing dynasty (1644-1912 AD) are presenting viable microbial contamination in the storeroom (**Figure 1**).

Media

In order to isolate microorganisms from the deteriorated objects (FD, PX1 and PX2), two media were prepared: Malt Extract Agar (MEA) medium supplemented with 50 μg/mL chloramphenicol to avoid bacterial growth and Trypticase Soy Agar (TSA) medium supplemented with 100 μg/mL nystatin to avoid fungal growth; The potato dextrose agar (PDA) medium was used for susceptibility testing.

Sampling

(I) Samples for microscopic analysis was obtained using double sided carbon adhesive tape. In brief, a strip of carbon adhesive tape is gently attached to the surface of three objects (refered to FD, PX1 and PX2) and then taken to the laboratory. Collected samples were divided into two aliquots for optical and scanning electron microscopy (SEM).

(II) In the same sampling points, samples for amplicon sequencing and qPCR were taken by minimal-invasive sampling techniques using sterile scalpels. There samples were scraped off from the surfaces showing visible mycelia

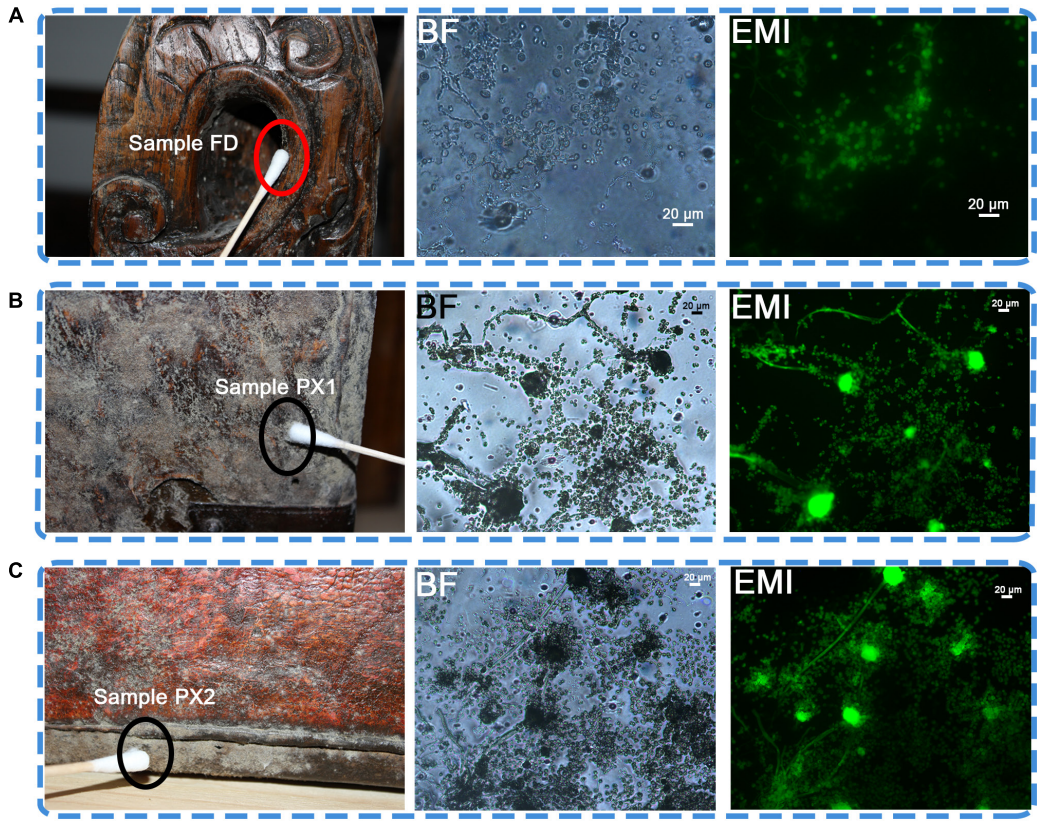


FIGURE 1 | Sampling sites for microbiological analyses on the surface of three objects showing visible microbial contamination. **(A)** Sample FD on a wooden desk. **(B,C)** Detail of sample PX1 and PX2. Visible signs of biofilm were present on the two luggage; BF, bright field. EMI, epifluorescence microscope images. In the bright filed, lots of fungal conidiophores and spores were observed. In the epifluorescence microscope images, the presence of a green fluorescence indicated the viability of microbial cells.

and then taken to the laboratory in an ice box for subsequent analyses.

(III) The Petri dishes (9 cm in diameter) containing MEA and TSA media were opened and some areas showing visible mycelia were streak-inoculated in the agar surface by using sterile cotton bud. Afterward these Petri dishes were brought to the laboratory for isolation and cultivation of microorganism.

Microscopic Analysis

The viability of microbial communities colonizing surface of storerooms objects was assessed by fluorescein diacetate (FDA) staining. Adhesive tape samples were stained with a FDA solution (20 mg of FDA in 1 mL of dimethyl sulfoxide, then diluted with phosphate buffered saline solution to 20 μg/ml) for 20 min of incubation in the dark at 20°C, then observed by Nikon Eclipse 80i epifluorescent microscope (blue excitation wave length, 450–490 nm). Active structures were assessed by the presence of a greenish fluorescence emanating from the cytoplasm of spores and hyphae, due to the liberation of fluorescein by enzymatic (hydrolytic) cleavage.

Tape samples were coated with gold and viewed using a scanning electron microscope (FEI Quanta 200). Images were obtained at magnifications between 1.0 k× and 2.0 k×, and at 15.0 kV for imaging.

Isolation and Identification of Microorganisms

These Petri dishes containing MEA and TSA media were incubated at 28°C for 5–30 days depending on the growth of microorganisms. Colonies showing different morphology and appearance were transferred to fresh plates to obtain further pure isolates.

DNA extraction of pure strains isolated from the surfaces and the air was performed using the CTAB method (Möller et al., 1992). Molecular identification of fungal strains was performed by amplification of ITS, 28S rRNA and RNA polymerase II largest subunit (RPB1) gene (White et al., 1990; Hofstetter et al., 2007; Vilgalys, 2018). In the case of bacterial identification, 16S rRNA gene was used (Muyzer et al., 1993). The primer sequences were summarized in **Supplementary Table S1**. PCR reaction mixtures consisted of a total volume of 50 μL containing 2 μL of genomic DNA, 5 μL of 10× Reaction Buffer, 4 μL of 2.5 mM dNTP mix, 2 μL of 10 μM forward primer, 2 μL of 10 μM reverse primer, 0.5 μL of 5 U/μL Transtaq-T DNA polymerase (TransGen Biotech, China), and ddH2O to 50 μL.

The PCR reaction programs were summarized in Table S2. PCR products were detected by electrophoresis in 1% agarose gels and were purified using a AxyPrep PCR Clean Up Kit (Axygen, United States).

The purified PCR products were sequenced by GENEWIZ (Beijing, China). The sequences obtained were analyzed using the National Center for Biotechnology Information (NCBI) BLAST program[1].

Studies of Airborne Communities

Airborne microorganisms were collected at two sites of storeroom C7 near the deteriorated objects. Air sampler ZR-2050 (Junray, China) with a rate flow 100 L/min was used for air sampling. Three replicates of 100 L of air was taken on Petri dishes with MEA and TSA at each site. Afterward these Petri dishes were brought to the laboratory for incubation at 28°C for 4–7 days. The viable microbial concentrations were calculated as colony forming units per 1 m^3 (CFU/m^3). The isolation and identification of microorganisms were performed according to the methods mentioned above.

Amplicon Sequencing

DNA Extraction

DNA extraction of collected samples (FD, PX1 and PX2) was performed using the MoBio PowerSoil® DNA Isolation Kit (Mo Bio Laboratories, United States) following the manufacturer's protocol. After DNA extraction, the DNA yield and purity (A260/A280 ratio) were assessed using the BioDrop μLite PC Spectrophotometer (Cambridge, United Kingdom). DNA of each sample were divided into two parts; one part was used for amplicon sequencing, while the other were used for qPCR.

PCR Amplification

Fungal communities were studied by amplifying internal transcribed spacer 1 (ITS1) fragments using primers ITS5-1737F/ITS2-2043R, while bacterial communities were investigated by amplifying 16S rRNA gene V4 regions with primers 515F/806R combined with adapter sequences and barcode sequences (Caporaso et al., 2011; Degnan and Ochman, 2012) (**Supplementary Table S2**). Amplifications were carried out in a 50 μL mixture including 25 μL of Master Mix (2X), a 0.5 μM final concentration of the forward and reverse primers, 10 ng of template DNA and nuclease-free water to 50 μL. The PCR conditions were 98°C for 1 min, followed by 30 cycles of 10 s at 98°C, 30 s at 50°C for 16S rRNA gene amplification or 55°C for ITS region amplification, and 30 s at 72°C, with a final extension of 5 min at 72°C.

To visualize PCR amplification success, an equal volume of 1X loading buffer (containing SYBR green) along with PCR products were loaded on a 2% agarose gel. Samples with amplicon bands in the range of 400–450 bp were chosen for further analyses. PCR products from different samples were pooled with equal molar amount. Then, mixture PCR products was purified with Qiagen Gel Extraction Kit (Qiagen, Germany).

Library Preparation and Sequencing

The purified amplicons were prepared for Illumina sequencing by constructing a library using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, United States) following the manufacturer's recommendations. The final library concentrations and quality were checked using a Qubit@ 2.0 Fluorometer (Thermo Scientific) and an Agilent Bioanalyzer 2100 system, respectively. Lastly, the library was sequenced on an Illumina Hiseq2500 PE250 platform.

Bioinformatic Analyses

Paired-end reads were assigned to samples based on unique barcodes and then trimmed of barcode and primer sequences. Paired-end reads were merged using FLASH v. 1.2.7, and the resultant sequences were used as raw tags (Magoč and Salzberg, 2011). Quality filtering of raw tags to obtain high-quality clean tags was performed according to the QIIME v. 1.7.0 quality control protocol (Caporaso et al., 2010). Fungal tags were compared with the Unite database v. 20140703, bacterial tags were compared to the SILVA Gold database v. 20110519 using the UCHIME algorithm v. 4.1 to detect chimaera sequences, and sequences flagged as chimaeras were then removed (Edgar et al., 2011). The resultant high-quality sequences were used for further analyses. OTU clustering analysis was performed using the Uparse software v. 7.0.1001 (Edgar, 2013). Sequences with ≥97% similarity in nucleotide identity were assigned to the same OTUs (Operational Taxonomic Units). Representative sequences for each OTU were then used for taxonomic annotation. For each representative fungal sequence, BLAST analysis was performed against the Unite database v. 20140703 in QIIME v. 1.7.0 to taxonomically annotate OTUs (Kõljalg et al., 2013). For bacterial OTUs, the Greengenes database[2] was used with the RDP classifier v. 2.2 algorithm for taxonomic annotation (DeSantis et al., 2006).

Phylogenetic Analysis

The isolated fungi and main OTUs were analyzed using the Molecular Evolutionary Genetics Analysis (MEGA, v. 7.0) software (Kumar et al., 2016) and aligned together with references sequences obtained from GenBank database using the Clustal W program included in the MEGA v. 7.0. Phylogenetic tree was conducted using the MEGA v. 7.0 based on the neighbour-joining method (Saitou and Nei, 1987). Confidence in tree topology was estimated using the bootstrap method (1,000 bootstrap replicates). The tree was visualized and edited using FigTree v. 1.4.3 software (Available at: http://tree.bio.ed.ac.uk/software/figtree/, accessed on 6 April 2018).

Quantitative Real-Time PCR

Fungal contamination was estimated quantifying the total amount of fungal DNA by qPCR using the primers NL1f/LS2r targeted on 28S rRNA gene (Bates and Garcia-Pichel, 2009). The total biomass of bacterial DNA was quantified by qPCR using the primers Eub338/Eub518 targeted on 16S rRNA gene (Fierer et al., 2005). Standard curves were constructed by plotting the logarithm values seven serial decimal dilutions of genomic DNA

[1] https://blast.ncbi.nlm.nih.gov/Blast.cgi

[2] http://greengenes.lbl.gov

in three replicates versus the threshold cycle (Ct) values generated from qPCR analysis. Genomic DNA of *Fusarium solani* and *Escherichia coli* was used as standard template for fungal and bacterial quantification, respectively.

The qPCR was performed in a StepOnePlus^{TM} Real-Time PCR Systems by using the Roche FastStart Universal SYBR Green Master (Rox). Each 20 μL reaction contained 1 μL of DNA template, 2 μL of 10 μM fungal primers NL1f/LS2r or bacterial primers Eub338/Eub518 (**Supplementary Table S1**), 10 μL SYBR Green mix and 7 μL H_2O. The cycling program consisted of an initial denaturing step at 95°C for 10 min, followed by 40 cycles of 95°C 15 s, 60°C 1 min for fungal primers NL1f/LS2r or 40 cycles of 95°C 15 s, 54°C 30 s, 72°C 20 s for bacterial primers Eub338/Eub518. A melt curve analysis was constructed by increasing the temperature from 60°C to 95°C.

Susceptibility Testing

The susceptibility of isolated fungi to four biocides (**Supplementary Table S3**) was tested by disk diffusion method. In brief, the plates containing PDA medium were inoculated with a spore suspension of the tested fungi, then five paper disks (6 mm in diameter) loaded with 30 μL 0.5% biocide were laid on the plates and incubated at 28°C for 7 days. The diameter of the inhibition zones excluding the disk was measured in centimeter. Duplicate tests were carried out for all biocide products.

Nucleotide Sequence Accession Number

The nucleotide sequences of strains have been deposited in the DDBJ/GenBank/EMBL database under the accession numbers MH169231- MH169238, MH171483-MH171491 for fungal ITS sequences, MG818932-MG818941 for fungal isolates and MG818942-MG818946 for bacterial isolates. The raw data generated from amplicon sequencing has been deposited into the NCBI Sequence Read Archive (SRA) database under the accession numbers SAMN08364417∼SAMN08364420.

RESULTS

Microscopic Observation

Optical microscopic observations of adhesive tape samples revealed the presence of fungal structures. Most of the morphological characteristics appeared as a mixture of spherical spores and fungal conidiophores, which were dominant compared to bacterial cells (**Figure 1**). The FDA assay conducted on samples FD, PX1 and PX2 showed the viability of the microbial community colonizing these storerooms items, as showed by the presence of a bright green fluorescence in the fungal conidiophores and spores.

All samples collected directly from storeroom objects using the adhesive tape technique examined under SEM, showed the presence of a main fungal species with characteristic structures. The fungus presents feature that allow its identification as *Eurotium halophilicum*. Typical "hairs" on the hyphae were pointed out by SEM observations (**Figure 2** and **Supplementary Figure S1**). Conidial heads and ellipsoidal conidia with a variable size (5–8 × 5–9 μm) were visible. The shape, ornamentation and dimensions of these conidia were consistent with those of the anamorphous state of *E. halophilicum*, namely *Aspergillus halophilicus* (Samson and Lustgraaf, 1978).

Comparison of Fungal Biomass and Bacterial Biomass

To determine whether or not the fungi were mainly deteriorative agents of the storeroom objects, qPCR was carried out to quantify fungal and bacterial biomass. The standard curves for *Fusarium solani* and *Escherichia coli* showed correlation coefficients >0.98 and qPCR efficiencies >90% (**Figures 3A,B**). The concentrations of fungal DNA (0.31–4.72 ng/μL) were much higher than for bacterial DNA (0.01–0.09 ng/μL) in three samples (**Figure 3C**). This finding confirmed the high levels of fungal contamination.

Microbial Diversity Characterized by Amplicon Sequencing and Cultivation Methods

An accurate study of the colonizing microorganisms is instrumental for the evaluation of the level of actual biological risk and to properly plan long-term preservation of thesis objects. Therefore, DNA was extracted from samples (FD, PX1 and PX2) for amplicon sequencing with specific primers targeting the fungal ITS regions and bacterial 16S rRNA gene fragments. The sample FD was for microbial isolation only because of the failure to construct a library of amplicon sequencing. ITS amplicon sequencing revealed that the overwhelming majority of the fungal taxa belonged to the phylum *Ascomycota* (99.11 and 99.97%) (**Supplementary Figure S2**). The fungal valid reads were assigned to 85 different operational taxonomic units (OTUs), of which 20 OTUs were annotated at the species level (**Figure 4A** and **Supplementary Figure S3**). *Eurotium halophilicum* was the most abundant fungi and accounted for 55.6 and 96.8% on two samples. *Aspergillus penicillioides* was the second most fungi on samples PX1 (29.0%) and PX2 (1.0%). *Aspergillus pseudoglaucus* comprised 2.7% in sample PX1, but only 0.02% in sample PX2. In addition, *Talaromyces funiculosus* were detected on samples PX1 (1.6%) and PX2 (0.83%). *Xeromyces bisporus* were detected on samples PX1 (0.02%) and PX2 (0.59%). Other fungal species only comprised the minute remainder.

The bacterial communities in the two samples were more diverse at the level of phyla than the fungal communities. The predominant phyla in PX1 were *Proteobacteria* (83.42%), Firmicutes (6.23%), Actinobacteria (6.21%) and Bacteroidetes (1.01%) (**Supplementary Figure S2**); however, *Proteobacteria* represented the largest single portion in PX2 (99.62%). At the species level, no single species dominated in bacterial communities and most of the microbes wasn't identified (**Figure 4B**).

A total of nine fungal strains could be isolated and identified using molecular methods from three storerooms objects (**Table 1**, **Supplementary Tables S4**, **S5** and **Supplementary Figure S4**). Originating from the sequenced isolates, 4 isolates were obtained from sample FD, 4 isolates from sample PX1 and 2 isolates from sample PX2. These isolated strains were most closely related to *Penicillium* spp., *Aspergillus* spp., *Chaetomium* spp., *Fusarium*

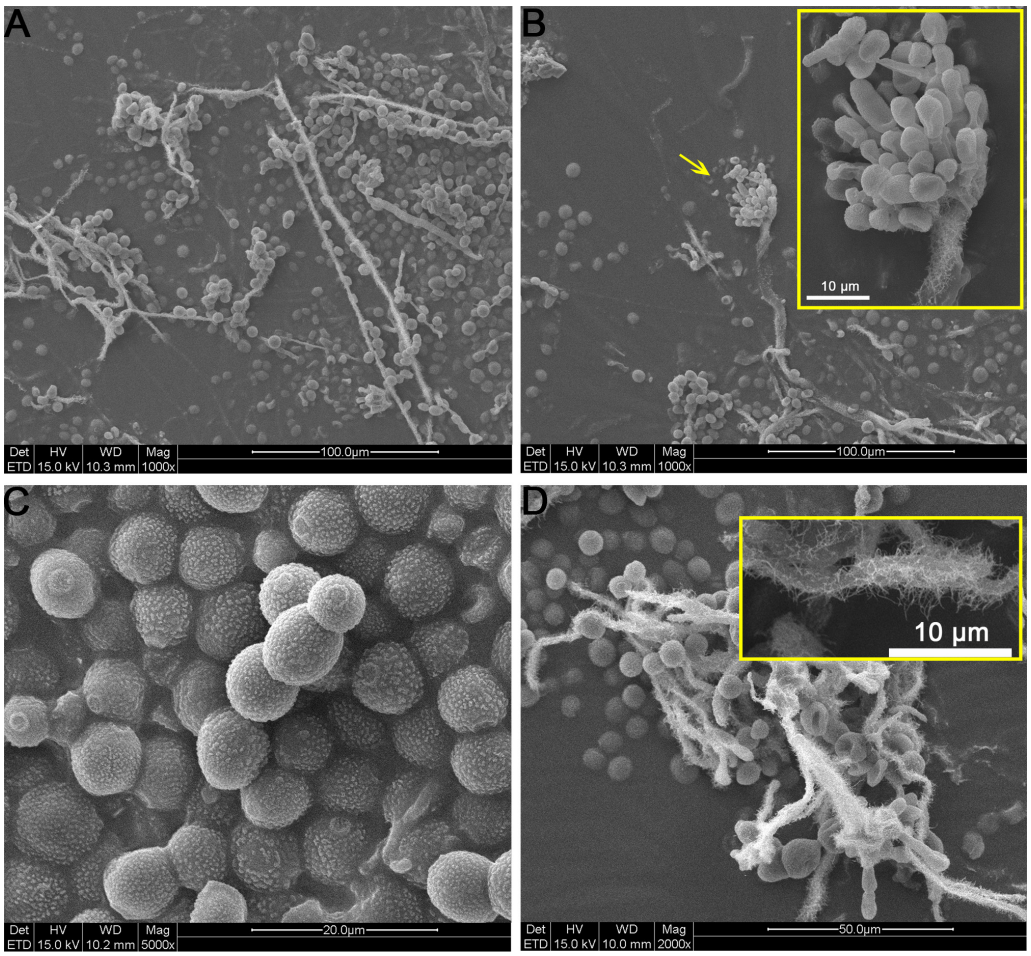


FIGURE 2 | Scanning electron microscopy (SEM) images of *E. halophilicum* elements observed on sample FD. **(A)** Mycelium and conidia. **(B)** Details of typical haired conidial head. **(C)** Hydrated conidia. **(D)** Details of haired hypha.

sp., and *Byssochlamys* sp. Most of the isolated bacterial strains belonged to *Pseudomonas* sp. and *Bacillus* spp. (**Table 1**).

The fungal airborne loads were 86 ± 5 and 103 ± 35 CFU/m^3 at two sites of storeroom C7. Cultural analyses of airborne communities showed the presence of fungi belonging in the genera *Aspergillus*, *Penicillium*, and *Chaetomium* (**Figure 4C**). The airborne loads for bacteria were about 233 ± 24 and 331 ± 59 CFU/m^3. The most frequent taxa were *Micrococcus luteus* (56.29%), *Pseudomonas* sp. (28.14%), and *Bacillus licheniformis* (1.8%) (**Figure 4D**).

Biocide Susceptibility of Fungal Strains

Chemical methods such as application of biocides is an important method to control microbial deterioration. Based on this, four biocide products (**Supplementary Table S2**) were chosen to test their efficacy against fungal isolates.

Biocide susceptibility of major fungi isolated from surfaces in this study was tested by using the disk diffusion method. Both biocides were applied to inhibit the fungal growth on culture plates at described concentrations (0.5%). The concentration is significantly lower than this suggested in manufactures' instructions of commonly used commercial products, in general at concentration 2%. Biocide products were more effective against TJM-F2 than against other strains. While almost no inhibition halo was observed for the use of the four products against TJM-F3 (**Figure 5**). In general, biocide product D 7, based on isothiazolinones, was the most effective against fungal isolates. Biocide products P 91 and 20 N, based on bronopol and isothiazolinones, had similar efficacy. Biocide K 100 combining methylisothiazolinone and benzyl alcohol was least efficient than the other products.

DISCUSSION

A preliminary investigation was performed to assess the nature of the microflora colonizing storeroom objects. On that occasion the non-invasive sampling using adhesive tape strip was used for microscopic and viability assays as it offered the possibility of gaining information on microbial colonization as well as viability of microorganisms without causing damage to the

surface. Three samples (FD, PX1 and PX2) were characterized by a marked presence of microorganisms, particularly fungi. Epifluorescence images revealed lots of fungal hyphae and spores were active.

qPCR techniques have been widely used for studying the levels of individual species and microbial quantification in medicine, agriculture and environmental sciences (Zhang and Fang, 2006). However, to the best of our knowledge, just a few studies successfully used qPCR to quantify microbial contaminations in cultural heritage materials. Martin-Sanchez et al. (2013) developed a qPCR protocol to detect and quantify *Ochroconis lascauxensis* in the Lascaux Cave in France, being this fungus the principal causal agent of the black stains threatening the Paleolithic paintings of this UNESCO World Heritage Site. The protocol required that microbial colonization should be due to a major or single fungus or bacterium (Martin-Sanchez et al., 2013). Ettenauer et al. (2014) developed a qPCR method to detect and quantify fungal abundance using the β-actin gene in five historical buildings materials. In our study, microbial contamination was estimated by qPCR using rRNA primers, confirming fungi are indeed the main causative agents behind the biodeterioration. In general, the qPCR methods targeting the rRNA regions were simple and rapid tools to quantify microbial abundance in cultural heritage materials.

Among the fungi detected by ITS amplicon sequencing, the majority of fungi was *Eurotium halophilicum*, whose conidial state is *Aspergillus halophilicus*. *E. halophilicum* is an obligate xerophilic fungus with high tolerance to water stress. The minimum water activity for its germination and growth of this 0.675, and growth does not occur above 0.935. Because of its particular requirements, the fungus, described for the first time by Christensen et al. (1959), has been recovered from house dust and dry food in association with *Aspergillus penicillioides* and dust mites (Christensen et al., 1959; Hocking and Pitt, 1988; Abdel-Hafez et al., 1990; Xu et al., 2011). More recently, it has been associated with paper and books biodeterioration in museums, libraries or archives. Volumes from an archive of the University of Milan showed whitish-gray discoloration caused by *E. halophilicum* (Polo et al., 2017). Some niches in museums were often colonized by the fungus. These niches are characterized by scarce ventilation and the presence of a water vapor gradient after sudden drop of temperature or night–day thermo hygrometric cycles. These peculiar, often very local, conditions in usually dry environments seemed to promote the development of xerophilic and osmophilic fungal species (Michaelsen et al., 2010; Pinzari and Montanari, 2011; Montanari et al., 2012; Micheluz et al., 2015). Similar contamination patterns and microscopic characteristics detected in these cases suggested that *E. halophilicum* had a large

TABLE 1 | Molecular identification of strains isolated from the surfaces and air.

	Nucleotide Blast reference strains				Surface			Air	
	Closet relative strain	**Similarity (%)**	**Accession number**	**Phylum**	**FD**	**PX1**	**PX2**	**Close to the FD**	**Close to the PX1 and PX2**
Fungi									
TJM-F1	*Penicillium oxalicum*	99%	KP868627	*Ascomycota*	√			√	√
TJM-F2	*Chaetomium globosum*	99%	KX379227	*Ascomycota*	√	√	√		√
TJM-F3	*Fusarium solani*	99%	KT876643	*Ascomycota*	√	√	√		
TJM-F4	*Aspergillus niger*	99%	KM979775	*Ascomycota*	√			√	
TJM-F5	*Penicillium chrysogenum*	99%	GU985086	*Ascomycota*		√			
TJM-F6	*Byssochlamys spectabilis*	98%	KR909186	*Ascomycota*		√			
TJM-F7	*Aspergillus flavus*	99%	LN482516	*Ascomycota*				√	
TJM-F8	*Penicillium sumatrense*	99%	LT558898	*Ascomycota*				√	√
TJM-F9	*Chaetomium* sp.	100%	KC963907	*Ascomycota*				√	√
Bacteria									
TJM-B1	*Bacillus* sp.	99%	MF276685	*Firmicutes*	√				
TJM-B2	*Bacillus megaterium*	100%	JX274543	*Firmicutes*	√	√			
TJM-B3	*Bacillus pumilus*	99%	KX375222	*Firmicutes*	√				
TJM-B4	*Bacillus licheniformis*	100%	FJ976541	*Firmicutes*		√		√	√
TJM-B5	*Micrococcus luteus*	99%	MF405214	*Actinobacteria*				√	√
TJM-B6	*Pseudomonas* sp.	100%	KY681920	*Proteobacteria*	√	√	√	√	√

"TJM-F1 ∼ TJM-F9" and "TJM-B1 ∼ TJM-B6" indicate strains identified by ITS and 16S rRNA gene. '√' indicates presence of the strain.

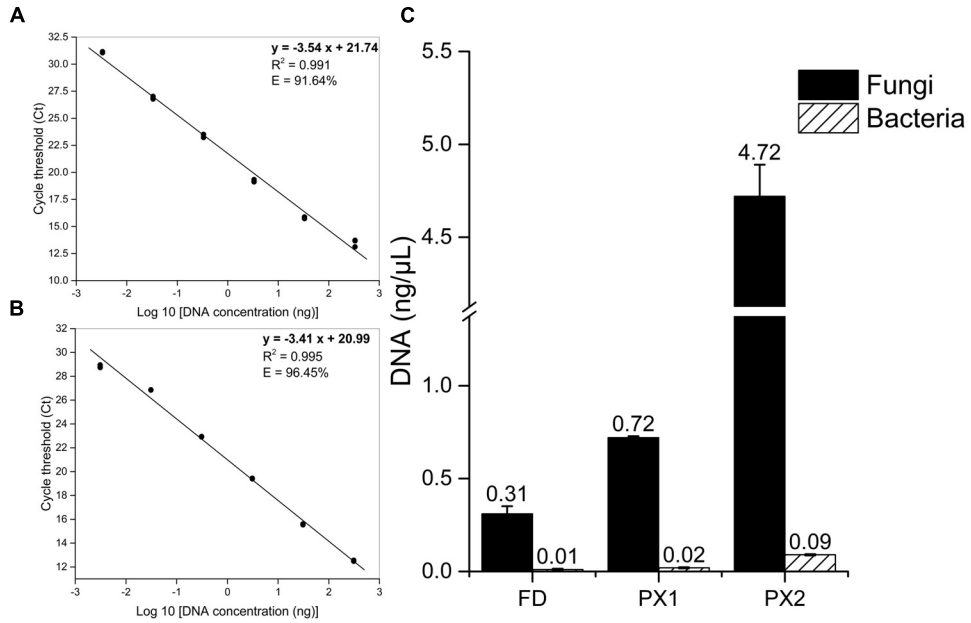


FIGURE 3 | Quantification of microbial biomass of samples by qPCR. **(A)** Standard curve for *F. solani* quantification using fungal primers NS1f/LS2r. **(B)** Standard curve for *E. coli* quantification using bacterial primers Eub338/Eub518. **(C)** Comparison between fungal and bacterial load.

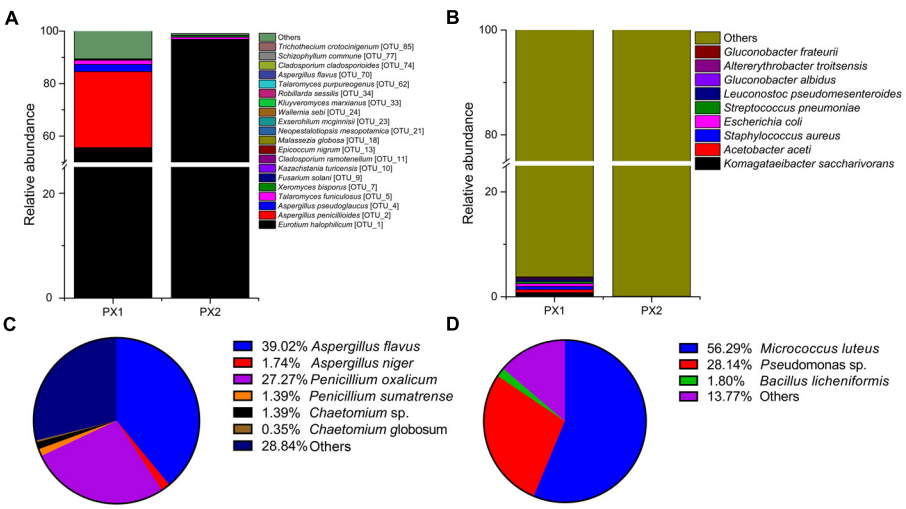


FIGURE 4 | Distribution patterns of fungal **(A)** and bacterial **(B)** species in the two samples. Major fungi **(C)** and bacteria **(D)** isolated from the air.

distribution in the particular environment such as museums or libraries. Finding reported there were consistent with single species *E. halophilicum* contamination by ITS amplicon sequencing. However, molecular identification of isolates in this study did not allow the detection of the dominant fungus *E. halophilicum*. This result was partly expected as the cultivation of *E. halophilicum* on typical media is extremely difficult.

Another xerophilic fungus, *A. penicillioides*, was also frequently isolated from contaminated books and manuscripts. Already in 1978, *A. penicillioides* was associated *E. halophilicum* by Samson and Lustgraaf as cohabiting in house dust, probably due to the similar behavior and low water requirements of both these species (Samson and Lustgraaf, 1978). In particular, *A. penicillioides* are common deteriorative agents of organic and synthetic materials and are often associated with the damage of museum objects as they can secrete a wide variety of enzymes that degrade cellulosic materials and cause discoloration (Abe, 2010; Michaelsen et al., 2010; Principi et al., 2011). In addition to the fungus, *Aspergillus* had a significant proportion in the airborne microbial communities. In consideration of the fact that they significantly affect the conservation of museum items and

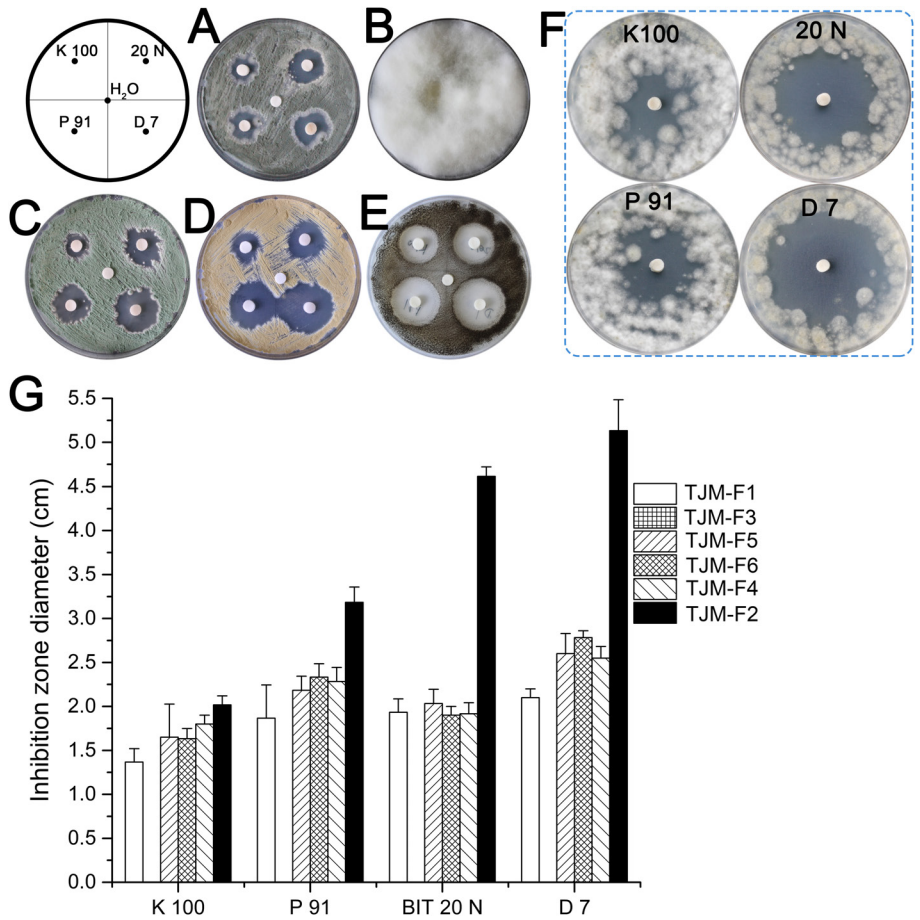


FIGURE 5 | Antifungal assays by disk diffusion method. The inhibition halo around the disk shows the effectiveness of biocides against fungal strains isolated from the surfaces. **(A)** TJM-F1. **(B)** TJM-F3. **(C)** TJM-F5. **(D)** TJM-F6. **(E)** TJM-F4. **(F)** TJM-F2. **(G)** The inhibition zone diameter. Vertical lines indicate standard deviation of three measurements.

the threat to employees of museums by forming mycotoxins or causing allergy diseases, the taxa should not be supposed to be in museums (Abe, 2010; Principi et al., 2011; Krijgsheld et al., 2013). *Fusarium solani* was also detected in the case. It is known to be deteriorative agents because of its damage to the wall paintings of Lascaux Cave (Martin-Sanchez et al., 2012). Therefore, these species also qualify as potential candidates as deteriorative agents of the stored objects.

Particular attention should be placed on the taxa *Chaetomium globosum* which was isolated from all sampling sites and the air by cultivation methods. The fungus is known to possess high cellulolytic activity and can efficiently destroy historical objects such as parchments, textiles, paintings, and wooden sculptures (Sterflinger and Piñar, 2013; Lech, 2016). In particular, *Chaetomium globosum* is a soft-rot fungi capable of degrading cellulose in the S2 layer of the secondary cell wall of wood (Pangallo et al., 2007). Thus the fungus must be regarded as posing a threat to the wooden objects.

With regard to bacteria, few surface strains were detected by 16S amplicon sequencing and cultivation methods. The most isolated bacterial phylum was *Firmicutes*, mostly *Bacillus* members, which were consistent with the literature associated the biodeterioration of paper heritage (Principi et al., 2011; Piñar et al., 2015a). However, the reports presented in the literature contained less information about the bacterial species than the fungal species recovered from paper documents. It could be the lesser role of bacteria in microbial contamination, particularly in terms of biodeterioration of cultural heritage materials.

The different results between amplicon sequencing and cultivation methods suggested that, using the amplicon sequencing, the species which were not abundant may not have been observed, as happened with *Chaetomium* sp. and *Penicillium* sp. However, these strains could be detected by cultivation methods. In contrast, the predominant fungal species, as identified by amplicon sequencing, may have been difficult to detect by cultivation only, as observed for the *E. halophilicum* and *A. penicillioides*. In general, NGS such as amplicon sequencing probably reveal the actual proportion of species present in the experimental system and provide a more realistic picture of the microbial communities, but the best choice may be the combination of traditional cultivation techniques and

culture-independent methods for the detection of the whole microbial communities in mycology.

To control microbial deterioration, the application of biocides is one of the important means in cultural heritage materials (Allsopp et al., 2004). Commercial biocides, like Biotin® T and Preventol® RI 80, have been usually applied by conservators to control biodeterioration of cultural heritage (Fonseca et al., 2010; De los Ríos et al., 2012). However, one of the major concerns involved in the application of biocides is the absence of appropriate monitoring. The potential recolonization has not been monitored after the application of biocides. It is easy to understand that a biocide can exert a selective pressure on the microbial community, so the microbial community may develop resistant mechanisms or the microbial community may be replaced by new microbial community which might have even more greater harm to the cultural heritage (Mitchell and McNamara, 2010). One notorious case is the French Lascaux Caves in which a series of biocide treatments, including penicillin, streptomycin and kanamycin, were applied. These treatments triggered new microbial outbreaks such as white fungal strains (*Fusarium*) and melanized fungal strains (*Ochroconis*) (Martin-Sanchez et al., 2012). Moreover, lots of chemical biocides have been banned because of the environmental and health hazards in the past decade (Coutinho et al., 2016).

So while four biocides tested in this study showed effectiveness against isolates in the laboratory analyses, further studies including human toxicity and respect to historic materials should be performed. In view of the fact that most artifacts in storerooms are in good condition, mechanical methods by hand or with tools such as scalpels or vacuum cleaners are recommended to mitigate current biodeterioration. Moreover, indoor climate is the most important factor for microbial growth. Steady environmental parameters (T 20 ± 2°C and RH 50 ± 3%) are recommended in museums. Slightly higher values (T 22 ± 2°C and RH 58.1 ± 5%) were observed in storeroom C7 and thus climate control should be adjusted below the recommended values. In addition, storeroom C7 is in a relatively confined space and lack of a dusting programmer, air-conditioning and poor ventilation, which increase the risk of fungal growth such as *E. halophilicum* in the case. Thus, the storeroom should be dusted regularly to prevent dirt from being nutritional substances for microorganisms and should improve insufficient condition.

CONCLUSION

We used conventional cultivation methods and modern molecular strategy (i.e., ITS amplicon sequencing and qPCR) as a tool to analyze microbial contamination in Tianjin Museum. The inhabiting members detected, mainly representatives of *E. halophilicum* and *A. penicillioides*, are the main cause of the biofilm on the surfaces of storeroom objects. These fungi, especially *E. halophilicum*, were also the main causative agents behind biodeterioration in other libraries from different countries. However, the fungus was difficult to isolate, even from contaminated surfaces with visible fungal growth. Therefore, further attempts to isolate *E. halophilicum* and study of its characteristics are required. The fungal isolates including *Penicillium* spp. *Aspergillus* spp., *Chaetomium* spp., *Fusarium* sp. and bacterial *Bacillus* members may be not directly responsible for the current biodeterioration, but these strains are known to could degrade organic materials and must be regarded as a threat to the storeroom objects. Base on biocide susceptibility assay, the active compound isothiazolinones was effective inhibiting the growth of fungal isolates. These data provide a valuable knowledge about storeroom fungi, and exemplify a type of preliminary test that may be conducted before planning any biocide treatment. However, considering possible negative impacts caused by the application of biocides, these treatments are a weaker option and not recommended in the current stage. Alternatively, mechanical methods combined with the control of environmental parameters could be conducted.

AUTHOR CONTRIBUTIONS

JP conceived and designed the experiments. ZL, YZ, CH, and FZ performed the experiments. ZL and YZ analyzed the data. ZL and YZ constructed the phylogenetic tree. ZL wrote the paper. YZ and GL assisted in sampling.

FUNDING

This study was funded by Research Project of Tianjin Cultural Relics Museum (TCHM2016012).

ACKNOWLEDGMENTS

We thank the conservators of the Tianjin Museum for the assistance with the samples. We gratefully acknowledge the assistance of Dr. Qiang Li from Zhejiang University in preparing the phylogenetic analysis and the two reviewers' comments, which were extremely valuable to improve this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00802/full#supplementary-material

FIGURE S1 | The presence of *E. halophilicum* structures on sample PX1 and PX2. Typical haired hypha, conidial head and conidia were observed on gold sputtered samples using SEM. **(A–C)** Sample PX1. **(D–F)** Sample PX2.

FIGURE S2 | Distribution patterns of fungal **(A)** and bacterial **(B)** phylum in the two samples.

FIGURE S3 | Neighbor-joining (NJ) phylogeny of OTUs based on ITS1 gene sequences (~400 bp), comprising 20 identified OTUs and 42 reference strains. The Genbank number of sequences used in the construction of this phylogenetic tree is given in brackets. These results are consistent with the OTUs annotation.

FIGURE S4 | Neighbor-joining (NJ) tree of fungal strains based on ITS gene sequences (~525–581 bp sequence used for each), including nine isolated fungi and 32 reference strains. The significance of each branch is indicated by the bootstrap percentage calculated for 1000 bootstraps. Strains TJM-F1, TJM-F5, and TJM-F8 belong to the genus *Penicillium*. Strains TJM-F2 and TJM-F9 are related to the genus *Chaetomium*. Strains TJM-F4 and TJM-F7 can be classified as the genus *Aspergillus*. Strains TJM-F3 and TJM-F6 belong to *Fusarium* and *Byssochlamys*, respectively.

TABLE S1 | Primer sequences used in this study.

TABLE S2 | PCR reaction programs.

TABLE S3 | Active principle of the four biocides used in this study.

TABLE S4 | Molecular identification of fungal strains based on 28S rRNA gene.

TABLE S5 | Molecular identification of fungal strains based on RPB1 gene.

REFERENCES

Abdel-Hafez, S. I., Moubasher, A. H., and Barakat, A. (1990). Keratinophilic fungi and other moulds associated with air-dust particles from Egypt. *Folia Microbiol.* 35, 311–325. doi: 10.1007/bf02821283

Abe, K. (2010). Assessment of the environmental conditions in a museum storehouse by use of a fungal index. *Int. Biodeterior. Biodegrad.* 64, 32–40. doi: 10.1016/j.ibiod.2009.10.004

Adamiak, J., Otlewska, A., Tafer, H., Lopandic, K., Gutarowska, B., Sterflinger, K., et al. (2017). First evaluation of the microbiome of built cultural heritage by using the Ion Torrent next generation sequencing platform. *Int. Biodeterior. Biodegrad.* (in press). doi: 10.1016/j.ibiod.2017.01.040

Allsopp, D., Seal, K. J., and Gaylarde, C. C. (2004). *Introduction to Biodeterioration*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511617065

Bates, S. T., and Garcia-Pichel, F. (2009). A culture-independent study of free-living fungi in biological soil crusts of the Colorado Plateau: their diversity and relative contribution to microbial biomass. *Environ. Microbiol.* 11, 56–67. doi: 10.1111/j.1462-2920.2008.01738.x

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4516–4522. doi: 10.1073/pnas.1000080107

Cappitelli, F., Pasquariello, G., Tarsitani, G., and Sorlini, C. (2010). Scripta manent? Assessing microbial risk to paper heritage. *Trends Microbiol.* 18, 538–542. doi: 10.1016/j.tim.2010.09.004

Christensen, C. M., Papavizas, G. C., and Benjamin, C. R. (1959). A new halophilic species of Eurotium. *Mycologia* 51, 636–640. doi: 10.2307/3755892

Coutinho, M. L., Miller, A. Z., Martin-Sanchez, P. M., Mirão, J., Gomez-Bolea, A., Machado-Moreira, B., et al. (2016). A multiproxy approach to evaluate biocidal treatments on biodeteriorated majolica glazed tiles. *Environ. Microbiol.* 18, 4794–4816. doi: 10.1111/1462-2920.13380

De los Ríos, A., Pérez-Ortega, S., Wierzchos, J., and Ascaso, C. (2012). Differential effects of biocide treatments on saxicolous communities: case study of the Segovia cathedral cloister (Spain). *Int. Biodeterior. Biodegrad.* 67(Suppl. C), 64–72. doi: 10.1016/j.ibiod.2011.10.010

Degnan, P. H., and Ochman, H. (2012). Illumina-based analysis of microbial community diversity. *ISME J.* 6, 183–194. doi: 10.1038/ismej.2011.74

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Ettenauer, J., Piñar, G., Tafer, H., and Sterflinger, K. (2014). Quantification of fungal abundance on cultural heritage using real time PCR targeting the β-actin gene. *Front. Microbiol* 5:262. doi: 10.3389/fmicb.2014.00262

Fierer, N., Jackson, J. A., Vilgalys, R., and Jackson, R. B. (2005). Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl. Environ. Microbiol.* 71, 4117–4120. doi: 10.1128/aem.71.7.4117-4120.2005

Fonseca, A. J., Pina, F., Macedo, M. F., Leal, N., Romanowska-Deskins, A., Laiz, L., et al. (2010). Anatase as an alternative application for preventing biodeterioration of mortars: evaluation and comparison with other biocides. *Int. Biodeterior. Biodegrad.* 64, 388–396. doi: 10.1016/j.ibiod.2010.04.006

Gutarowska, B., Celikkol-Aydin, S., Bonifay, V., Otlewska, A., Aydin, E., Oldham, A. L., et al. (2015). Metabolomic and high-throughput sequencing analysis-modern approach for the assessment of biodeterioration of materials from historic buildings. *Front. Microbiol.* 6:979. doi: 10.3389/fmicb.2015.00979

Hocking, A. D., and Pitt, J. I. (1988). Two new species of xerophilic fungi and a further record of *Eurotium halophilicum*. *Mycologia* 80, 82–88. doi: 10.2307/3807497

Hofstetter, V., Miadlikowska, J., Kauff, F., and Lutzoni, F. (2007). Phylogenetic comparison of protein-coding versus ribosomal RNA-coding sequence data: a case study of the Lecanoromycetes (Ascomycota). *Mol. Phylogenet. Evol.* 44, 412–426. doi: 10.1016/j.ympev.2006.10.016

Hugenholtz, P., and Pace, N. R. (1996). Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol.* 14, 190–197. doi: 10.1016/0167-7799(96)10025-1

Kim, J., Lim, J., and Lee, C. (2013). Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations. *Biotechnol. Adv.* 31, 1358–1373. doi: 10.1016/j.biotechadv.2013.05.010

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. doi: 10.1111/mec.12481

Kraková, L., Chovanová, K., Selim, S. A., Šimonovičová, A., Puškarová, A., Maková, A., et al. (2012). A multiphasic approach for investigation of the microbial diversity and its biodegradative abilities in historical paper and parchment documents. *Int. Biodeterior. Biodegrad.* 70, 117–125. doi: 10.1016/j.ibiod.2012.01.011

Krijgsheld, P., Bleichrodt, R., van Veluw, G. J., Wang, F., Müller, W. H., Dijksterhuis, J., et al. (2013). Development in *Aspergillus*. *Stud. Mycol.* 74, 1–29. doi: 10.3114/sim0006

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Laiz, L., Piñar, G., Lubitz, W., and Saiz-Jimenez, C. (2003). Monitoring the colonization of monuments by bacteria: cultivation versus molecular methods. *Environ. Microbiol.* 5, 72–74. doi: 10.1046/j.1462-2920.2003.00381.x

Lech, T. (2016). Evaluation of a parchment document, the 13th century incorporation charter for the city of Krakow, Poland, for Microbial Hazards. *Appl. Environ. Microbiol.* 82, 2620–2631. doi: 10.1128/AEM.03851-15

Liu, Z., Wang, Y., Pan, X., Ge, Q., Ma, Q., Li, Q., et al. (2017). Identification of fungal communities associated with the biodeterioration of waterlogged archeological wood in a Han dynasty tomb in China. *Front. Microbiol.* 8:1633. doi: 10.3389/fmicb.2017.01633

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Martin-Sanchez, P. M., Bastian, F., Alabouvette, C., and Saiz-Jimenez, C. (2013). Real-time PCR detection of *Ochroconis lascauxensis* involved in the formation of black stains in the Lascaux Cave, France. *Sci. Total Environ.* 443(Suppl. C), 478–484. doi: 10.1016/j.scitotenv.2012.11.026

Martin-Sanchez, P. M., Nováková, A., Bastian, F., Alabouvette, C., and Saiz-Jimenez, C. (2012). Use of biocides for the control of fungal outbreaks in subterranean environments: the case of the Lascaux Cave in France. *Environ. Sci. Technol.* 46, 3762–3770. doi: 10.1021/es2040625

Michaelsen, A., Piñar, G., and Pinzari, F. (2010). Molecular and microscopical investigation of the microflora inhabiting a deteriorated Italian manuscript dated from the thirteenth century. *Microb. Ecol.* 60, 69–80. doi: 10.1007/s00248-010-9667-9

Micheluz, A., Manente, S., Tigini, V., Prigione, V., Pinzari, F., Ravagnan, G., et al. (2015). The extreme environment of a library: xerophilic fungi inhabiting indoor niches. *Int. Biodeterior. Biodegrad.* 99(Suppl. C), 1–7. doi: 10.1016/j.ibiod.2014.12.012

Mitchell, R., and McNamara, C. J. (2010). *Cultural Heritage Microbiology: Fundamental Studies in Conservation Science*. Washington, DC: ASM Press.

Möller, E. M., Bahnweg, G., Sandermann, H., and Geiger, H. H. (1992). A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res.* 20, 6115–6116. doi: 10.1093/nar/20.22.6115

Montanari, M., Melloni, V., Pinzari, F., and Innocenti, G. (2012). Fungal biodeterioration of historical library materials stored in Compactus movable shelves. *Int. Biodeterior. Biodegrad.* 75(Suppl. C), 83–88. doi: 10.1016/j.ibiod.2012.03.011

Muyzer, G., de Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59, 695–700.

Pangallo, D., Bučková, M., Kraková, L., Puškárová, A., Šaková, N., Grivalský, T., et al. (2015). Biodeterioration of epoxy resin: a microbial survey through culture-independent and culture-dependent approaches. *Environ. Microbiol.* 17, 462–479. doi: 10.1111/1462-2920.12523

Pangallo, D., Šimonovičová, A., Chovanová, K., and Ferianc, P. (2007). Wooden art objects and the museum environment: identification and biodegradative characteristics of isolated microflora. *Lett. Appl. Microbiol.* 45, 87–94. doi: 10.1111/j.1472-765X.2007.02138.x

Piñar, G., Sterflinger, K., Ettenauer, J., Quandt, A., and Pinzari, F. (2015a). A combined approach to assess the microbial contamination of the archimedes palimpsest. *Microb. Ecol.* 69, 118–134. doi: 10.1007/s00248-014-0481-7

Piñar, G., Sterflinger, K., and Pinzari, F. (2015b). Unmasking the measles-like parchment discoloration: molecular and microanalytical approach. *Environ. Microbiol.* 17, 427–443. doi: 10.1111/1462-2920.12471

Pinzari, F., and Montanari, M. (2011). "Mould growth on library materials stored in compactus-type shelving units," in *Sick Building Syndrome: in Public Buildings and Workplaces*, ed. S. A. Abdul-Wahab (Berlin: Springer), 193–206.

Polo, A., Cappitelli, F., Villa, F., and Pinzari, F. (2017). Biological invasion in the indoor environment: the spread of *Eurotium halophilicum* on library materials. *Int. Biodeterior. Biodegrad.* 118, 34–44. doi: 10.1016/j.ibiod.2016.12.010

Principi, P., Villa, F., Sorlini, C., and Cappitelli, F. (2011). Molecular studies of microbial community structure on stained pages of Leonardo da Vinci's Atlantic Codex. *Microb. Ecol.* 61, 214–222. doi: 10.1007/s00248-010-9741-3

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Samson, R. A., and Lustgraaf, B. V. D. (1978). Aspergillus penicilloides and Eurotium halophilicum in association with house-dust mites. *Mycopathologia* 64, 13–16. doi: 10.1007/bf00443082

Shokralla, S., Spall, J. L., Gibson, J. F., and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* 21, 1794–1805. doi: 10.1111/j.1365-294X.2012.05538.x

Sterflinger, K. (2010). Fungi: their role in deterioration of cultural heritage. *Fungal Biol. Rev.* 24, 47–55. doi: 10.1016/j.fbr.2010.03.003

Sterflinger, K., and Piñar, G. (2013). Microbial deterioration of cultural heritage and works of art — tilting at windmills? *Appl. Microbiol. Biotechnol.* 97, 9637–9646. doi: 10.1007/s00253-013-5283-1

Sterflinger, K., and Pinzari, F. (2012). The revenge of time: fungal deterioration of cultural heritage with particular reference to books, paper and parchment. *Environ. Microbiol.* 14, 559–566. doi: 10.1111/j.1462-2920.2011.02584.x

Vilgalys, R. (2018). *Conserved Primer Sequences for PCR Amplification of Fungal rDNA [Online]*. Available at: https://sites.duke.edu/vilgalyslab/rdna_primers_for_fungi/

Ward, D. M., Weller, R., and Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63–65. doi: 10.1038/345063a0

White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *PCR Protocols a Guide to Methods and Applications*, eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (New York, NY: Academic Press), 315–322.

Xu, A., Wang, Y., Wen, J., Liu, P., Liu, Z., and Li, Z. (2011). Fungal community associated with fermentation and storage of Fuzhuan brick-tea. *Int. J. Food Microbiol.* 146, 14–22. doi: 10.1016/j.ijfoodmicro.2011.01.024

Zhang, T., and Fang, H. H. P. (2006). Applications of real-time polymerase chain reaction for quantification of microorganisms in environmental samples. *Appl. Microbiol. Biotechnol.* 70, 281–289. doi: 10.1007/s00253-006-0333-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Liu, Zhang, Zhang, Hu, Liu and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 24 August 2017
doi: 10.3389/fmicb.2017.01633

Identification of Fungal Communities Associated with the Biodeterioration of Waterlogged Archeological Wood in a Han Dynasty Tomb in China

Zijun Liu[1†], Yu Wang[1†], Xiaoxuan Pan[2], Qinya Ge[2], Qinglin Ma[2], Qiang Li[3], Tongtong Fu[1], Cuiting Hu[1], Xudong Zhu[1] and Jiao Pan[1]*

[1] Key Laboratory of Molecular Microbiology and Technology for Ministry of Education, Department of Microbiology, College of Life Sciences, Nankai University, Tianjin, China, [2] Chinese Academy of Cultural Heritage, Beijing, China, [3] Laboratory of Cultural Relics Conservation Materials, Department of Chemistry, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Florence Abram,
NUI Galway, Ireland

Reviewed by:
Virginie Chapon,
Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), France
Maher Gtari,
Tunis El Manar University, Tunisia
Anna Otlewska,
Lodz University of Technology, Poland

***Correspondence:**
Jiao Pan
panjiaonk@nankai.edu.cn

[†] These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *13 June 2017*
Accepted: *11 August 2017*
Published: *24 August 2017*

Citation:
Liu Z, Wang Y, Pan X, Ge Q, Ma Q, Li Q, Fu T, Hu C, Zhu X and Pan J (2017) Identification of Fungal Communities Associated with the Biodeterioration of Waterlogged Archeological Wood in a Han Dynasty Tomb in China. Front. Microbiol. 8:1633. doi: 10.3389/fmicb.2017.01633

The Mausoleum of the Dingtao King (termed 'M2') is a large-scale *huangchang ticou* tomb that dates to the Western Han Dynasty (206 B.C.–25 A.D.). It is the highest-ranking Han Dynasty tomb discovered to date. However, biodeterioration on the surface of the tomb M2 is causing severe damage to its wooden materials. The aim of the present study was to give insight into the fungal communities colonized the wooden tomb. For this purpose, seven samples were collected from different sections of the tomb M2 which exhibited obvious biodeterioration in the form of white spots. Microbial structures associated with the white spots were observed with scanning electron microscopy. Fungal community structures were assessed for seven samples via a combination of high-throughput sequencing and culture-dependent techniques. Sequencing analyses identified 114 total genera that belonged to five fungal phyla. *Hypochnicium* was the most abundant genus across all samples and accounted for 98.61–99.45% of the total community composition. Further, *Hypochnicium* sp. and *Mortierella* sp. cultures were successfully isolated from the tomb samples, and were distinguished as *Hypochnicium* sp. WY-DT1 and *Mortierella* sp. NK-DT1, respectively. Cultivation-dependent experiments indicated that the dominant member, *Hypochnicium* sp. WY- DT1, could grow at low temperatures and significantly degraded cellulose and lignin. Thus, our results taken together suggest that this fungal strain must be regarded as a serious threat to the preservation of the wooden tomb M2. The results reported here are useful for informing future contamination mitigation efforts for the tomb M2 as well as other similar cultural artifacts.

Keywords: Han Dynasty tomb, *huangchang ticou*, biodeterioration, high-throughput sequencing, fungal community, *Hypochnicium* sp.

INTRODUCTION

The 'M2' Mausoleum of the Dingtao King dates to the Western Han Dynasty and is located 200 m northwest of the Lijiacun Village in the Maji Township of the Dingtao District of Shandong Heze City (**Figure 1**). The site has been referred to as one of China's 10 most important archeological discoveries due to its value toward research on the imperial mausoleums of the Western Han

Dynasty. The M2 tomb represents the best-preserved large-scale *huangchang ticou* tomb that has been found in China and is very valuable for research on the *huangchang ticou* (黃腸題湊) burial system (Cui et al., 2013).

Huangchang ticou was a burial form in ancient China that emerged in the spring and autumn period (770–476 B.C.) and became prevalent during the Western Han Dynasty (206 B.C.–A.D. 25). *Huangchang ticou* consist of piled wooden walls of cypress heartwood timbers that encompass the inner and outer coffins of the tomb occupant. The burial form material is usually cypress with bark that has been removed (Cui et al., 2013). *Huangchang* indicates the color and shape of the material and refers to the cypress heartwood timbers that appear faint yellow. *Ticou* refers to the shape and structure of the piled-up material, which largely comprises wooden cross-sections of yellow cypress that face the inside and outside of the chambers and form the *ticou* wall (Huang, 1998). *Huangchang ticou* was a burial standard for emperors like jade shrouds, catalpa coffins, burial lounges, and exterior burial pits for carriages or kitchen wares and grains. Later, the standard was extended to empresses, imperial concubines, favored ministers, and vassal kings and queens. Currently, there are about 10 *huangchang ticou* tombs from the Han Dynasty that are being excavated in China.

The Mausoleum of the Dingtao King was excavated in October of 2010. The tomb resembles the Chinese character "甲" and its roof and ticou walls are sealed with bricks. The whole chamber is square with 22.8 m long walls. The wooden coffin chambers exhibit a large-scale huangchang ticou architectural complex that consists of four main parts—a front chamber, a central chamber, a rear chamber, and its eight side chambers. The structure also consists of 12 outer storage chambers, corridors, passages connecting the main chambers, four doorways, and ticou walls (**Supplementary Figure S1**). As a result of underground water intrusion, the tomb is highly waterlogged, and has consequently remained well-conserved.

After its discovery, fungal mycelia were observed on the surface of the tomb M2, as noted by white spots that were first observed in September 2012 (**Figure 2A** and **Supplementary Figure S2**). Since the onset of our investigation into the biodeterioration of this wooden tomb in March 2015, the white spots have spread to every chamber. In addition to the white spots, insects including fungus gnats, spiders, and millipedes have been found in the tomb that may also contribute to the propagation of the white spots. Microorganisms, and especially filamentous fungi, are able to cause biodeterioration of archeological wood (Irbe et al., 2012; Ortiz et al., 2014; Gutarowska et al., 2015; Piñar et al., 2016). Thus, the

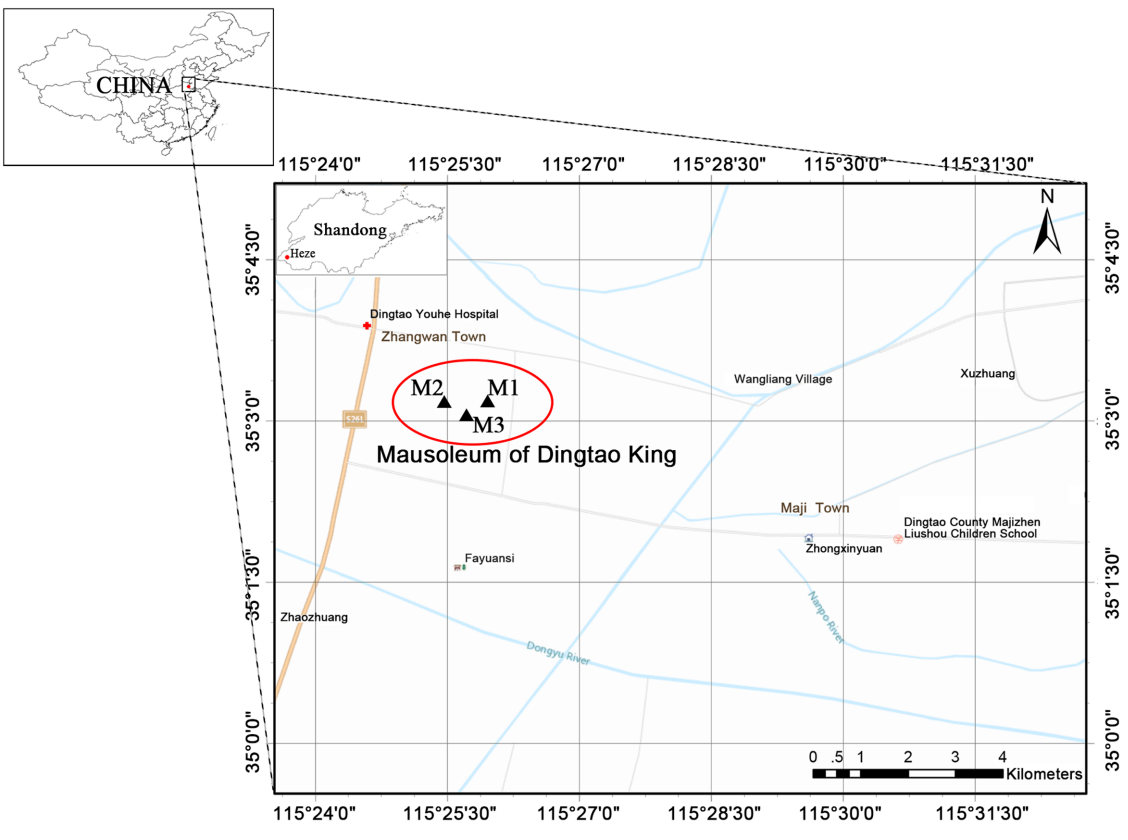


FIGURE 1 | The location of the Dingtao King's Mausoleum in Shandong Heze, China. The mausoleum consists of three tombs: M1, M2, and M3. The M1 and M3 tombs are stone-chamber tombs, while the M2 is a large-scale *huangchang ticou* tomb. The map was drawn using ArcGIS v.10.2 (www.esri.com/arcgis).

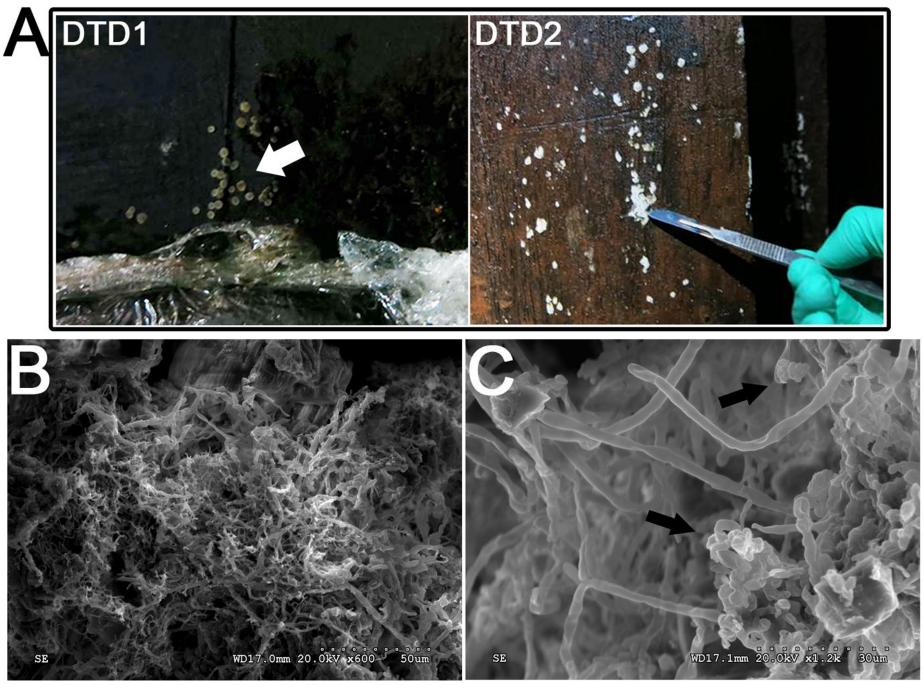


FIGURE 2 | Morphological observations of the heavy fungal colonization of the wooden tomb. **(A)** Sample DTD1, which exhibited a sporocarp appearance was used for microbial isolation and identification. Sample DTD2 was collected from white spots on the surface of the tomb. **(B,C)** Scanning electron micrograph of a white spot. Visualization of fungal hyphae and spores which appear ellipsoidal and spherical. The small dots indicate the bar.

biodeterioration of the wooden tomb due to microorganisms that can degrade lignin and cellulose must be investigated with urgency.

In order to inform the preservation of the wooden tomb M2, the aims of our study were to characterize the fungal communities of the white spots. The spots were analyzed through a combination of scanning electron microscopy (SEM) and high-throughput sequencing. Further, we isolated the predominant fungal populations and assessed their ability to degrade lignin and cellulose. Finally, the results of this study are summarized with regards to conservation recommendations and biodeterioration control for the wooden tomb.

MATERIALS AND METHODS

Sample Collection

The total surface area of the tomb is about 3,000 m^2 and the total volume of the wood that was used is ∼2,200 m^3. The primary construction material was cypress and pine. The temperature inside the tomb remains between 10 and 16°C year-round. The mound of the tomb is ∼11 m deep and is affected by Yellow River silting.

Wood samples DTD1–DTD7 were collected from different locations in the tomb and the sampling sites are shown in **Figure 2A** and **Supplementary Figure S2**. All samples were collected using minimally invasive sampling techniques with sterile scalpels and then taken to the laboratory in an ice box for subsequent analyses. Six white spot samples (DTD2–DTD7) were subdivided where one was used for SEM while the other two were used for cultivation and biodiversity analyses. DTD1 was used only for microbial isolation and identification.

Scanning Electron Microscopy

Minute samples from white spots were adhered to a conductive carbon tabs stuck on standard vacuum-clean stub and coated with gold. Gold coated samples were observed using a SEM (Hitachi S3600N). Images were obtained at magnifications between 600× and 2.5 k×, and at 20.0 kV for imaging.

DNA Extraction and PCR Amplification

Total community genomic DNA was extracted from the samples using the MoBio PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, United States) following the manufacturer's protocol. Extracted DNA was diluted to 1 ng/μL using sterile water and then stored at −80°C for subsequent analyses.

Fungal ITS region sequencing followed previously described protocols (Kozich et al., 2013). Amplification of the fungal ITS1 gene region was performed using the ITS5-1737F/ITS2-2043R primers with barcodes attached that were unique to each sample (Supplementary Table S1). All PCR reactions were carried out using the Phusion® High-Fidelity PCR Master Mix with GC Buffer (New England Biolabs, United Kingdom). Amplifications were carried out in a 30 μL mixture that included 15 μL of Master Mix (2X), a 0.5 μM final concentration of the forward and reverse

primers, 10 ng of template DNA and nuclease-free water to 30 μL. PCR conditions consisted of an initial denaturation at 98°C for 1 min, followed by 30 cycles of 98°C for 10 s, 55°C for 30 s, and elongation at 72°C for 30 s, with a final extension at 72°C for 5 min.

To visualize PCR amplification success, an equal volume of 1X loading buffer (containing SYBR green) along with PCR products were loaded on a 2% agarose gel. Samples with amplicon bands in the range of 400–450 bp were chosen for further analyses. To pool samples, PCR products were mixed in equidensity ratios. Pooled PCR products were then purified using the Qiagen Gel Extraction Kit (Qiagen, Germany).

High-Throughput Sequencing

The purified amplicons were prepared for Illumina sequencing by constructing a library using the TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, United States) following the manufacturer's recommendations. The final library concentrations and quality were checked using a Qubit@ 2.0 Fluorometer (Thermo Scientific) and an Agilent Bioanalyzer 2100 system, respectively. Lastly, 250 bp paired-end reads were generated for the library on a Hiseq2500 PE250 platform at the Novogene Bioinformatics Technology, Co., Ltd. (Beijing, China).

Bioinformatic Analyses

Paired-end reads were assigned to samples based on their unique barcodes and then trimmed of barcode and primer sequences. Paired-end reads were merged using FLASH v.1.2.7 (Magoč and Salzberg, 2011) and the resultant sequences were considered as raw tags. Quality filtering of raw tags to obtain high-quality clean tags (Bokulich et al., 2013) was performed according to the QIIME v.1.7.0 (Caporaso et al., 2010) quality control protocol. Fungal tags were then compared against the UNITE reference database using the UCHIME algorithm (Edgar et al., 2011) to detect chimeric sequences, and sequences flagged as chimeras were then removed (Haas et al., 2011). The resultant high-quality sequences were used for further analyses. OTU clustering analysis was performed using Uparse v.7.0.1001 (Edgar, 2013). Sequences with ≥97% similarity in nucleotide identity were assigned to the same OTUs. Representative sequences for each OTU were then used for further annotation. Each representative fungal sequence was assigned a taxonomic classification using the BLAST algorithm and the UNITE Database (Kõljalg et al., 2013) in QIIME. Multiple sequence alignment was then conducted using the MUSCLE software package v.3.8.31 (Edgar, 2004). OTU abundances were then normalized across samples based on the sequence count in the sample with the least amount of sequences in order to limit bias from unequal sequencing depths. Subsequent analysis of beta diversity was performed based on the normalized abundance data. Beta diversity analysis was used to evaluate differences in community composition using the unweighted Unifrac distance metric as calculated with QIIME (v.1.7.0).

The raw sequencing data generated from this study has been deposited into the NCBI Sequence Read Archive (SRA) under the accession numbers SAMN06603620–SAMN06603625.

Fungal Isolation and Identification

Fungal potato dextrose agar (PDA) medium was prepared for isolation and identification of fungi: 20% potatoes, 2% dextrose, 2% agar, and 1 L of tap water. These plates were incubated at 28°C for 5–30 days depending on the growth of fungi. Grown colonies were transferred to fresh plates containing PDA medium to obtain further pure isolates.

DNA extraction of pure fungal strains was performed using the CTAB method (Möller et al., 1992). Fungal 28S rRNA and ITS1-5.8S rRNA-ITS2 genes were amplified using the primers LR0R/LR7 and ITS1/ITS4, respectively (White et al., 1990; Kruys et al., 2015). PCR reaction mixtures consisted of a total volume of 50 μL containing 2 μL of genomic DNA, 5 μL of 10× Reaction Buffer, 4 μL of 2.5 mM dNTP mix, 2 μL of 10 μM forward primer, 2 μL of 10 μM reverse primer, 0.5 μL of 5 U/μL Transtaq-T DNA polymerase (TransGen Biotech, China), and ddH_2O to 50 μL. The PCR conditions were 95°C for 3 min, followed by 32 cycles of 30 s at 95°C, 30 s at 55°C for ITS region amplification, and 20 s at 72°C, with a final extension of 5 min at 72°C. Parameters for amplifying the 28S rRNA genes were identical except extension was conducted at 72°C for 1 min. PCR products were detected by electrophoresis in 1% agarose gels and purified using a AxyPrep PCR Clean Up Kit (Axygen, United States).

The purified PCR products were sequenced by GENEWIZ (Beijing, China) and sequence identities were analyzed using the National Center for Biotechnology Information (NCBI) BLASTn program[1] and the GenBank database. Cladograms showing evolutionary relationships were conducted using the Molecular Evolutionary Genetics Analysis software package (MEGA, v.5.05) using the neighbor-joining method. Confidence in tree topology was estimated with 1,000 bootstrap replicates.

Enzymatic Characteristics of Dominant Fungi

To test ligninolytic enzyme and cellulase activity, two different media were prepared: (i) fungal PDA plates containing 0.04% (v/v) guaiacol (ii) CMC agar medium consisting of 0.2% $NaNO_3$, 0.1% K_2HPO_4, 0.05% $MgSO_4$, 0.05% KCl, 0.2% carboxymethylcellulose (CMC) sodium salt, 0.02% peptone, 1.7% agar, and 1 L of tap water. Gram's iodine consisted of 2.0 g KI and 1.0 g iodine in 300 ml distilled water which was used to flood CMC plates for 3–5 min. All media were autoclaved for 20 min at 121°C.

Strain WY-DT1 was cultivated on PDA-guaiacol and CMC plates at 14°C for 7 days. Strain NK-DT1 was cultivated on PDA-guaiacol and CMC plates at 28°C for 4 days.

RESULTS

Microscopic Observation

To investigate microbial deterioration of tomb M2, representative samples of the white spots that were present on the surface of the *ticou* wall were observed by SEM. An abundance of mycelia

[1] https://blast.ncbi.nlm.nih.gov/Blast.cgi

was identified and fungal hyphae with a diameter of 2–3 μm were clearly discerned (**Figures 2B,C** and **Supplementary Figure S2**). The micrographs showed numerous spores on the surface of the sample, which could not be attributed to a genus due to their fragmentary state. The abundance of mycelium and spores indicated that significant fungal colonization of the tomb had occurred.

Fungal Community Analysis

High-throughput sequencing was carried out on an Illumina Hiseq2500 PE250 platform to assess the diversity and variability of the fungal communities colonizing the wooden tomb. A total of 1,457,734 fungal reads were recovered after filtering low-quality reads and chimeras. The distribution of identified and unidentified fungal phyla among the six samples is summarized in **Figure 3**. In total, five fungal phyla were detected in the six samples. Basidiomycota were the most dominant phylum in all six samples, accounting for 99.10, 99.52, 99.38, 99.71, 99.59, and 99.75% of the fungal communities, with an average relative abundance of 99.50%. Ascomycota were also present in the six samples and comprised 0.44, 0.33, 0.37, 0.23, 0.04, and 0.19% of the fungal communities with an average relative abundance of 0.27%. The other three phyla that were present, Zygomycota, Chytridiomycota, and Rozellomycota accounted for the remaining 0.23% of the total abundances. In total, there were 114 fungal genera that were detected among the samples (Supplementary Table S2), and the dominant genera were similar in all samples (**Table 1**). Among the 10 most abundant fungal genera, *Hypochnicium*, *Cortinarius,* and *Geminibasidium* were present in all samples and *Hypochnicium* was the most abundant genus across all samples, accounting for 98.61–98.45% of the community totals, with an average abundance of 99.22%. Other fungal genera only comprised the minute remainder.

Cultivation of the Dominant Fungal Populations

A filamentous fungal isolate, 'WY-DT1,' was isolated from the white spot samples. The isolate grew slowly and the color of the colonies were white to cream and hypochnoid, while hyphae diameter was ∼2–3 μm (**Supplementary Figure S3**). Analysis of the isolate's ITS sequence indicated that it displayed 99% sequence similarity with *Hypochnicium* spp. and was thus identified as *Hypochnicium* sp. WY-DT1 (KP980549, Supplementary Table S3). Phylogenetic analyses indicated that strain WY-DT1 belongs to the genus *Hypochnicium* and belongs to a distinct subclade along with *Hypochnicium bombycinum* (FN552537) and *Hypochnicium lyndoniae* (JX124704).

To confirm that the isolated strain *Hypochnicium* sp. WY-DT1 was the fungus responsible for wood deterioration within the tomb, strain WY-DT1 was inoculated onto PDA-guaiacol plates. Lignocellulolytic enzymes catalyze the oxidative polymerization of guaiacol to form reddish brown zones in the medium and wood degrading ability is directly proportional to the size and depth of the reddish brown zones (Viswanath et al., 2008). That is, the larger and deeper the reddish brown zones, the stronger the wood degrading ability. Crimson circles indicate that strain *Hypochnicium* sp. WY-DT1 can significantly degrade lignin at low temperatures. To further test whether the strain can degrade cellulose, it was inoculated on to carboxymethylcellulose (CMC) plates (**Supplementary Figure S3**). The presence of clear and distinct zones indicated that the strain has the ability to metabolize cellulose. Taken together, the cultivation results indicated that strain WY-DT1 is capable of wood biodeterioration and may be the fungus responsible for tomb deterioration.

Another fungal strain, NK-DT1, was isolated from the sample DTD1 (**Supplementary Figure S4**). Cultures were fast growing, white to grayish-white, downy, and often exhibited a broadly zonate or lobed surface appearance on PDA plates. NK-DT1 displayed 99% ITS sequence similarity to *Mortierella* spp. (Supplementary Table S3). Morphological and molecular biological results led to the determination that the strain belonged to the *Mortierella* genus and was thus designated *Mortierella* sp. NK-DT1 (KY779731). The *Mortierella* genus was detected in five samples (excluding DTD6) and also possessed high cellulolytic activity (Supplementary Table S2 and **Figure S4**).

DISCUSSION

Scanning electron microscopy revealed clear evidence for the association of filamentous fungi with the white spots. The high abundance of spores that were identified by SEM could indicate how the white spots were able to spread so quickly throughout the wooden tomb. The vast numbers of spores that were present could be an indication of the fungi's harmful contribution to the tomb's rapid deterioration.

High-throughput sequencing on the Illumina Hiseq2500 platform revealed 114 fungal genera that belonged to five phyla. The use of high-throughput sequencing can provide deeper assessments of the microbial communities that are related to the biodeterioration of cultural heritage items. For instance, using less-throughput techniques (i.e., clone-library based sequencing), only 14 fungal genera were identified in a study of wooden stairs (Piñar et al., 2016). The Basidiomycota phylum accounted for nearly all of the reads in each sample analyzed here (99.10–99.75% relative abundances of the communities), with an average relative abundance of 99.50%. Basidiomycota, Ascomycota, Zygomycota, Chytridiomycota, and Rozellomycota are often responsible for the biodeterioration of wooden relics (Irbe, 2010; Irbe et al., 2012; Palla and Barresi, 2017).

Although a number of fungal genera were detected in the samples from the tomb, the *Hypochnicium* genus was dominant in all samples, and accounted for more than 98% of the total fungal communities. *Hypochnicium* belongs to the Basidiomycota phylum and the reproductive hyphae exhibit clamp connections that are colorless and thin or exhibit thick walls. The hyphae have a mixed arrangement and produce branches at the clamp connections. There are a total of 33 currently accepted *Hypochnicium* species, with 12 species of *Hypochnicium* having been identified in China alone (Qin and He, 2013). The genus is one of the corticoid fungi which have

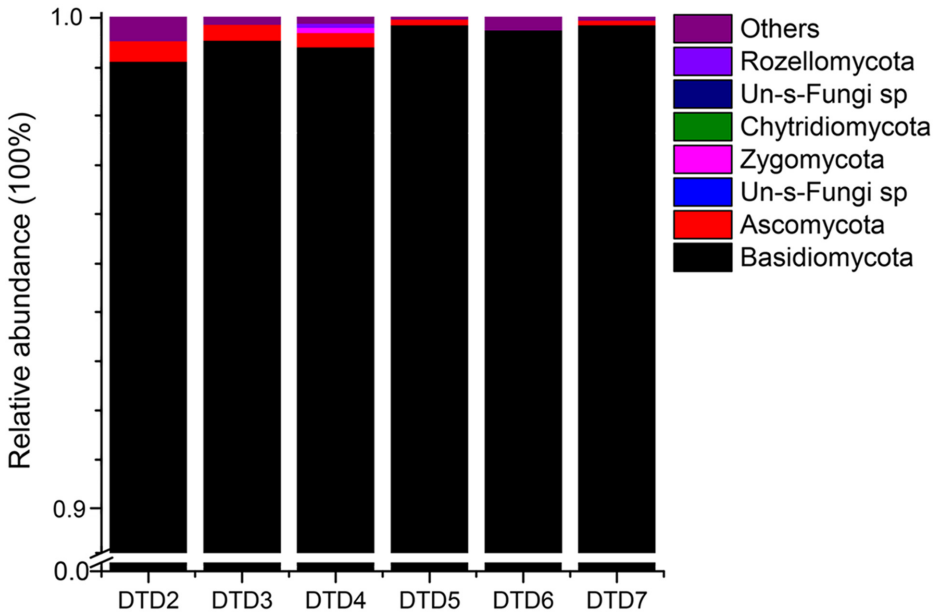


FIGURE 3 | Relative abundances of the fungal phyla among samples. Relative abundances (out of 100%) for each sample are shown based on ITS sequencing. Fungal phyla are colored according to the legend on the right.

TABLE 1 | Genus-level relative abundances and community composition of samples DTD2–7.

Genus	DTD2 (%)	DTD3 (%)	DTD4 (%)	DTD5 (%)	DTD6 (%)	DTD7 (%)
Hypochnicium	98.61	99.41	99.19	99.36	99.30	99.45
Ramariopsis	0	<0.01	0	<0.01	0.28	0
Cortinarius	0.21	0.01	0.02	0.13	<0.01	0.14
Geminibasidium	0.14	<0.01	0.01	0.14	<0.01	0.09
Eupenicillium	0.01	<0.01	0.10	<0.01	0	<0.01
Oidiodendron	0.06	0.01	0.01	0.01	0	<0.01
Trechispora	0.05	0.01	0.03	<0.01	0	<0.01
Hygrophorus	<0.01	0.01	0.04	0	0	0.01

proved to be responsible for active wood decay in buildings (Irbe et al., 2012). *Hypochnicium* can also exist in extreme polar environments such as Antarctica, which indicates that some members of the genus may have lower growth temperature requirements (Held and Blanchette, 2017). Of relevance to cultural heritage studies, *Hypochnicium* sp. is one of deteriorative agents of the Latvian Ethnographic Open Air Museum (Irbe et al., 2012).

That *Hypochnicium* can exhibit the ability to grow at low temperatures may explain why *Hypochnicium* comprised a large proportion of the fungal communities in the tomb which features temperatures between 10 and 16°C. Our cultivation analyses indicate that *Hypochnicium* sp. WY-DT1 possesses lignolytic and cellulolytic enzymes that could be responsible for the digestion of complex organic components of wood including cellulose, hemicelluloses, and lignin. It follows that the wooden structure of the tomb M2 could feasibly provide nutrients for the optimal growth and wood-degrading activity of the *Hypochnicium* sp. WY-DT1 isolate that was identified here.

Another isolate that we recovered, *Mortierella* sp. NK-DT1, belongs to the Zygomycota phylum. *Mortierella* spp. typically live as saprotrophs in soils, growing on decaying leaves and other organic material and they also possess the ability to degrade cellulose (Štursová et al., 2012). Our results indicated that *Mortierella* sp. NK-DT1 exhibited high cellulase activity. Accordingly, the fungal strain is likely to also contribute damage to the wooden structure of the tomb and must be regarded as a threat along with the *Hypochnicium* sp. WY-DT1 that was identified. In addition to the *Hypochnicium* and *Mortierella* species that were identified, other genera that comprised minor components of the fungal communities (e.g., as *Penicillium* and *Aspergillus*), and exhibit cellulolytic activities, were also present and could be potential deteriorative threats for the wooden tomb.

Fungi in the genera *Antrodia*, *Athelia*, *Gloeophyllum*, *Hyphoderma*, *Hyphodontia*, *Botryobasidium,* and *Postia* have all been associated with wood in archeological heritage materials. The majority of fungi in wooden structures were corticoid species which they caused a typical white rot. The biodeterioration

potential of these genera has been described and discussed in context of their respective threatened materials (Irbe et al., 2012). Wood from the tomb of King Midas (700 B.C., Turkey) showed soft rot decay. The extreme environmental conditions of the tomb such as low moisture and high pH inhibited white and brown rot fungi but favored soft rot. Wood from Egyptian tombs (3000–1000 B.C.) displayed degradation caused by brown rot fungi and in a few cases by soft rot fungi (Blanchette et al., 1991). Unlike these previous cases, we found a significant proportion of white rot fungi which in the case was *Hypochnicium* in the tomb. The distinctiveness of this genera's dominance in the M2 tomb suggests that the environment of the tomb (high moisture, low temperature) has led to the specific colonization and potential adaptation of these species to the tomb environment. Further analyses could indicate which adaptations have led to the prevalence of this species in this particular environment.

Cellulose and hemicellulose are normally surrounded by lignin within wood which acts as a barrier between carbohydrates and cellulolytic organisms. Fungi, including white, brown, and soft rot types, are generally known to decompose lignin and hemicellulose in lignocellulosic biomass (Sánchez, 2009). Among the white rot fungi, more than 1,500 species have the ability to degrade lignin (Tian et al., 2012). These fungi are even applied in the biological pre-treatment of biomass in the process of biofuel production (Itoh et al., 2003). In addition to biological pre-treatment, they can also be used in other bioconversion processes including wastewater treatment, bio-pulping, and the bioconversion of forest and agricultural wastes to animal feeds (Nilsson et al., 2006; Levin et al., 2007; van Kuijk et al., 2015). In the context of the tomb, the lignolytic activity of these isolates may be the main causative mechanism underpinning biodegradation, despite their beneficial application in many industrial processes.

Protective measures should be applied to control the current biodeterioration of the M2 tomb and aid conservation efforts. The application of biocides is one of the most important means to control microbial deterioration (Polo et al., 2010; Ríos et al., 2012; Diaz-Herraiz et al., 2013; Coutinho et al., 2016). However, the application of biocides can exert a selective pressure on microbial populations, and the populations may then develop biocide-resistance mechanisms, or the affected populations may be replaced by others that could be even more harmful to the artifacts (Mitchell and McNamara, 2010). One such example is in the French Lascaux Caves, where a series of biocide treatments were applied that triggered new microbial outbreaks (Martin-Sanchez et al., 2012). Thus, the application of biocides in the tomb is currently a less optimal approach and not recommended at present. Therefore, frequent cleaning and other mechanical removal procedures should be the current top priority to mitigate fungal biodeterioration.

Lastly, humidity is one of the most important factors to consider when preserving artifacts. Optimal humidity levels for white-rot and brown-rot fungi are in the range of 40–80% (Blanchette, 2000). Sterile, deionized water has been sprayed throughout the M2 tomb to maintain continuously waterlogged conditions. During this process, our observations have indicated that the white spots grew faster when spraying ceased. Thus, maintaining waterlogged conditions may be an important strategy in order to control the contamination of the tomb by the white spot fungal communities.

CONCLUSION

Our study represents the first high-throughput community sequencing and cultivation analyses of fungi associated with an ancient wooden tomb that exhibited obvious biodeterioration. SEM revealed clear evidence of filamentous fungi associated with the white spots. Although five fungal phyla including Basidiomycota, Ascomycota, Zygomycota, Chytridiomycota, and Rozellomycota were identified, Basidiomycota represented by *Hypochnicium* genus dominated the fungal communities with an average abundance of 99.22% in six samples. For the isolates obtained, we found the two species, including *Hypochnicium* sp. WY-DT1 and *Mortierella* sp. NK-DT1, possess the ability to degrade cellulose or lignin. Finally, in addition to controlling the environment as a means to inhibit microorganism deterioration activity, efficient monitoring and protective measures should be applied to mitigate current fungal biodeterioration. Although the investigation provided a basic understanding on the dominant members of the fungal community that are likely to be responsible for contamination and potential deterioration of the M2 tomb, the reason behind the *Hypochnicium* contamination is yet unclear. Additional research is required in order to better understand the adaptations and ecological traits that allowed the genus to dominate the M2 tomb-contaminating communities.

AUTHOR CONTRIBUTIONS

All authors contributed to this work. JP conceived and planned the research. ZL performed statistical analyses and wrote the manuscript. YW and QL extracted DNA, cultured and tested the isolates. XP and QG organized the sampling trips and provided access to the sampling sites. TF and CH assessed lignocellulose activity. QM and XZ edited the manuscript.

FUNDING

This work as supported by the Project of Microorganisms Monitoring and Prevention of the Mausoleum of the Dingtao King (2015G003-2).

ACKNOWLEDGMENT

The authors are grateful to the conservators of the Bureau of Cultural Relics of Shandong and the Bureau of Cultural Relics of the Dingtao District.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01633/full#supplementary-material

FIGURE S1 | The structure of the tomb M2. **(A)** Cross-section drawing of the tomb. **(B)** The plan of the tomb. S1–S12 are 12 outer storage chambers. HB, HN, QB, QN, ZB1, ZB2, ZN1, and ZN2 are eight side chambers. WD, SD, and ND refer to west doorway, south doorway and north doorway, respectively. FC, front chamber; CC, central chamber; RC, rear chamber. Red stars indicate the sampling positions.

FIGURE S2 | White spots on the *ticou* wall of the tomb. **(A)** The samples with the white spots (DTD3–DTD7) were collected for cultivation analyses and high-throughput sequencing in March 2015. **(B,C)** Scanning electron micrograph of white spots.

FIGURE S3 | Colony morphological appearances of strain *Hypochnicium* sp. WY-DT1 on different media. Colony **(A)** and micro-morphology features **(B)** of the culprit fungus at 100× magnification. The white bar represents 10 μm. **(C)** Neighbor-joining phylogenetic tree based on the ITS sequence (approximately 640 bp) of strain WY-DT1, including representatives of the most closely related strains and additional members of the genus *Hypochnicium*. Bootstrap values are given at the nodes as a percentage of 1,000 bootstrap replicates. **(D)** The culprit deteriorative fungus was cultivated on PDA-guaiacol plates. The crimson circle indicated that the fungus can degrade lignin significantly at low temperatures. **(E)** Effect of Gram's iodine flooding on cellulolytic zone in CMC plates. The transparent circle indicates that the fungus can produce cellulose.

FIGURE S4 | Colony morphological appearances of strain *Mortierella* sp. NK-DT1 on different media. **(A,B)** Colony features of NK-DT1. **(C)** Strain NK-DT1 was cultivated on PDA-guaiacol plates. Results showed the fungus have no ability to degrade lignin. **(D)** The transparent circle indicates that NK-DT1 can also produce cellulose.

REFERENCES

Blanchette, R. A. (2000). A review of microbial deterioration found in archaeological wood from different environments. *Int. Biodeterior. Biodegrad.* 46, 189–204. doi: 10.1016/S0964-8305(00)00077-9

Blanchette, R. A., Cease, K. R., Abad, A. R., Sams, G. K., Simpson, E., and Koestler, R. J. (1991). "Biodeterioration of archaeological wood," in *Proceedings of the 8th International Biodeterioration and Biodegradation Symposium: Biodeterioration and biodegradation 8,Windsor, Ontario, Canada, 26-31 August 1990*, ed. H. W. Rossmoore (Amsterdam: Elsevier), 385–388.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Coutinho, M. L., Miller, A. Z., Martin-Sanchez, P. M., Mirão, J., Gomez-Bolea, A., Machado-Moreira, B., et al. (2016). A multiproxy approach to evaluate biocidal treatments on biodeteriorated majolica glazed tiles. *Environ. Microbiol.* 18, 4794–4816. doi: 10.1111/1462-2920.13380

Cui, S., Cai, Y., Wu, S., Jia, C., and Ding, X. (2013). The Han tomb near Lingsheng Lake in Dingtao County, Shandong. *Chin. Archaeol.* 13, 19–27. doi: 10.1515/char-2013-0003

Diaz-Herraiz, M., Jurado, V., Cuezva, S., Laiz, L., Pallecchi, P., Tiano, P., et al. (2013). The actinobacterial colonization of Etruscan paintings. *Sci. Rep.* 3:1440. doi: 10.1038/srep01440

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Gutarowska, B., Celikkol-Aydin, S., Bonifay, V., Otlewska, A., Aydin, E., Oldham, A. L., et al. (2015). Metabolomic and high-throughput sequencing analysis—modern approach for the assessment of biodeterioration of materials from historic buildings. *Front. Microbiol.* 6:979. doi: 10.3389/fmicb.2015.00979

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi: 10.1101/gr.112730.110

Held, B. W., and Blanchette, R. A. (2017). Deception Island, Antarctica harbors a diverse assemblage of wood decay fungi. *Fungal Biol.* 121, 145–157. doi: 10.1016/j.funbio.2016.11.009

Huang, Z. (1998). On princes'tombs of the han period. *Acta Archaeol. Sin.* 1, 11–34.

Irbe, I. (2010). "Wood decay fungi in Latvian buildings including cultural monuments," in *Proceedings of the International Conference held by COST Action IE0601, Braga, November 5–7, 2008*, ed. J. Gril (Firenze: Firenze University Press), 94–100.

Irbe, I., Karadelev, M., Andersone, I., and Andersons, B. (2012). Biodeterioration of external wooden structures of the Latvian cultural heritage. *J. Cult. Herit.* 13, S79–S84. doi: 10.1016/j.culher.2012.01.016

Itoh, H., Wada, M., Honda, Y., Kuwahara, M., and Watanabe, T. (2003). Bioorganosolve pretreatments for simultaneous saccharification and fermentation of beech wood by ethanolysis and white rot fungi. *J. Biotechnol.* 103, 273–280. doi: 10.1016/S0168-1656(03)00123-8

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. doi: 10.1111/mec.12481

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi: 10.1128/AEM.0104313

Kruys, Å, Huhndorf, S. M., and Miller, A. N. (2015). Coprophilous contributions to the phylogeny of Lasiosphaeriaceae and allied taxa within Sordariales (Ascomycota, Fungi). *Fungal Divers.* 70, 101–113. doi: 10.1007/s13225-014-02963

Levin, L., Villalba, L., Da Re, V., Forchiassin, F., and Papinutti, L. (2007). Comparative studies of loblolly pine biodegradation and enzyme production by Argentinean white rot fungi focused on biopulping processes. *Process Biochem.* 42, 995–1002. doi: 10.1016/j.procbio.2007.03.008

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Martin-Sanchez, P. M., Nováková, A., Bastian, F., Alabouvette, C., and Saiz-Jimenez, C. (2012). Use of biocides for the control of fungal outbreaks in subterranean environments: the case of the Lascaux Cave in France. *Environ. Sci. Technol.* 46, 3762–3770. doi: 10.1021/es2040625

Mitchell, R., and McNamara, C. J. (2010). *Cultural Heritage Microbiology : Fundamental Studies in Conservation Science*. Washington, DC: ASM Press.

Möller, E. M., Bahnweg, G., Sandermann, H., and Geiger, H. H. (1992). A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res.* 20, 6115–6116. doi: 10.1093/nar/20.22.6115

Nilsson, I., Möller, A., Mattiasson, B., Rubindamayugi, M. S. T., and Welander, U. (2006). Decolorization of synthetic and real textile wastewater by the use of white-rot fungi. *Enzyme Microb. Technol.* 38, 94–100. doi: 10.1016/j.enzmictec.2005.04.020

Ortiz, R., Navarrete, H., Navarrete, J., Párraga, M., Carrasco, I., de la Vegac, E., et al. (2014). Deterioration, decay and identification of fungi isolated from wooden structures at the Humberstone and Santa Laura saltpeter works: a world heritage site in Chile. *Int. Biodeterior. Biodegrad.* 86, 309–316. doi: 10.1016/j.ibiod.2013.10.002

Palla, F., and Barresi, G. (2017). *Biotechnology and Conservation of Cultural Heritage*. Berlin: Springer. doi: 10.1007/978-3-319-46168-7

Piñar, G., Dalnodar, D., Voitl, C., Reschreiter, H., and Sterflinger, K. (2016). Biodeterioration risk threatens the 3100 year old staircase of hallstatt (Austria): possible involvement of halophilic microorganisms. *PLoS ONE* 11:e0148279. doi: 10.1371/journal.pone.0148279

Polo, A., Cappitelli, F., Brusetti, L., Principi, P., Villa, F., Giacomucci, L., et al. (2010). Feasibility of removing surface deposits on stone using biological and chemical remediation methods. *Microb. Ecol.* 60, 1–14. doi: 10.1007/s00248-009-96336

Qin, W. M., and He, S. H. (2013). Three species of corticioid fungi (Basidiomycota) new to China. *Mycosystema* 32, 297–301.

Ríos, A. D. L., Pérez-Ortega, S., Wierzchos, J., and Ascaso, C. (2012). Differential effects of biocide treatments on saxicolous communities: case study of the Segovia cathedral cloister (Spain). *Int. Biodeterior. Biodegrad.* 67, 64–72. doi: 10.1016/j.ibiod.2011.10.010

Sánchez, C. (2009). Lignocellulosic residues: biodegradation and bioconversion by fungi. *Biotechnol. Adv.* 27, 185–194. doi: 10.1016/j.biotechadv.2008.11.001

Štursová, M., Žifčáková, L., Leigh, M. B., Burgess, R., and Baldrian, P. (2012). Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiol. Ecol.* 80, 735–746. doi: 10.1111/j.1574-6941.2012.01343.x

Tian, X. F., Fang, Z., and Guo, F. (2012). Impact and prospective of fungal pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Biofuels Bioprod. Bioref.* 6, 335–350. doi: 10.1002/bbb

van Kuijk, S. J. A., Sonnenberg, A. S. M., Baars, J. J. P., Hendriks, W. H., and Cone, J. W. (2015). Fungal treated lignocellulosic biomass as ruminant feed ingredient: a review. *Biotechnol. Adv.* 33, 191–202. doi: 10.1016/j.biotechadv.2014.10.014

Viswanath, B., Chandra, M. S., Pallavi, H., and Reddy, B. R. (2008). Screening and assessment of laccase producing fungi isolated from different environmental samples. *Afr. J. Biotechnol.* 7, 1129–1133.

White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *PCR Protocols: A Guide to Methods and Applications*, eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (New York, NY: Academic Press), 315–322.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Liu, Wang, Pan, Ge, Ma, Li, Fu, Hu, Zhu and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 28 February 2017
doi: 10.3389/fbioe.2017.00014

PCR Primer Design for 16S rRNAs for Experimental Horizontal Gene Transfer Test in *Escherichia coli*

Kentaro Miyazaki[1,2], Mitsuharu Sato[1,2] and Miyuki Tsukuda[1,2]*

[1] *Department of Life Science and Biotechnology, Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan,* [2] *Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan*

We recently demonstrated that the *Escherichia coli* ribosome is robust enough to accommodate foreign 16S rRNAs from diverse gamma- and betaproteobacteria bacteria (Kitahara et al., 2012). Therein, we used the common universal primers Bac8f and UN1541r to obtain a nearly full-length gene. However, we noticed that these primers overlap variable sites at 19[A/C] and 1527[U/C] in Bac8f and UN1541r, respectively, and thus, the amplicon could contain mutations. This is problematic, particularly for the former site, because the 19th nucleotide pairs with the 916th nucleotide, which is a part of the "central pseudoknot" and is critical for function. Therefore, we mutationally investigated the role of the base pair using several 16S rRNAs from gamma- and betaproteobacteria. We found that both the native base pairs (gammaproteobacterial 19A–916U and betaproteobacterial 19C–916G) and the non-native 19A–916G pair retained function, whereas the non-native 19C–916U was defective 16S rRNAs. We next designed a new primer set, Bac1f and UN1542r, so that they do not overlap the potential mismatch sites. 16S rRNA amplicons obtained from the environmental metagenome using the new primer set were dominated by proteobacterial species (~85%). Subsequent functional screening identified various 16S rRNAs from proteobacteria, all of which contained native 19A–916U or 19C–916G base pairs. The primers developed in this study are thus advantageous for functional characterization of foreign 16S rRNA in *E. coli* with no artifacts.

Keywords: bacterial phylogeny, 16S rRNA, ribosome, horizontal gene transfer, molecular clock, functional complementation, metagenome, central pseudoknot

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, National Scientific Council (CONICET), Argentina

Reviewed by:
Kei Kitahara, Hokkaido University, Japan
Trevor Carlos Charles, University of Waterloo, Canada

****Correspondence:***
Kentaro Miyazaki
miyazaki-kentaro@aist.go.jp

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *19 December 2016*
Accepted: *09 February 2017*
Published: *28 February 2017*

Citation:
Miyazaki K, Sato M and Tsukuda M (2017) PCR Primer Design for 16S rRNAs for Experimental Horizontal Gene Transfer Test in Escherichia coli. Front. Bioeng. Biotechnol. 5:14. doi: 10.3389/fbioe.2017.00014

INTRODUCTION

The bacterial ribosome consists of 3 rRNA molecules and 54 proteins and plays a crucial role in translating mRNA-encoded information into proteins. Because of the structural complexity of the ribosome (Schuwirth et al., 2005), it is believed that each ribosomal component coevolves to maintain function (Jain et al., 1999). In particular, because the 16S and 23S rRNAs form the structural core of the ribosome (Schuwirth et al., 2005), they are believed to be least likely to experience horizontal gene transfer between species (Jain et al., 1999). On the basis of the species-specific nature of rRNA and their omnipresence in all bacteria, the rRNA genes, especially those for 16S rRNA, have

long been used as an "ultimate chronometer" (Woese, 1987) for phylogenetic classification of bacterial species (Lane et al., 1985; Woese, 1987).

Despite the apparent species-specific nature of 16S rRNAs, we recently found that the *Escherichia coli* ribosome is able to accommodate foreign 16S rRNA (Kitahara et al., 2012). Namely, using *E. coli* Δ7, a null mutant of the *rrn* (ribosomal RNA) operon, as a host strain, we have shown that various 16S rRNA genes, including those from a different phylogenetic class (i.e., betaproteobacteria), were able to complement growth. The lowest identity of functional 16S rRNA gene to that of *E. coli* was as low as 80%, implying that hundreds of simultaneous nucleotide changes are permitted in the maintenance of ribosome function. The basis for this high mutability is the conservation of the RNA secondary structures, which is consistent with a previous finding that 16S rRNA is typically recognized by ribosomal proteins via salt bridges between phosphate oxygen atoms of the RNA backbone, but nucleotide bases are not strictly discriminated (Brodersen et al., 2002). Furthermore, insertion/deletion is allowed in some RNA helices (e.g., h6, 10, and 17) that are not involved in protein binding. Understanding the sequence and structural variations of 16S rRNA that are accommodated in the *E. coli* ribosome should be helpful for our understanding of the evolution of rRNA and the sequence–structure–function relationships of the ribosome.

In our previous study, to PCR amplify foreign 16S rRNA genes, we used Bac8f(A) or Bac8f(C) [the most commonly used "Bac8f" (Eden et al., 1991)] for the forward primer and UN1541r(U) or UN1541r(C) for the reverse primer (**Figure 1**; oligonucleotide sequences summarized in **Table 1**) (Kitahara et al., 2012), which allowed amplification of a nearly full-length gene. These primers can cover the majority of bacterial 16S rRNA genes and thus are commonly used for phylogenetic and/or community analysis (Lane et al., 1985; Weisburg et al., 1991; Amann et al., 1995). However, we noticed that the amplicons obtained using the primer set contained mutations at certain frequencies, which could affect the functionality of *in vivo*-reconstituted mutant ribosomes. In *E. coli* 16S rRNA, nucleotides 17–19 pair with nucleotides 916–918 to form a short helix (h2) (**Figure 2**). The helix is involved in the formation of the "central pseudoknot," whose structure is highly conserved in both prokaryotes and eukaryotes. This unique structure is essential for translational initiation and is highly susceptible to point mutations (Brink et al., 1993; Dammel and Noller, 1993; Poot et al., 1998). Despite this structural conservation, however, the 19th nucleotide varies depending on the species, 19A or 19C, which pairs with 916U or 916G, respectively (**Figure 2**). Thus, if Bac8f(A) or Bac8f(C) is used as a primer, there is a possibility of generating a mismatch between the 19th and 916th nucleotides. Similarly, the 1,527th position is also variable (C or U) and can generate a mismatch in the amplicons (**Figure 1B**), although this site may not be involved in function. Thus, in our specific system for functional investigation of 16S rRNAs, it is essential to develop a new primer set to avoid the introduction of artificial mutations. In addition, we need to take the RNA processing issue into consideration. For proper processing of a precursor transcript into mature rRNAs (16S, 23S, and 5S rRNAs), the processing sites (i.e., RNase cleavage sites) need to be similar to the *E. coli* sequences (Gutgsell and Jain, 2012).

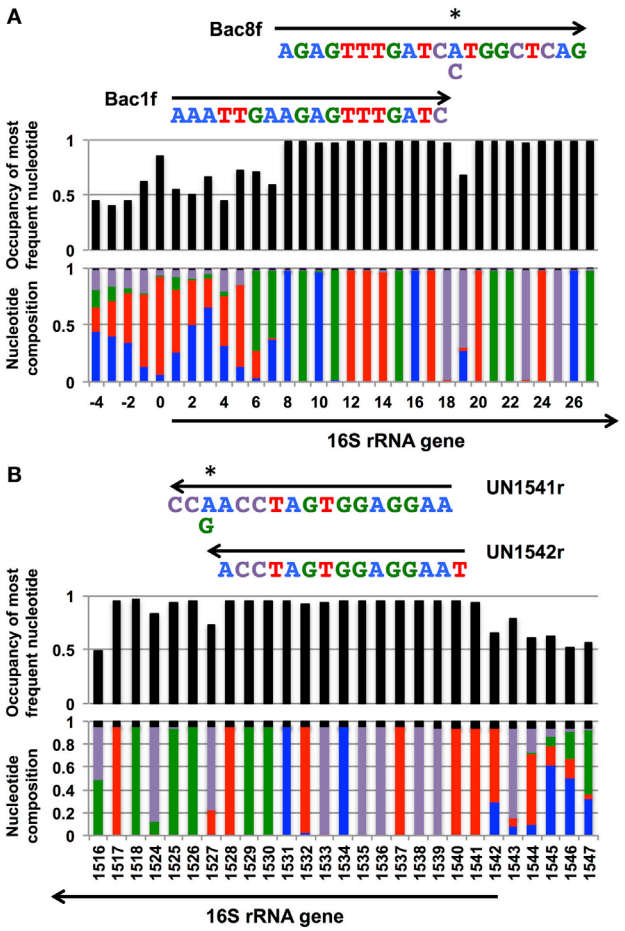


FIGURE 1 | Coverage rates and nucleotide composition at the (A) 5′- and (B) and 3′-ends of the bacterial 16S rRNA gene. The annealing regions for Bac1f, Bac8f, and UN1541r are shown by arrows with the sequences. Wobble nucleotides are marked by asterisks. Colors: blue, A; red, T; green, G; purple, C; black, gap.

Taking these points into consideration, we first evaluated the effects of mismatches between the 19th and 916th nucleotides. In addition to the native 19A–916U and 19C–916G pairs, the non-native 19A–916G pair retained function, whereas the non-native 19C–916U was detrimental. Second, we designed new primers, Bac1f and UN1542r, which encompassed nucleotide positions 1–18 for Bac1f and 1542–1528 for UN1542r, so that they did not overlap the potential mismatch sites. These primers were used for PCR amplification of 16S rRNA genes and the resultant library was functionally screened. DNA sequencing of the 16S rRNA genes in the functional clones confirmed the absence of unwanted mismatches in the functional genes.

TABLE 1 | PCR primers used for amplification of 16S rRNA genes.

Primer	Sequence	Reference
Bac1f	5′-AAATTGAAGAGTTTGATC-3′	This study
Bac8f(A)[a]	5′-AGAGTTTGATCATGGCTCAG-3′	Kitahara and Miyazaki (2011); Kitahara et al. (2012); Weisburg et al. (1991)
Bac8f(C)[a]	5′-AGAGTTTGATCCTGGCTCAG-3′	Kitahara and Miyazaki (2011); Kitahara et al. (2012); Weisburg et al. (1991)
UN1542r	5′-TAAGGAGGTGATCCA-3′	This study
UN1541r(U)[a]	5′-AAGGAGGTGATCCAACC-3′	Kitahara and Miyazaki (2011); Kitahara et al. (2012)
UN1541r(C)[a]	5′-AAGGAGGTGATCCAGCC-3′	Kitahara and Miyazaki (2011); Kitahara et al. (2012); Weisburg et al. (1991)
Bac1r	5′-GATCAAACTCTTCAATTTAAAAGTTTGACGCTCAAAG-3′	This study
Bac8r(A)	5′-CTGAGCCATGATCAAACTCTTC-3′	Kitahara and Miyazaki (2011); Kitahara et al. (2012)
Bac8r(C)	5′-CTGAGCCAGGATCAAACTCTTC-3′	This study
UN1542f	5′-TGGATCACCTCCTTACCTTAAAGAAGCGT-3′	This study

[a]*Underlined nucleotides are the sites that can generate a mismatch.*

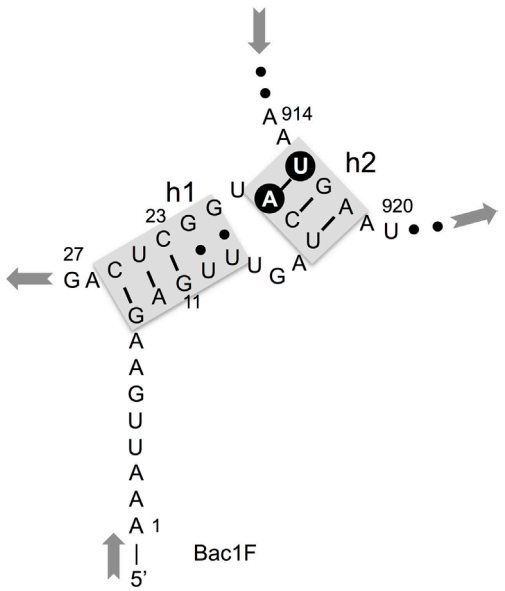


FIGURE 2 | Central pseudoknot structure in *Escherichia coli* 16S rRNA. Nucleotide numbers 19 and 916 are labeled in white letters in a black circle.

MATERIALS AND METHODS

Reagents

KOD-Plus-Neo and KOD FX-Neo DNA polymerases were purchased from Toyobo (Osaka, Japan). Trimethoprim (Tmp), ampicillin (Amp), kanamycin (Km), and sucrose (Suc) were purchased from Wako Pure Chemicals (Tokyo, Japan). Zeocin™ (Zeo) and the In-Fusion Cloning Kit were purchased from Invitrogen (Carlsbad, CA, USA). Lennox LB medium [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl] was purchased from Merck (Tokyo, Japan). The Extrap Soil DNA Kit Plus ver.2 was purchased from J-Bio21 (Tsukuba, Japan). FastDNA Kit was purchased from BIO101 (La Jolla, CA, USA). Oligonucleotide primers (**Table 1**) were purchased from Sigma (Hokkaido, Japan).

Bacterial Strains and Culture Conditions

The following bacterial strains were purchased from the Biological Resource Center (NBRC), National Institute of Technology and Evaluation, Japan: *Serratia ficaria* (NBRC 102596), *Caldimonas manganoxidans* (NBRC 16448), *Hydrogenophaga flava* (NBRC 102514), *Hydrogenophilus thermoluteolus* (NBRC 14978), *Oxalicibacterium horti* (NBRC 13594), *Oligella urethralis* (NBRC 14589), and *Ralstonia pickettii* (NBRC 102503). *Burkholderia sacchari* was a laboratory stock. Competent *E. coli* JM109 cells were purchased from RBC Bioscience (Taipei, Taiwan). Antibiotics were added when necessary at the following concentrations: Tmp, 10 μg/ml; Amp, 100 μg/ml; Km, 25 μg/ml; and Zeo, 50 μg/ml. Agar (1.5% [w/v]) was added to solidify the medium. Suc was added at 5% (w/v) for counterselection purposes when necessary.

Genomic DNA Purification, PCR Amplification of 16S rRNA Genes, and Library Construction

Genomic DNA of bacterial isolates was purified using the FastDNA Kit. The *Nitrosomonas europaea* was a gift from Dr. Naohiro Noda (AIST, Japan). Environmental metagenomic DNA (soils, fermented products, and seawater) was purified using the Extrap Soil DNA Kit Plus ver. 2. The 16S rRNA genes were amplified by PCR using a set of primers, Bac1f, Bac8f(A), or Bac8f(C) and UN1542r. The reaction mixture contained 100 ng of template DNA (bacterial genome or environmental metagenome), 1× PCR buffer, 0.4 mM each of dNTPs, 0.25 μM each of primers and 1 U of KOD FX-Neo DNA polymerase in a total volume of 50 μl. The mixture was heated at 94°C for 2 min and subjected to 30 rounds of thermal cycling at 98°C for 10 s, 48°C for 30 s for Bac1f or 57°C for 30 s for Bac8f(A) and Bac8f(C), and 68°C for 1.5 min and final incubation at 68°C for 5 min. The amplicon was separated by agarose gel (0.8% [w/v]) electrophoresis; a single band was excised from the gel, purified and dissolved in 30 μl of water.

An expression vector for the 16S rRNA gene was modified from pRB103 (Kitahara and Suzuki, 2009; Kitahara et al., 2012) by deleting the genes for tRNA, 23S rRNA, and 5S rRNA, replacing the antibiotic selection marker from Zeo to Tmp and replication origin from pSC101 to p15A. The resultant plasmid was named pMS205aTp1 (map illustrated in Figure S1A in Supplementary Material). The entire vector (without the 16S rRNA gene) was PCR-amplified using the primer set Bac1R, Bac8r(A), or Bac8r(C) and UN1542f. The PCR mixture contained 1× PCR buffer, 0.2 mM each of dNTPs, 1.5 mM $MgSO_4$, 0.25 μM each of primers, 10 ng of pMS205aTp1, and 1 U of KOD-Neo-DNA

polymerase in a total volume of 50 µl. The mixture was heated at 94°C for 2 min and subjected to 25 cycles at 94°C for 10 s, 60°C for 30 s, and 68°C for 2.5 min, followed by a final incubation at 68°C for 5 min. The products were treated with *Dpn*I (10 U, 37°C, 6 h), gel-purified and dissolved in 30 µl of water.

The 16S rRNA gene (*ca.*, 200 ng) and the linearized pMS205aTp1 (*ca.*, 200 ng) fragments were combined and ligated using the In-Fusion Cloning Kit in a total volume of 10 µl. After incubation at 50°C for 1 h, the reaction products (2 µl) were introduced into competent *E. coli* JM109 cells (100 µl) and grown on LB/Tmp agar plates at 37°C overnight. Some of the colonies were randomly picked and used for sequence determination for phylogenetic analysis. Rest of the colonies were combined and plasmids were extracted to yield a library.

Functional Screening of 16S rRNA Genes

Escherichia coli MY201 *rna*$^-$ is a derivative of *E. coli* Δ7 *rna*$^-$ (Kitahara and Suzuki, 2009), which contains the growth rescue plasmid pMY201 (modified from pRB101 by substituting the pSC101 ori to p15A ori, Figure S1B in Supplementary Material) and pML103Δ (expression plasmid for 23S rRNA, 5S rRNA, and tRNAs, created by deleting the 16S rRNA gene from pRB103, map illustrated in Figure S1C in Supplementary Material). Competent MY201 cells were transformed with a pMS205aTp1 library and grown on LB/Km/Zeo/Tmp agar plates at 37°C for overnight. Colonies were collected, mixed in 1 ml of LB broth, vigorously vortexed, appropriately diluted, and spread over LB/Km/Zeo/Tmp/Suc agar plates. Some of the colonies were randomly picked and used for sequence determination for phylogenetic analysis. Rest of the colonies grown on the plates were collected and used for further studies.

Growth Assay

Mutant *E. coli* strains were grown in 1 ml of LB/Km/Zeo/Tmp/Suc broth in a 96 deep-well plate. The plate was incubated at 37°C with vigorous agitation (1,200 rpm) in an MBR-024 microplate shaker (Taitec, Saitama, Japan). After 14 h, 1 µl of the culture was transferred to a fresh LB/Km/Zeo/Tmp/Suc broth (1 ml) in 96-well plate and grown at various temperatures (30, 37, or 42°C) with vigorous agitation (1,200 rpm). After 14 h, 200 µl of the culture was transferred to a 96-well plate and OD_{600} was measured.

DNA Sequencing and BLAST Search

DNA sequencing was carried out using the Sanger method with an Applied Biosystems (Foster City, CA, USA) automatic DNA sequencer (ABI PRISM 3130xl Genetic Analyzer) and an Applied Biosystems BigDye (ver. 3.1) kit. Blast search (Altschul et al., 1990) was carried out using the NCBI nucleotide database "16S rRNA sequences (Bacteria and Archaea)" with the program selection optimized for "Highly similar sequences (megablast)."

Dataset and Sequence Alignment of 16S rRNA Genes

All 16S rRNA gene sequences (plus 50 additional nucleotides at the 5′ and 3′ ends) were retrieved from the genomic sequences in the NCBI database (as of August 2014) (**Table 2**). A total of 9,624 genes were identified in 2,476 genomes of 23 phyla. Multiple sequence alignment of these genes was performed using the MAFFT v7 program (Katoh and Standley, 2013).

TABLE 2 | List of 16S rRNA genes retrieved from the NCBI database.[a]

Phylum or group	Number of genomes	Number of 16S rRNA genes
Actinobacteria	268	833
Aquificae	13	27
Bacteroidetes–Chlorobi group	99	287
Caldiserica	1	1
Chlamydiae–Verrucomicrobia group	112	200
Chloroflexi	20	34
Chrysiogenetes	1	3
Deferribacteres	4	7
Cyanobacteria	75	168
Dictyoglomi	2	4
Deinococcus–Thermus	22	51
Elusimicrobia	1	1
Fibrobacteres–Acidobacteria group	7	12
Firmicutes	532	3,304
Fusobacteria	9	38
Nitrospirae	4	7
Planctomycetes	7	18
Proteobacteria	1,144	4,369
Spirochaetes	58	106
Synergistetes	4	11
Tenericutes	74	111
Thermodesulfobacteria	2	3
Thermotogae	17	29
Total	2,476	9,624

[a]*All 16S rRNA gene sequences were extracted from all available genomic sequences in NCBI database on August 1, 2014.*

Nucleotide Sequence Accession Numbers

The nucleotide sequences for 16S rRNA gene have been deposited in GenBank/EMBL/DDBJ under the accession numbers LC213146–LC213207, LC213207–LC213252, and LC213253–LC213296.

RESULTS AND DISCUSSION

Primer Design

Nucleotide composition around the 5′- and 3′-end regions of all bacterial 16S rRNA genes (**Table 2**) is shown in **Figure 1**. For the 5′-end, the sequence surrounding the 19th nucleotide, particularly from the 8th to 27th, is highly conserved (**Figure 1A**), which corresponds to the Bac8f primer-binding site. The Bac8f primer covers 97% of bacterial 16S rRNA sequences (**Figure 1A**), confirming the appropriateness of the primer for phylogenetic/community analysis (Lane et al., 1985; Amann et al., 1995). However, due to the presence of a potential mismatch site at the 19th nucleotide position, this is not appropriate for our specific purpose (i.e., functional analysis), and thus, we designed a new primer Bac1f, which encompasses the 1st to 18th nucleotides. Although the very beginning of the sequence (from first to seventh nucleotides) is highly variable among all bacteria (**Figure 1A**), the region is critical for RNA processing (Gutgsell and Jain, 2012), so we strictly followed the *E. coli* sequence for this site.

We next checked the coverage rate of the Bac1f primer for each phylum (Figure S2 in Supplementary Material). As described above, the 5′ end of the 16S rRNA sequence varies among bacteria (**Figure 1A**). Nevertheless, Bac1f showed relatively high specificity to some bacterial 16S rRNAs that included Bacteroidetes–Chlorobi (Figure S2B in Supplementary Material), Chlamydiae-Verrucomicrobia (Figure S2C in Supplementary Material), and Proteobacteria (Figure S2G in Supplementary Material). In our previous study (Kitahara et al., 2012), no functional 16S rRNAs were obtained from phyla other than proteobacteria. Thus, in practice, although the Bac1f primer has some bias to specific phylogenetic groups, this bias is advantageous to enriching libraries with a potentially functional fraction and to reducing background.

Figure S3 in Supplementary Material summarizes the coverage rate of the Bac1f primer for each class of proteobacteria. Overall, there is a slight preference for alpha-, beta-, and gamma-classes of proteobacteria, and the delta-epsilon class has a larger number of potential mismatches.

Effects of Non-Natural Base Pairing between the 19th and 916th Nucleotides on Ribosomal Activity

To investigate how non-natural base pairing between the 19th and 916th nucleotides affects ribosomal activity, we used 16S rRNA genes from the following bacteria: gammaproteobacterial *E. coli* (Eco) and *S. ficaria* (Sfi) and betaproteobacterial *B. sacchari* (Bsa), *C. manganoxidans* (Cma), *H. flava* (Hfl), *H. thermoluteolus* (Hth), *N. europaea* (Neu), *O. horti* (Oho), *O. urethralis* (Our), and *R. pickettii* (Rpi).

PCR amplification was carried out using three types of forward primers: Bac1f, Bac8f(A), or Bac8f(C); UN1542r was used as a common reverse primer. A specifically amplified fragment was then cloned back into pMS205aTp1. After confirming the sequence of the entire 16S rRNA gene, the resultant plasmid was transferred into *E. coli* MY201 *rna*$^{-}$ (Δ7 strain). After *sacB*-based counterselection to eliminate rescue plasmids expressing *E. coli* 16S rRNA, all the clones were successfully obtained at 37°C, implying that mutation in the 19th–916th base pair is more or less permissive under this condition.

We next examined the growth properties of each mutant at various temperatures (30°C, 37°C, and 42°C). **Figure 3A** illustrates OD_{600} after growth for 14 h. In general, for the native base pairs (19A–916U for gammaproteobacterial and 19C–916G for betaproteobacterial 16S rRNAs) all clones had higher OD_{600} at 37°C than at 30°C. For the clones carrying non-native pairs (**Figure 3B**, 19C–916U for gammaproteobacterial and 19A–916G for betaproteobacterial 16S rRNAs), no growth perturbation was observed for betaproteobacterial clones and they appeared to gain a broadened temperature optimum; final OD_{600} shifted upward at 30°C. In contrast, gammaproteobacterial clones (Eco and Sfi) showed greatly reduced OD_{600} values at all temperatures. Virtually no growth was observed at 30°C, indicating the appearance of a cold-sensitive (or heat-tolerant) phenotype.

Poot et al. (1998) have analyzed the role of h2 through mutagenesis. Using *E. coli* 16S rRNA as a template, they introduced a point mutation to alter the native 19A–916U base pair to 19A–916G, 19C–916G, and 19C–916U. They used the mutant ribosome in an *in vitro* translational assay (at 42°C) and found that the former two mutants retained nearly full activity (>80%), whereas 19C–916U had much reduced (30%) activity. Although the assay systems are different, the general conclusion of both their study and ours is that 19C–916U is defective.

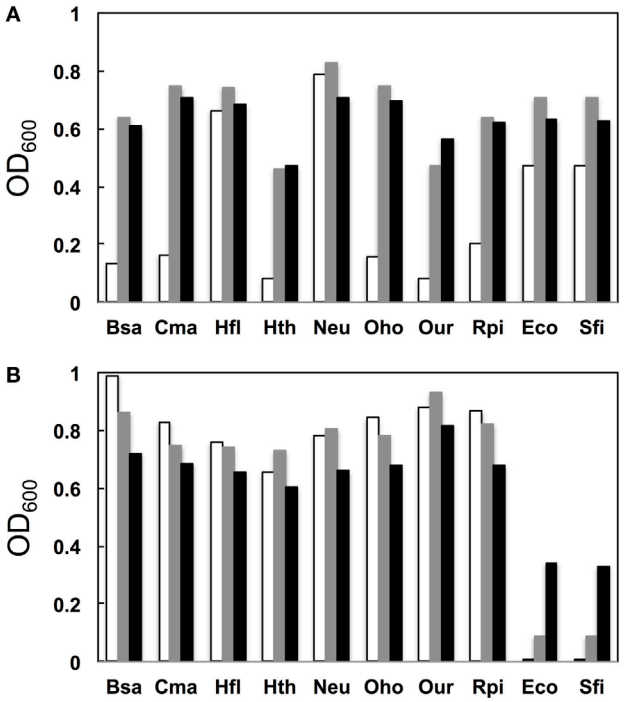


FIGURE 3 | Growth profile. **(A)** Native base pairs between nucleotides 19 and 916 [19C–916G for betaproteobacterial *Burkholderia sacchari* (Bsa), *Caldimonas manganoxidans* (Cma), *Hydrogenophaga flava* (Hfl), *Hydrogenophilus thermoluteolus* (Hth), *Nitrosomonas europaea* (Neu), *Oxalicibacterium horti* (Oho), *Oligella urethralis* (Our), and *Ralstonia pickettii* (Rpi) and 19A–916U for gammaproteobacterial *Escherichia coli* (Eco) and *Serratia ficaria* (Sfi)] and **(B)** non-native base pairs between nucleotides 19 and 916 (19A–916G for betaproteobacterial Bsa, Cma, Hfl, Hth, Neu, Oho, Our, and Rpi and 19C–916U for gammaproteobacterial Eco and Sfi). *y* axis represents OD_{600} after cultivation at various temperatures (30°C, open bars; 37°C, shaded bars; 42°C, solid bars) for 14 h.

Metagenomic Screening for Functional 16S rRNA Genes in *E. coli*

We next used environmental metagenomes as a source for 16S rRNA genes. For all primer sets, specific amplification was obtained (Figure S4 in Supplementary Material). To investigate sequence diversity, the cloned genes were phylogenetically characterized. As shown in **Figure 4**, when Bac1f was used as a forward primer, the gene was dominated by proteobacterial 16S rRNAs [~84% (32/38)], but was much less for those obtained using the Bac8f(A) and Bac8f(C) primers [~56% (25/45) and ~73% (35/48), respectively].

The genes were then subjected to functional screening. Functional 16S rRNA genes were collected, and their microbial

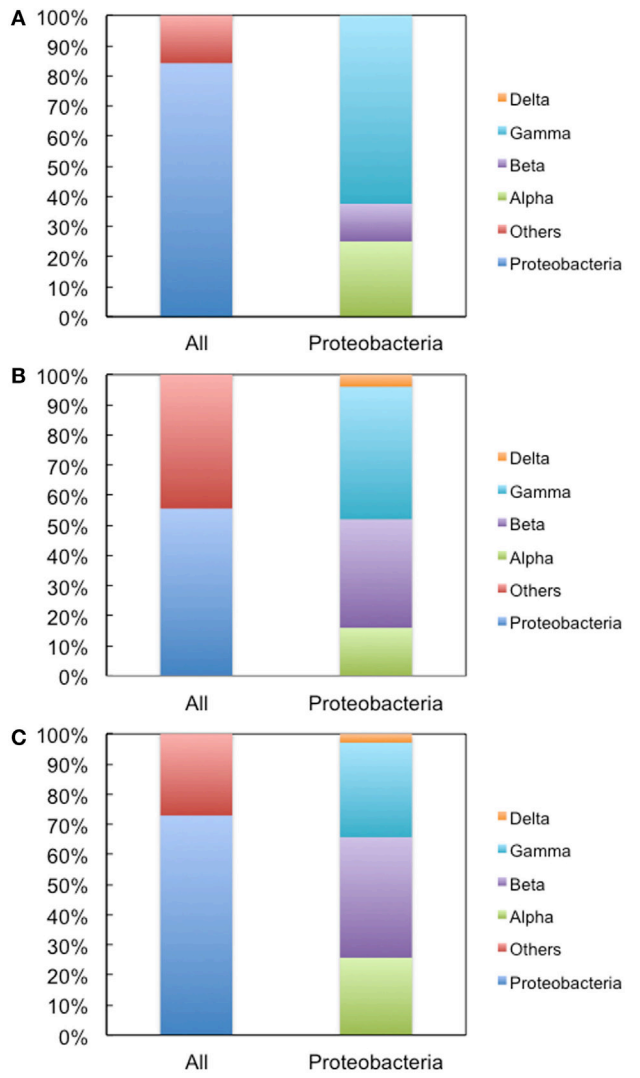


FIGURE 4 | Phylogenetic analysis of 16S rRNA genes cloned in pMS205aTp1 (before functional selection) amplicons. The forward primers used were **(A)** Bac1f, **(B)** Bac8f(A), and **(C)** Bac8f(C). "All" is the classification based on phylum and "Proteobacteria" is the class composition in the proteobacterial phylum.

origins and the base pair patterns between the 19th and 916th nucleotides were investigated. As shown in **Table 3**, when Bac1f was used, 85% (52/61) of functional 16S rRNAs were from gammaproteobacteria and the rest were from betaproteobacteria. Phylogenetic tree of these sequences is illustrated in Figure S5A in Supplementary Material. Base pair patterns were 78% 19A–916U (48/61) and 22% (13/61) 19C–916G. It is noteworthy that no artificially shuffled base pairs (A–G and C–U) were observed, implying that the newly designed primers did not introduce non-native base pairs and are adequate for functional studies. There are some mismatches between proteobacterial 16S rRNA and Bac1f sequences (Figure S3 in Supplementary Material), but in practice, we succeeded in retrieving various functional genes, suggesting that the mismatches at the 5′-end can affect annealing efficiency, but still remain effective for amplification. The lack of any alphaproteobacterial 16S rRNAs in our functional 16S rRNA collection may be due to functional incompatibility in *E. coli*.

TABLE 3 | Sequence analysis of functional 16S rRNA genes in *Escherichia coli* Δ7 retrieved from the metagenome.

	Bac1f	Bac8f(A)	Bac8f(C)
Predicted microbial origins			
Gammaproteobacteria	52	39	13
Betaproteobacteria	9	6	29
Deltaproteobacteria	0	0	2
Total	61	45	44
Base pairs between the 19th and 916th nucleotides			
A–G	0	21	0
A–U	48	23	1
C–G	13	1	42
C–U	0	0	1
Total	61	45	44

When Bac8f(A) was used as a forward primer, 87% (39/45) of functional 16S rRNAs were of gammaproteobacteria and the rest (13%; 6/45) were of betaproteobacteria. Phylogenetic tree of these sequences is illustrated in Figure S5B in Supplementary Material. When the base pair patterns were investigated, non-canonical base pairs were frequently observed. Approximately half (21/45) of the sequences contained the 19A–916G base pair, which may have resulted from the mis-annealing of the primer to template 16S rRNA genes containing 19C and 916G. Because this artificial base pair is permissive (or even encouraged) under normal growth conditions (**Figure 3B**), it is reasonable to find these species; simple PCR conditioning may not easily remove these mis-annealed products.

When Bac8f(C) was used as a forward primer, 66% (29/44) of functional 16S rRNAs were of betaproteobacteria, 30% (13/44) were of gammaproteobacteria, and 5% (2/44) were of deltaproteobacteria. Phylogenetic tree of these sequences is illustrated in Figure S5C in Supplementary Material. In this library, non-natural base pair patterns were rarely observed. Most were of the 19C–916G base pair and the non-native 19C–916U base pair was found in one clone (closest relative was gammaproteobacterial *Rahnella aquatilis*, NR_074921). This low-level occurrence of the 19C–916U base pair agreed with the detrimental (yet still non-lethal) effect of the pair on cell growth (**Figure 3B**).

In conclusion, we developed a new primer set, Bac1f and UN1542r, for the functional study of 16S rRNAs in *E. coli*. The effective utilization of the primers was demonstrated by retrieval of a range of functional 16S rRNAs from the proteobacterial lineage, all of which contained a native base pair between the 19th and 916th nucleotides.

AUTHOR CONTRIBUTIONS

KM, MS, and MT designed the study, conducted the data analysis, and wrote the manuscript.

ACKNOWLEDGMENTS

The authors thank Dr. Naohiro Noda (AIST, Japan) for the genomic DNA of *Nitrosomonas europaea*. This work was

supported by a JSPS Grant-in-Aid for Scientific Research (B) (Grant 26292048 to KM), Grant-in-Aid for Scientific Research on Innovative Areas (Grant 26670219 to KM), Grant-in-Aid for Challenging Exploratory Research (Grant 15H01072 to KM), and Grant-in-Aid for JSPS Fellows 26-7760 (to MT).

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/s0022-2836(05)80360-2

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.

Brink, M. F., Verbeet, M. P., and De Boer, H. A. (1993). Formation of the central pseudoknot in 16S rRNA is essential for initiation of translation. *EMBO J.* 12, 3987–3996.

Brodersen, D. E., Clemons, W. M. Jr., Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* 316, 725–768. doi:10.1006/jmbi.2001.5359

Dammel, C. S., and Noller, H. F. (1993). A cold-sensitive mutation in 16S rRNA provides evidence for helical switching in ribosome assembly. *Genes Dev.* 7, 660–670. doi:10.1101/gad.7.4.660

Eden, P. A., Schmidt, T. M., Blakemore, R. P., and Pace, N. R. (1991). Phylogenetic analysis of *Aquaspirillum* magnetotacticum using polymerase chain reaction-amplified 16S rRNA-specific DNA. *Int. J. Syst. Bacteriol.* 41, 324–325. doi:10.1099/00207713-41-2-324

Gutgsell, N. S., and Jain, C. (2012). Gateway role for rRNA precursors in ribosome assembly. *J. Bacteriol.* 194, 6875–6882. doi:10.1128/jb.01467-12

Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3801–3806. doi:10.1073/pnas.96.7.3801

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010

Kitahara, K., and Suzuki, T. (2009). The ordered transcription of RNA domains is not essential for ribosome biogenesis in *Escherichia coli*. *Mol. Cell* 34, 760–766. doi:10.1016/j.molcel.2009.05.014

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fbioe.2017.00014/full#supplementary-material.

Kitahara, K., and Miyazaki, K. (2011). Specific inhibition of bacterial RNase T2 by helix 41 of 16S ribosomal RNA. *Nat. Commun.* 2, 549. doi:10.1038/ncomms1553

Kitahara, K., Yasutake, Y., and Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19220–19225. doi:10.1073/pnas.1213609109

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. U.S.A.* 82, 6955–6959. doi:10.1073/pnas.82.20.6955

Poot, R. A., Van Den Worm, S. H., Pleij, C. W., and Van Duin, J. (1998). Base complementarity in helix 2 of the central pseudoknot in 16S rRNA is essential for ribosome functioning. *Nucleic Acids Res.* 26, 549–553. doi:10.1093/nar/26.2.549

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., et al. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827–834. doi:10.1126/science.1117230

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* 173, 697–703. doi:10.1128/jb.173.2.697-703.1991

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51, 221–271.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Miyazaki, Sato and Tsukuda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

METHODS
published: 25 May 2018
doi: 10.3389/fmicb.2018.01094

Clean Low-Biomass Procedures and Their Application to Ancient Ice Core Microorganisms

Zhi-Ping Zhong[1,2], Natalie E. Solonenko[2], Maria C. Gazitúa[2], Donald V. Kenny[1], Ellen Mosley-Thompson[1,3], Virginia I. Rich[2,4], James L. Van Etten[5], Lonnie G. Thompson[1,6] and Matthew B. Sullivan[2,7]**

[1] Byrd Polar and Climate Research Center, The Ohio State University, Columbus, OH, United States, [2] Department of Microbiology, The Ohio State University, Columbus, OH, United States, [3] Department of Geography, The Ohio State University, Columbus, OH, United States, [4] Department of Soil, Water and Environmental Science, The University of Arizona, Tucson, AZ, United States, [5] Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska–Lincoln, Lincoln, NE, United States, [6] School of Earth Sciences, The Ohio State University, Columbus, OH, United States, [7] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, United States

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
Charles K. Lee, The University of Waikato, New Zealand
Stefano Campanaro, Università degli Studi di Padova, Italy

***Correspondence:**
Lonnie G. Thompson
thompson.3@osu.edu
Matthew B. Sullivan
sullivan.948@osu.edu

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *22 December 2017*
Accepted: *07 May 2018*
Published: *25 May 2018*

Citation:
Zhong Z-P, Solonenko NE, Gazitúa MC, Kenny DV, Mosley-Thompson E, Rich VI, Van Etten JL, Thompson LG and Sullivan MB (2018) Clean Low-Biomass Procedures and Their Application to Ancient Ice Core Microorganisms. Front. Microbiol. 9:1094. doi: 10.3389/fmicb.2018.01094

Microorganisms in glacier ice provide tens to hundreds of thousands of years archive for a changing climate and microbial responses to it. Analyzing ancient ice is impeded by technical issues, including limited ice, low biomass, and contamination. While many approaches have been evaluated and advanced to remove contaminants on ice core surfaces, few studies leverage modern sequencing to establish *in silico* decontamination protocols for glacier ice. Here we sought to apply such "clean" sampling techniques with *in silico* decontamination approaches used elsewhere to investigate microorganisms archived in ice at ∼41 (D41, ∼20,000 years) and ∼49 m (D49, ∼30,000 years) depth in an ice core (GS3) from the summit of the Guliya ice cap in the northwestern Tibetan Plateau. Four "background" controls were established – a co-processed sterile water artificial ice core, two air samples collected from the ice processing laboratories, and a blank, sterile water sample – and used to assess contaminant microbial diversity and abundances. Amplicon sequencing revealed 29 microbial genera in these controls, but quantitative PCR showed that the controls contained about 50–100-times less 16S DNA than the glacial ice samples. As in prior work, we interpreted these low-abundance taxa in controls as "contaminants" and proportionally removed them *in silico* from the GS3 ice amplicon data. Because of the low biomass in the controls, we also compared prokaryotic 16S DNA amplicons from pre-amplified (by re-conditioning PCR) and standard amplicon sequencing, and found the resulting microbial profiles to be repeatable and nearly identical. Ecologically, the contaminant-controlled ice microbial profiles revealed significantly different microorganisms across the two depths in the GS3 ice core, which is consistent with changing climate, as reported for other glacier ice samples. Many GS3 ice core genera, including *Methylobacterium*, *Sphingomonas*, *Flavobacterium*, *Janthinobacterium*, *Polaromonas*, and *Rhodobacter*, were also abundant in previously studied ice cores, which suggests wide distribution across glacier environments. Together these findings help further establish "clean" procedures for studying low-biomass ice microbial communities and contribute to a baseline understanding of microorganisms archived in glacier ice.

Keywords: clean, low biomass, *in silico* decontamination, glacier ice, microbial community

INTRODUCTION

The cryosphere covers approximately 20% of the Earth's surface, and includes glaciers, snow, ice sheets, permafrost, lake ice, river ice, and sea ice (Fountain et al., 2012). Although microorganisms have been known to be present in glacier ice for nearly a century (McLean, 1919; Darling and Siple, 1941), such early findings were largely ignored until microorganisms were investigated in the deep Vostok ice core in the 1980s (Abyzov et al., 1982; Abyzov, 1993). This motivated further studies of microorganisms in ice cores collected from polar glaciers, such as the Greenland and Antarctic ice sheets (Priscu et al., 1998; Karl et al., 1999; Miteva et al., 2004; Tung et al., 2005; Knowlton et al., 2013), as well as some low-latitude ice caps, such as Guliya, Geladangdong, Zuoqiupu, and Noijinkangsang in China (Christner et al., 2000; Liu et al., 2016), Pastoruri in Peru (Gonzalez-Toril et al., 2015), Sajama in Bolivia (Christner et al., 2000), and Mount Humboldt in Venezuela (Ball et al., 2014).

These studies explored the mechanisms by which microorganisms could be archived in glacier ice, and used culture-dependent and -independent methods to reveal which microorganisms were archived. Microbial cells are buried and archived in glacier ice by three major processes: (i) emission from various sources (e.g., vegetation, soils, water, and rocks) and transportation in the air over the ice sheet by atmosphere currents; (ii) deposition onto the glacier ice surface; and (iii) gradual incorporation into the deeper ice layers as snow accumulates continuously (Santibanez-Avila, 2016) during the post-depositional period. Thus, microorganisms immured in ice cores represent those in the atmosphere at the time of deposition and hence reflect environmental conditions during the same time period (Priscu et al., 2007; Xiang et al., 2009). Previous investigations of the microbial community in polar glaciers (e.g., Miteva et al., 2009, 2015; Santibanez-Avila, 2016) and low-latitude glaciers (e.g., Yao et al., 2008; Chen et al., 2016) have suggested that microbial diversity and abundance preserved in deep ice cores are correlated with dust particle concentrations, local climate conditions, and global atmospheric circulation. Usually the biomass is very low in most glacier ice samples, with the estimated number of microbial cells ranging from 10^2 to 10^4 cells ml^{-1} (Miteva, 2008). Bacterial strains have often been recovered and isolated from glacier ice (Christner et al., 2000; Miteva et al., 2004; D'Elia et al., 2008; Zhang et al., 2008). Most of these isolated bacteria were psychrotolerant (D'Elia et al., 2008), which had optimal growth temperatures well above freezing and could be preserved under cold environments such as glacier ice for a long time (Willerslev et al., 2004). A growing number of studies have demonstrated the possibility for *in situ* microbial activity in glacier ice. The concentration of methane at several depths in the lowest 90 m is up to an order of magnitude higher than that at other depths in a 3,053-m-deep Greenland Ice Sheet Project 2 ice core (Tung et al., 2005). The excess methane at those depths was produced via *in situ* metabolism of the methanogenic archaea, which expended their metabolic energy to mainly repair damaged DNA and amino acids rather than for growth (Tung et al., 2005). Iron-reducing bacteria were reported to account for producing most of the excess CO_2 by reducing Fe^{3+} to Fe^{2+} and oxidizing the organic acids ions to CO_2 in ice at some depths of the bottom 13 m of the Greenland Ice Sheet Project 2 ice core (Tung et al., 2006). Some dominant genera (e.g., *Acinetobacter*, *Sphingomonas*, and *Comamonas*) within *Proteobacteria* and *Firmicutes* might be capable of post-depositional biological production of N_2O *in situ* at some depths of the North Greenland Eemian Ice Drilling ice core (Miteva et al., 2016). These reports suggested that excess gases (i.e., CO_2, CH_4, and N_2O) at some depths in the ice cores are due to ongoing *in situ* production by microorganisms. However, microbial activity is presumed to be very low in the deep glacial ice (Maccario et al., 2015). Furthermore, there is no direct evidence to indicate that microorganisms are active *in situ* in the ancient ice cores.

These advances have come in spite of glacier ice being a challenging medium in which to study microbial communities. First, microbial biomass in glacier ice is low (cell concentrations range from 10^2 to 10^4 cells ml^{-1}) and often only small volumes of ice are available (Miteva, 2008). Second, it is difficult to disrupt spore-forming and non-sporulating Gram-positive cells, which are frequently detected in glacier ice cores (Christner et al., 2000; Abyzov et al., 2004; Steven et al., 2008; Knowlton et al., 2013). These problems hamper obtaining microbial DNA of sufficient quantity and quality for culture-independent studies. In addition, because of its low biomass, contamination from sampling, storage, and preparation conditions is a major issue for studies of microbial communities in ice (Ram, 2009).

The surface ice of the ice core probably contains microbial contaminants that were introduced during drilling or handling ice cores in the fields or labs. Therefore, it is important to remove microbial contaminants on the surface of glacier ice cores (surface decontamination) for collecting low-contaminant ice samples. Considerable effort has been put forth to develop clean sampling technology and a number of surface decontamination strategies have been proposed and summarized in detail (Rogers et al., 2004; Christner et al., 2005). Briefly, these methods either killed microorganisms with chemical regents (Rogers et al., 2004), washed and removed the microorganisms in surface ice (e.g., Karl et al., 1999; Priscu et al., 1999; Christner et al., 2005), or collected the ice core interior by using a melting device (e.g., Abyzov, 1993; Christner et al., 2000). After surface decontamination, microbial contaminants can also be introduced into ice samples from environments (e.g., laboratory personnel, tools, reagents, and air) during the processing of ice including ice sampling, concentrating cells, and DNA extraction and sequencing. Specifically, DNA extraction methods can have a profound effect in studying microbial communities, and it has been a major source of variation in microbial Metagenomic work for low-biomass samples (Morgan et al., 2010; Woyke et al., 2011; Salter et al., 2014; Glassing et al., 2016). "Background" controls help reveal potential contaminants introduced during the processing of ice after surface decontamination (Willerslev et al., 2004; Hebsgaard et al., 2005), and the studied sample datasets can be *in silico* decontaminated by removing microbiota found in "background" controls. "Background" controls were included for microbial investigations of glacier ice in some

reports, whereas some found that these controls did not yield any amplification products and suggested "clean" ice processing procedures (Sheridan et al., 2003; Yao et al., 2008; Zhang et al., 2008; Liu et al., 2009). Open air culture plates were used to check potential air contaminants that could be cultivated and were removed from the ice samples (Ram, 2009; Knowlton et al., 2013; Miteva et al., 2015). Two "background" controls (nanopure water and autoclaved nanopure water) were conducted during a metagenomic study of two ice samples from the Greenland Ice Sheet Project 2 ice core (Knowlton et al., 2013). A total of 55,254 and 52,078 high-quality 454 reads were generated for two ice and two control samples, respectively. Only 33 sequences that were unique to the ice were selected for further microbial analysis after removing the sequences that were in common with the control samples and were considered as potential contaminants (Knowlton et al., 2013). In another study, nine ice samples were excluded from further microbial analysis since they had a high abundance (68.7 ± 24.8%) of an operational taxonomic unit (OTU) that was also abundant (73.1%) in a "background" control sample conducted in parallel to the ice DNA extractions (Cameron et al., 2016).

All of these studies removed suspected "contaminants" in ice samples by conducting "background" controls and obtained decontaminated data for further microbial analysis. It is a challenge to determine if the removed microorganisms were from ice or contaminants, and it has been suggested to not remove OTUs identified in "background" controls due to cross-contamination if they are biologically expected in the given sample type (Salter et al., 2014).

Here we sought to establish low-biomass, culture-independent "clean" procedures to survey microorganisms in glacier ice and then together with several "background" controls and published *in silico* decontamination methods use them to identify and quantify microbial diversity at two depths in an ice core from the Guliya ice cap in northwestern Tibet. The Tibetan Plateau is a mountainous area (average altitude of ∼4,500 m) that covers about 2.5 million km^2 of the Eurasian continent (Cui and Graf, 2009). It contains the third largest reservoir of glacial ice on Earth (Qiu, 2008) and is the major water source for Southern and Eastern Asia (Cui and Graf, 2009; Immerzeel et al., 2010). The Guliya ice cap is located at the northwestern Kunlun Mountains of the Tibetan Plateau and is the highest (6,700 m), largest (>200 km^2), thickest (308.6 m), and coldest (−18.6°C) ice cap among all the ice caps in middle-low latitude regions (Yao et al., 1992; Thompson et al., 1995). Previous studies on the Guliya ice cap focused primarily on the formation, structure, geochemistry, and dating of the ice, and found that the Guliya ice cap preserved the history of past climate change over tens to hundreds of thousands of years (Yao et al., 1992, 2004; Thompson et al., 1995, 1997, 2000; Wang et al., 2002; Wu et al., 2004). The microbial community in this ice cap has been largely unexplored except for two culture-dependent studies, which recovered viable bacterial strains immured in glacial ice that was more than 500,000 years old. These recovered isolates belonged to the alphaproteobacterial, betaproteobacterial, actinobacterial, and low-G+C Gram-positive bacterial lineages (Christner et al., 2000, 2003).

MATERIALS AND METHODS

Site Characterization and Field Sampling

The Guliya summit 3 (GS3) ice core was drilled in October 2015 from the summit of the Guliya ice cap (35°17′ N, 81°29′ E, ∼6700 m above sea level, **Figure 1A**). This ice core was 10 cm in diameter, 50.80 m in length (**Figure 1B**), and the bedrock temperature was about −15°C. Ice core sections were sealed in plastic tubes, put into cardboard tubes covered with aluminum, and transferred at −20°C by truck from the drill site to freezers in Lhasa, by airplane to freezers in Beijing, by airplane to Chicago, and then by freezer truck to the Byrd Polar and Climate Research Center at Ohio State University where it is stored at −34°C.

Ice Core Sampling and Physiochemical Conditions

The GS3 ice core sections were transferred from −34°C to the sampling temperature of −5°C overnight to reduce the possibility of fracturing during surface decontamination by cutting and washing. The decontamination procedures used included washing and removing the surface of the ice core as described previously (Karl et al., 1999; Priscu et al., 1999) with some modifications that added an additional removal of the ice core's outermost layer by cutting with a band saw. Briefly, ∼6 mm of the outermost layer was removed from the ice cores with a band saw. The inner ice core was cut into 3–4 cm sections in a cold room (−5°C) and the sections were thoroughly washed with filtered (0.22-μm-pore-sized filter) and sterilized water to remove 3–5 mm of the surface layer after which they were melted in covered containers in a Class 100 clean room at room temperature for about 4 h. Although prior bacterial cultivation work was conducted in the same cold and clean rooms, no biological experiments of any kind had been conducted in them for more than 10 years. Sections of melted ice from the depth of 41.10–41.84 m of the GS3 ice core were combined as one sample (D41), and those from 49.51 to 49.90 m were combined as another sample (D49; **Figure 1B**), for microbial analysis. Dust, chemical ions, and oxygen isotopes were analyzed as described previously (Davis and Thompson, 2006). The approximate age of each ice section was dated by matching the oxygen isotopic ratios with those from another 310.6-m ice core of similar age collected from the Guliya ice cap in 1992 (Thompson et al., 1995, 1997).

"Background" Controls

Four "background" control samples were used to investigate possible sources of background contamination during processing. First, we assessed what microorganisms were in the air from the cold and clean rooms used for ice core processing. Specifically, cells from 28.3 and 28.8 m^3 of air were collected from the cold room (named Air_ColdRoom) and the clean room (Air_CleanRoom), respectively. Cell collection in the air started at the same time as the processing of the GS3 ice core sections, and continued after ice core processing for a total of 4 days of sampling. The air samples were passed through sterilized polycarbonate 0.8-μm-pore-sized filters (Cat No. ATTP02500, Isopore), as well as a Button Aerosol

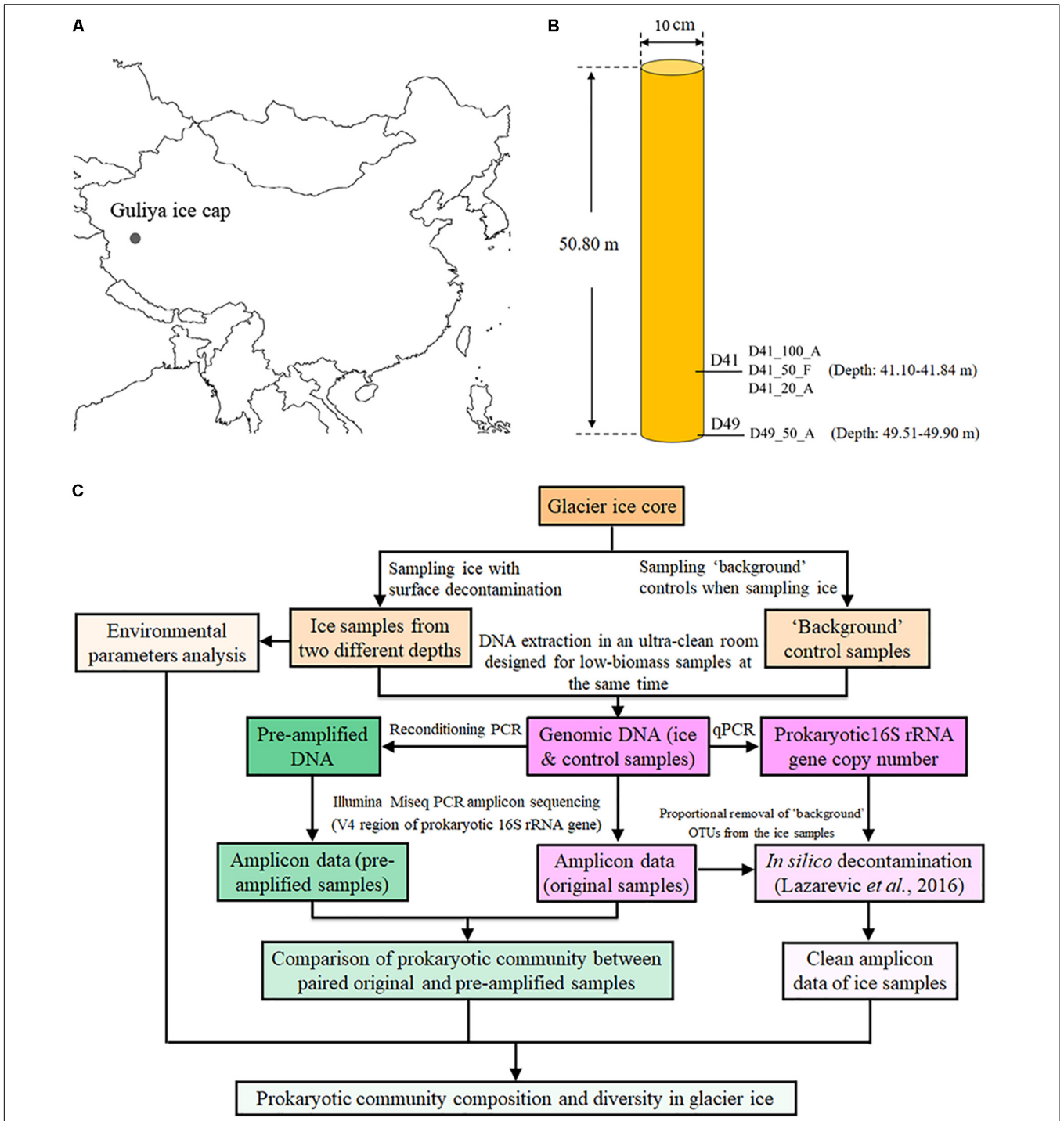


FIGURE 1 | Location **(A)**, sampling sites **(B)**, and an overview of experimental design **(C)** for investigating the microbial communities of the GS3 ice core drilled from the Guliya ice cap. The sample names from this study are coded as follows for the example of D41_100_A: D41, the depth of the ice sample (41 m under the surface); 100, the ice volume for DNA extraction (100 ml). Abbreviations of two methods for concentrating cells: A, Amicon Ultra Concentrators; F, filters with 0.22-μm pore size. Three samples including D41_100_A, D41_50_F, and D41_20_A were collected from the same mixture of melted ice from 41.40 to 41.84 m deep, while D49_50_F was sampled from ice 49.51–49.90 m deep.

Sampler (SKC Inc.), which is reported to have higher recovery efficiency of bacteria (specific recovery efficiency not provided) in indoor and outdoor air compared to three other samplers including the IOM Inhalable Dust Sampler, the NIOSH Personal Bioaerosol Cyclone Sampler, and the 37-mm Filter Cassette sampler (Wang et al., 2015). These two controls evaluated

background contamination due to exposure to air during the ice processing. Second, an artificial ice core made from 0.22-μm filtered (Cat No. MPGP04001, Millipak® Express 40 Filter, Merck KGaA) and autoclaved (121°C for 30 min) water was frozen (−34°C for 12–24 h) and then processed in parallel with the GS3 ice core samples through the entire analysis. This control facilitated evaluation of contamination from the instruments used to process the ice. Finally, a blank control was established by extracting DNA directly from 400 ml of 0.22-μm filtered and autoclaved water (as above). This control allowed evaluation of contamination downstream of the ice processing, including the molecular procedures (DNA extraction, PCR, library preparation, and sequencing).

Genomic DNA Extraction

A total of 400 ml of artificial ice (Artificial_ice), 400 ml of the blank control (Blank), and 50 ml each of the two ice samples (D41_50_F and D49_50_F) were filtered through sterilized polycarbonate 0.22-μm-pore-sized filters (Cat No. GTTP02500, Isopore) to collect microorganisms including all bacterial/archaeal cells, with cell sizes exceeding 0.22 μm. The filters were used to isolate DNA. DNA was also isolated from cells concentrated from 100 (D41_100_A) and 20 ml (D41_20_A) of Sample D41 to 0.6 ml by 100 kDa Amicon Ultra Concentrators (EMD Millipore, Darmstadt, Germany), with a pre-filtration by 3.0-μm-pore-size filters to remove big dust particles to avoid clogging the concentrators. Community DNA was isolated from these four ice samples (D41_100_A, D41_20_A, D41_50_F, and D49_50_F) and the four "background" controls (Artifical_ice, Blank, Air_ColdRoom, and Air_CleanRoom) with a DNeasy Blood & Tissue Kit (Cat No. 69506, QIAGEN) according to the manufacturer's instructions, with an additional step of beating with beads to disrupt bacterial spores and Gram-positive cells before cell lysis by homogenizing at 3,400 rpm for 1 min with 100 mg of autoclaved (121°C for 30 min) 0.1-mm-diameter glass beads (Cat No. 13118-400, QIAGEN) in a MiniBeadBeater-16 (Model 607, BioSpec Products). DNA was stored at −80°C. DNA denaturants (DNA AWAY, Cat No. 7010, Thermo Scientific) and 70% ethanol were used to eliminate potential naked DNA and cell contaminants on the surface of gloves, lab benches, and some tools used in this study.

Real-Time Quantitative Polymerase Chain Reaction (qPCR)

Total bacterial and archaeal biomass was estimated using real-time qPCR for the four ice samples and the four "background" controls after isolating DNA. Primer sets 1406f (5′-GYACWCACCGCCCGT-3′) and 1525r (5′-AAGGAGGTGWTCCARCC-3′) were used to amplify bacterial and archaeal 16S rRNA genes (Vanwonterghem et al., 2014). Each 20-μl reaction contained: 10 μl 2× QuantiTect SYBR Green PCR Master Mix (Cat No. 204143, QIAGEN), 0.5 μl of each primer (1406f/1525r, 10 mM), 3 μl template DNA, and 6 μl RNase-free water. Thermocycling consisted of an initial polymerase activation and template DNA denaturation at 95°C for 15 min, followed by 40 cycles of 95°C for 15 s, 55°C for 30 s, and 72°C for 15 s. A melt curve was produced by running a cycle of 95°C for 15 s, 55°C for 15 s, and 95°C for 15 s. A standard curve was generated with a PCR product using primers 1406f/1525r from *Cellulophaga baltica* strain 18 (NCBI accession number of the complete genome, CP009976). All reactions were performed in triplicate, using an Illumina Eco cycler (Cat No. 1010180).

Reconditioning PCR

Reconditioning PCR, reported to reduce PCR artificial bias (Thompson et al., 2002; Lenz and Becker, 2008), was conducted for each sample to pre-amplify the V4 region of prokaryotic 16S rRNA genes with primer sets 515f/806r (Caporaso et al., 2011), which was selected for amplicon sequencing to investigate the microbial community. A Phusion High-Fidelity DNA Polymerase Kit (Cat No. F530L, Thermo Scientific) was used for reconditioning PCR. The 20 μl PCR reaction consisted of: 4 μl 5× Phusion HF Buffer (containing $MgCl_2$), 0.4 μl 10 mM dNTP, 1 μl of each primer (515f/806r, 10 mM), 0.2 μl high-fidelity DNA polymerase, 2 μl template DNA, and 11.4 μl of water. For all eight samples, the first round amplification consisted of a 40-s denaturing step at 98°C, followed by 28 cycles of 8 s at 98°C, 20 s at 48°C, and 15 s at 72°C, with a final extension of 8 min at 72°C. To recondition the PCR products, the amplified reactions were diluted fivefold into a fresh reaction mixture of the same composition and cycled eight times using the same conditions as the first round PCR. All reactions were conducted in triplicate, which were combined as one sample after each PCR. The combined reaction mixtures after reconditioning PCR were purified by Agencourt AMPure XP Beads (Cat No. A63881, Beckman Coulter) and collected in 50 μl of buffer, according to the manufacturer's instructions.

Tag-Encoded Amplicon Pyrosequencing of Microbial Community

Bar-coded primers 515f/806r (Caporaso et al., 2011) were selected to amplify the V4 hypervariable regions of 16S rRNA genes of bacteria and archaea for both original and pre-amplified samples. Resulting amplicons were sequenced by the Illumina MiSeq platform (paired-end reads), as described previously (Caporaso et al., 2011, 2012). These experiments were performed at Argonne National Laboratory.

Sequence Analysis

Sequences with an expected error >1.0 or length <245 nt were excluded from the analyses (Edgar, 2013). The remaining sequences were truncated to a constant length (245 nt). Various analyses were conducted using the Quantitative Insights Into Microbial Ecology (QIIME, version 1.9.1) software package (Caporaso et al., 2010) with default parameters, except that chimera filtering, OTU clustering, and singleton excluding were performed with QIIME through the UPARSE pipeline (Edgar, 2013). A phylogenetic tree was constructed with a set of sequence representatives of the OTUs using the method of FastTree (Price et al., 2009). Chimeras were identified and filtered by UPARSE with the UCHIME algorithm using the ChimeraSlayer reference database (Haas et al., 2011), which is considered to be

sensitive and quick (Edgar et al., 2011). Reads were clustered into OTUs at 97% sequence similarity by UPARSE. A representative sequence from each OTU was selected for taxonomic annotation using the Ribosomal Database Project (RDP) classifier (Wang et al., 2007) from the RDP Release 11.5 database. Taxonomic assignments with <80% confidence were marked as unclassified taxa. Mitochondrial and chloroplastic sequences were excluded from further analysis.

Relative abundance of the microbial profiles at the genus level was calculated for each sample. The differences in microbial community composition between each paired original and pre-amplified samples were tested for significance using a two-tailed paired *t*-test. A heatmap was generated based on the number of sequences per OTU per 30,000 sequences using functions in the Pheatmap package version 1.0.8 (Kolde, 2015) in R version 3.4.2 (R Core Team, 2012). A new profile of OTU composition for the ice samples was generated after *in silico* decontamination as described previously (Lazarevic et al., 2016). Briefly, an R-OTU value was designated as the ratio between the mean "absolute" abundance of OTUs in "background" controls and ice samples; then, an approximated estimation of the "absolute" abundance of OTUs was calculated by multiplying the relative abundance of each OTU by the 16S rRNA gene copy number in a given sample (determined by qPCR). The OTUs with R-OTU values >0.01 were considered to be contaminants and were removed from the ice samples. The significance of the difference in microbial community between D41 and D49 samples was evaluated by Analysis of Similarity Statistics (ANOSIM) (Clarke, 1993), which was performed using functions in the Vegan package version 2.4-4 (Dixon, 2003) in R version 3.4.2 (R Core Team, 2012).

Comparison of Microbial Profiles Between Guliya Ice Cap and Several Other Ice Caps

The microbial profiles of the Guliya ice cap were compared to those from other glaciers and ice fields previously characterized by next-generation sequencing of the overlapped region (V4) of the 16S rRNA gene. The selected samples included two to four samples from each of three Tibetan Plateau ice caps (Geladangdong, Noijinkangsang, and Zuoqiupu) (Liu et al., 2016) and the Greenland ice sheet (Miteva et al., 2015, 2016). Sequence files (.fastq) of each sample were obtained from NCBI Sequence Read Archive using the SRA Toolkit[1] and combined with that of the Guliya ice cap samples from this study. Sequences were analyzed as described in the previous section. In addition, samples were also clustered by the unweighted pair group method with the arithmetic mean (UPGMA) based on weighted UniFrac distances, which accounts for changes in relative taxon abundance (Caporaso et al., 2010). Principal coordinates analysis (PCoA) using weighted UniFrac metrics was performed to distinguish general distribution patterns of microbial profiles among all samples. The Mantel test was conducted to evaluate the linkage between the microbial community structure and environmental parameters.

[1]https://www.ncbi.nlm.nih.gov/books/NBK158900/

Nucleotide Sequence Accession Numbers

The nucleotide sequences discovered during this study have been deposited in the NCBI Sequence Read Archive under accession number SRP114723.

RESULTS AND DISCUSSION

A GS3 ice core, which was 50.80 m in length and contained ice up to ∼30,000 years old, was retrieved from the Guliya ice cap, China in 2015 (**Figures 1A,B**) to monitor past climate change and archived microbial profiles. In this study, four "background" controls including a sterile water artificial ice core (named as Artificial_ice), two air samples collected from the ice processing laboratories (Air_ColdRoom and Air_CleanRoom), and a blank sterile water sample (Blank) were co-processed with four real ice samples to check "background" microbial profiles and their abundances, and establish "clean" sampling and amplicon sequencing protocols. Subsequently the procedures, together with published *in silico* decontamination methods, were used to investigate the microbial profiles archived in ice at two depths in the GS3 ice core (overview in **Figure 1C**).

Establishment of Microbial "Contaminants" From Four "Background" Controls

To obtain clean amplicon sequencing reads of ice samples, the first step was to identify how much biomass and what microbial taxa (contaminants) were in the four "background" controls. Total microbial cell abundance was first measured by epifluorescence microscopy after the cells were concentrated on a 0.22-μm-pore-sized filter (Cat No. GTTP02500, Isopore) and stained by SYBR Gold as described previously (Noble and Fuhrman, 1998). A total of 11.3×10^3 and 8.4×10^3 cells were observed for Air_ColdRoom and Air_CleanRoom, respectively; while less than 100 and 10 cells were detected on the filters of Artificial_ice and Blank, respectively (data not shown). The DNA extraction process could also introduce contaminations, such as those from investigators (e.g., human skin and respiratory), tools, and reagents (Woyke et al., 2011; Salter et al., 2014). Therefore, qPCR was performed to assess how much biomass of microbial DNA was obtained from the four "background" controls after DNA extraction, by calculating the copy number of 16S rRNA genes with reference to a standard curve. The 16S rRNA gene copies per microliter in the 50-μl volumes of each sample revealed 49, 49, 51, and 27 copies/μl in Air_ColdRoom, Air_CleanRoom, Artificial_ice, and Blank, respectively (Supplementary Figure S1).

Microbial profiles in the four "background" controls were investigated using Illumina Miseq PCR amplicon sequencing. The QC data were normalized to 15,000 sequences for each sample (i.e., each MiSeq sequencing library) for further analysis. These sequences were affiliated with 169 bacterial genera, 94

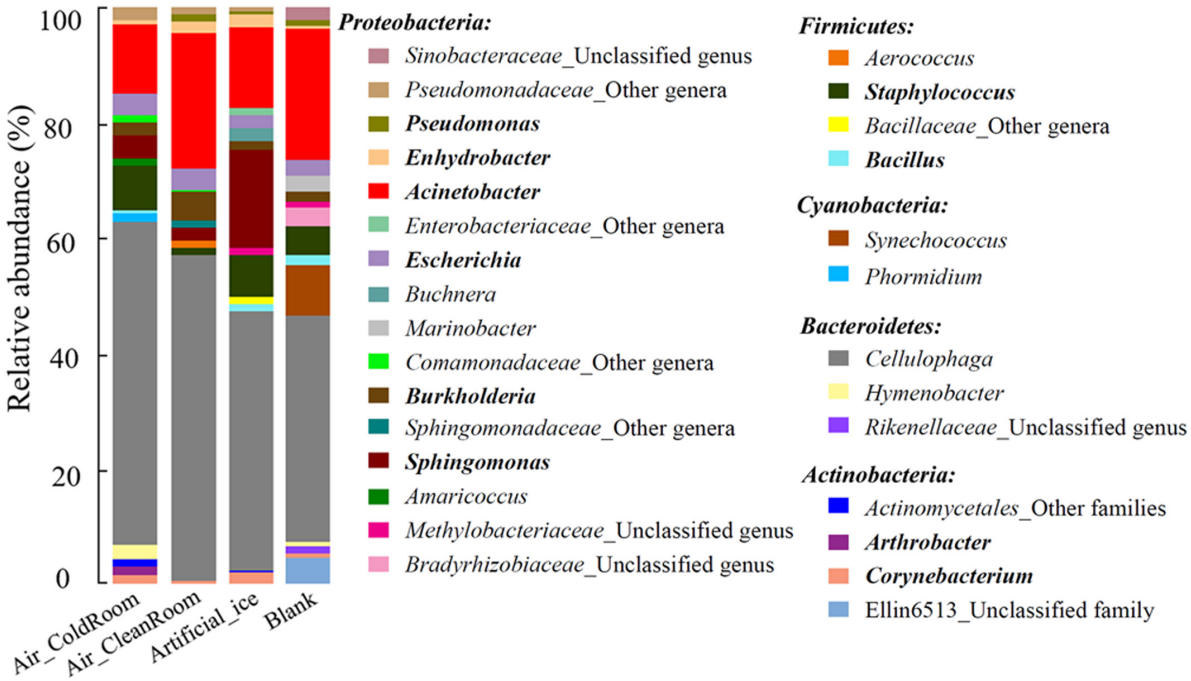


FIGURE 2 | Microbial community structure of the 29 most abundant genera in the four "background" controls. Genera belonging to the same phylum are described under the phylum name. The "other genera/families" represent unclassified sequences and could not be assigned to a single genus/family. Genera previously reported as contaminant taxa are indicated in bold. The four "background" controls: Air_ColdRoom and Air_CleanRoom, two air samples collected from a cold and clean room, respectively, in which the ice samples were processed; Artificial_ice, an artificial ice sample made by sterile water and processed along with the glacier ice samples; Blank, a blank sample with 400 ml sterile water.

of which had recognized names (Supplementary Table S1). The 29 most abundant genera, each of which accounted for ≥1.0% of the sequences in at least one sample, comprised 82.9–88.8% of each community and were selected to illustrate the microbial communities of the four "background" controls (**Figure 2**). These genera belonged to five phyla, *Proteobacteria*, *Firmicutes*, *Cyanobacteria*, *Bacteroidetes*, and *Actinobacteria*, which contained 16, 4, 2, 3, and 4 genera, respectively (**Figure 2**). Many exogenous sequences assigned as unexpected taxa from contamination have been detected during the analysis of low-biomass environmental microbiota (Biesbroek et al., 2012; Lazarevic et al., 2014), cultures (Salter et al., 2014; Lazarevic et al., 2016), and diluted mock microbial communities (Willner et al., 2012). These contaminants might come from lab air (Othman, 2015; Lauder et al., 2016), investigators (e.g., human skin and respiratory) (Knights et al., 2011), tools, and reagents used for DNA extraction, PCR amplification, and sequencing (Corless et al., 2000; Barton et al., 2006; Glassing et al., 2016). Some contaminant genera detected in this study overlapped with previously described contaminant groups, including the genera *Sphingomonas* (Barton et al., 2006; Laurence et al., 2014), *Burkholderia* (Laurence et al., 2014), *Escherichia* (Tanner et al., 1998; Laurence et al., 2014; Salter et al., 2014), *Acinetobacter* (Tanner et al., 1998; Barton et al., 2006), *Enhydrobacter* (Salter et al., 2014), *Pseudomonas* (Grahn et al., 2003), *Corynebacterium* (Salter et al., 2014), *Arthrobacter* (Salter et al., 2014), *Bacillus* (Grahn et al., 2003), and *Staphylococcus* (Othman, 2015; **Figure 2**). These findings indicate that many microbial taxa are common contaminants in microbial community studies. Two additional genera, *Cellulophaga* and *Synechococcus*, were detected and interpreted as contaminants in this study (**Figure 2**). Many isolates belonging to these two genera have been used previously as type strains to investigate virus–host interactions in our laboratory (Deng et al., 2014; Dang et al., 2015), which is why we interpreted these as low-level laboratory contaminants. It is likely then that such contamination would vary from laboratory to laboratory for low biomass samples, which is consistent with prior findings (Willerslev et al., 2004).

Paired Original and Pre-amplified Samples Capture Almost Identical Microbial Profiles

A previous report (Salter et al., 2014) indicated that it is difficult to determine the composition of a microbial community if the number of microorganisms used for DNA extraction is less than 10^3–10^4 cells. Considering the low biomass in our "background" controls (10^0–10^4 cells), we pre-amplified the targeted region (V4) of bacterial and archaeal 16S rRNA genes by reconditioning PCR (Thompson et al., 2002) in all of the four original "background" controls before standard amplicon sequencing. These pre-amplified samples were subjected to standard amplicon sequencing together with

the original "background" controls. Microbial profiles were compared between each pair of original and pre-amplified samples to determine whether reliable microbial community values were obtained for the original and their pre-amplified "background" controls.

All of the four original and four pre-amplified libraries were normalized to 15,000 sequences for further analysis. The 36 most abundant genera, each of which accounted for >1.0% of sequences in at least one sample, comprised 85.7% of the total 120,000 sequences in eight samples. These groups were designated as "major genera" and used to exemplify the microbial community of all the original and pre-amplified samples (Supplementary Table S2 and Supplementary Figure S2). All of these "major genera" were detected in each pair of original and pre-amplified samples, and accounted for almost all of each microbial community (Supplementary Table S2 and Supplementary Figure S2). For example, the 16 most abundant genera, including *Cellulophaga*, *Acinetobacter*, *Staphylococcus*, *Sphingomonas*, *Escherichia*, *Hymenobacter*, *Burkholderia*, an unclassified genus within the family *Pseudomonadaceae*, *Corynebacterium*, *Arthrobacter*, *Amaricoccus*, an unclassified genus within the family *Comamonadaceae*, *Enhydrobacter*, *Propionibacterium*, *Stenotrophomonas*, and *Streptococcus*, were all similarly represented in the original sample Air_ColdRoom and its pre-amplified sample Air_ColdRoom_28+8 ("28+8" represents 28 and 8 cycle times at the first and the reconditioning PCR rounds, respectively) (Supplementary Table S2 and Supplementary Figure S2). These 16 genera comprised 96.7 and 95.6% of the microbial community in Air_ColdRoom and Air_ColdRoom_28+8, respectively. In addition, results from the two-tailed paired *t*-test showed pre-amplification with reconditioning PCR does not significantly alter the microbial community in original samples (*p*-values were 0.60–0.92 for the above four pairs of original and pre-amplified samples, respectively). The similar community composition in each pair of original and pre-amplified samples indicates that the reliable microbial profile values were obtained for both original and pre-amplified "background" controls, and that reconditioning PCR captures a microbial community that is almost identical to the original samples with low biomass. Lenz and Becker (2008) used standard PCR and reconditioning PCR to analyze polymorphic loci and investigate genetic variation in the major histocompatibility complex (MHC) class IIB genes of the three-spined stickleback (*Gasterosteus aculeatus*). They reported that 24% of the clones were artificial allele chimeras generated by the hybrids of two or three different alleles that occurred in the same individual, using standard PCR, while the number of artificial chimeras was reduced 10-fold by reconditioning PCR (Lenz and Becker, 2008). The results from this study and previous reports confirm that reconditioning PCR reduces amplification bias from multi-template PCR products before library construction and the results more closely reflect the genetic diversity of the original samples (Thompson et al., 2002). In addition, our previous studies of viromes suggest that the degree of amplification has little impact on the resulting metagenomes (Duhaime et al., 2012; Solonenko et al., 2013; Solonenko and Sullivan, 2013).

Proportional Removal (*In Silico* Decontamination) of "Contaminants" From the GS3 Ice Core Samples

With the established contaminant taxa from the four "background" controls, we next *in silico* removed these contaminants in the amplicon sequencing dataset of the four Guliya ice samples to generate "clean" sequencing reads of these ice samples by the following procedure. A recently published *in silico* decontamination strategy, that combines the information of both the relative abundance of each OTU and the 16S rRNA gene copy number in a given sample (proportional removal), effectively removes the contaminant sequences derived from the "background" controls in the samples of interest (i.e., ice samples) as described in the section "Materials and Methods" (Lazarevic et al., 2016). To use this *in silico* decontamination strategy for the ice samples and the "background" controls, we first quantified the 16S rRNA gene copy number in the ice samples and checked the differences in the OTU compositions between the ice samples and the "background" controls.

The 16S rRNA gene copies per microliter in the 50 μl volumes from each ice sample were 4.60×10^3, 0.97×10^3, 0.95×10^3, and 1.25×10^3 copies/μl in D41_100_A, D41_20_A, D41_50_F, and D49_50_F, respectively (Supplementary Figure S1). Thus the biomass in the ice samples was about 50–100-times higher than that in all four "background" controls (27–51 copies/μl, Supplementary Figure S1). The amplicon data of the four Guliya ice samples and four "background" controls were normalized to 30,000 sequences for further analysis. The 32 most abundant OTUs (relative abundance was >1.0% in at least one sample) comprised 88.6% of the total sequences (240,000) of the ice samples and the "background" controls, and were selected to illustrate their OTU compositions (**Figure 3**). Total sequences belonging to eight OTUs, including OTU_1, OTU_4, OTU_5, OTU_9, OTU_3, OTU_953, OTU_188, and OTU_12, accounted for 93.4–98.9% of all sequences in the 32 OTUs for each ice sample, but only made up 0.3–2.9% of the "background" controls (**Figure 3**). In contrast, the other 24 OTUs contributed 1.1–6.4 and 97.1–99.7% of the sequences in the 32 OTUs from the ice samples and "background" controls, respectively (**Figure 3**). These results indicate that the most abundant OTUs in the ice samples were notably different from those in the "background" controls, and that the latter 24 OTUs are probably contaminants and should be *in silico* removed from the ice samples before taxonomic analysis. Sequences belonging to the most abundant OTUs in the ice samples (i.e., OTU_1, OTU_3, and OTU_4) were also detected in small amounts in the "background" controls (**Figure 3**). Similar results were also observed in a study that investigated the bacterial community in mock and control samples (Lazarevic et al., 2016), indicating possible cross-contamination during DNA extraction from samples with much higher biomass to samples with lower biomass (e.g., from ice to "background" controls in this study). Thus, special caution should be taken with regard to the suspicious "contaminating" microorganisms that are also discovered to be present in the investigated environments.

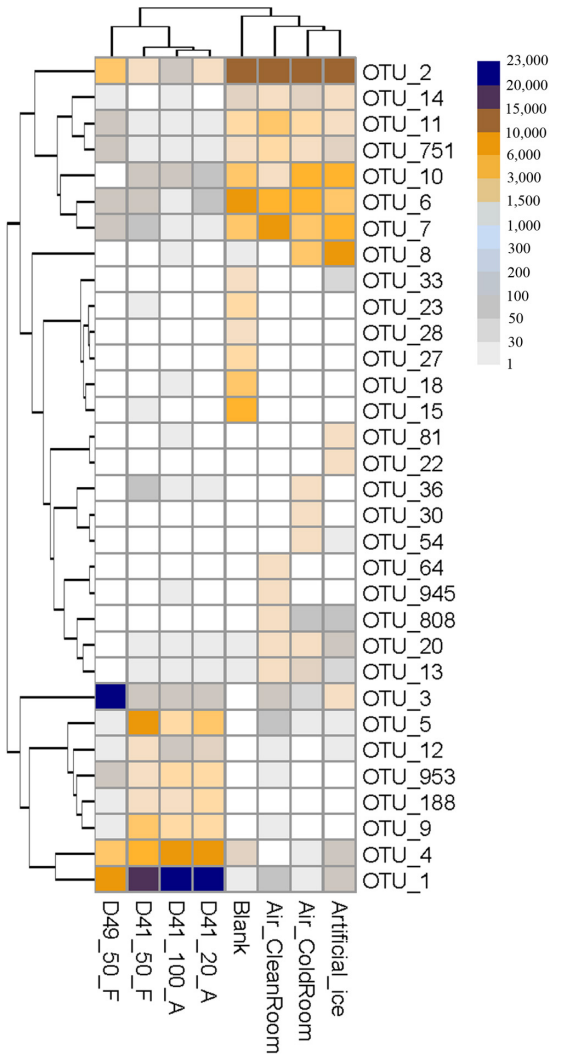


FIGURE 3 | Heatmap showing the sequence number of each OTU per 30,000 sequences for the Guliya ice samples and "background" controls. All OTUs accounted for >1.0% of sequences (i.e., >300 sequences) in at least one sample. OTUs were defined as reads with 97% sequence similarity.

The dataset of ice samples in this study was decontaminated *in silico* with the proportional removal strategy mentioned above using R-OTU cut-off values of 0.01 by removing OTUs with this ratio exceeding 0.01 (Lazarevic et al., 2016). After *in silico* decontamination, 93.2–97.8% of the reads in the ice samples were retained, while only 0.2–2.3% of the reads were retained in the "background" controls; this small number of reads might represent cross-contamination of the ice samples with much higher biomass (Supplementary Figure S1), as discussed above. An important but largely unrecognized source of laboratory-based contamination is PCR product carryover because the amount of contaminant DNA might be larger than the DNA in the glacier ice samples (Kwok and Higuchi, 1989; Willerslev et al., 2004; Willerslev and Cooper, 2005). DNA molecules and cells from laboratory personnel, tools, reagents, and air can also introduce contaminants. Thus, it is important to include no-template control "blank" samples in experiments with low biomass to control for this low-level source of contamination (Hebsgaard et al., 2005). "Background" controls and the subsequent removal of "suspect" contaminants were included in some prior culture-dependent and -independent microbial studies with glacier ice cores (Ram, 2009; Knowlton et al., 2013; Miteva et al., 2015; Cameron et al., 2016). These reports reflect the laboratory contamination in the glacier ice samples and indicate the necessity to *in silico* remove the contaminants. The challenge is to determine if the removed microorganisms originated from ice or contaminants. The proportional removal approach used in this study may efficiently find the OTUs derived from cross-contamination, in contrast to those derived from reagents, and thus not remove them from the dataset; this process may improve the taxonomic representation in the low-biomass ice samples.

We note, however, that there are variations across taxa in DNA extraction and recovery efficiency (Yuan et al., 2012), which is associated with the qPCR-quantified 16S rRNA gene abundance in this study. Our method can be used to proportionally adjust the contaminants based on their relative amounts if the DNA extraction and recovery efficiency are similar or nearly identical for the same taxon across samples. We also realize that it is hard to quantify the amount of contamination from air to ice samples, and that the volume of collected air or other factors also influence the 16S rRNA gene concentration that is used to calculate the "absolute" abundance of each OTU. The retrieved biomass of air samples was only a small fraction of that observed in the ice samples (Supplementary Figure S1), although the air was continuously sampled for 4 days. This suggests that the cold and clean rooms were quite clean for processing the low-biomass ice samples in this study. The DNA-denaturing regent was used to "clean" the surface of the bench, gloves, and some tools before processing the glacier ice samples, but not used for removing naked DNA from the filtered and/or autoclaved water, or from reagents we used. However, the "background" controls can help identify and remove such contamination *in silico* from any possible contaminant naked DNA in the ice. With more attention paid to conducting "background" controls and *in silico* decontamination for microbial investigations of low-biomass glacier ice, as well as the usage of internal standards to better control for DNA extraction and recovery efficiency (Mumy and Findlay, 2004), we will be able to remove "background" contaminants more efficiently and obtain cleaner ice microbial data in future studies.

Microbial Profiles Differ Between Ice Samples From Two Different Depths of the GS3 Ice Core

With the "clean" reads after *in silico* decontamination, we then examined the microbial communities of three and one ice samples collected from 41 and 49 m depths in the GS3 ice core, respectively. These "clean" reads in the four ice samples contained 169 bacterial genera, 70 of which had recognized names (Supplementary Table S3). The 13 most abundant genera, each of which accounted for >0.1% of sequences in at least one ice

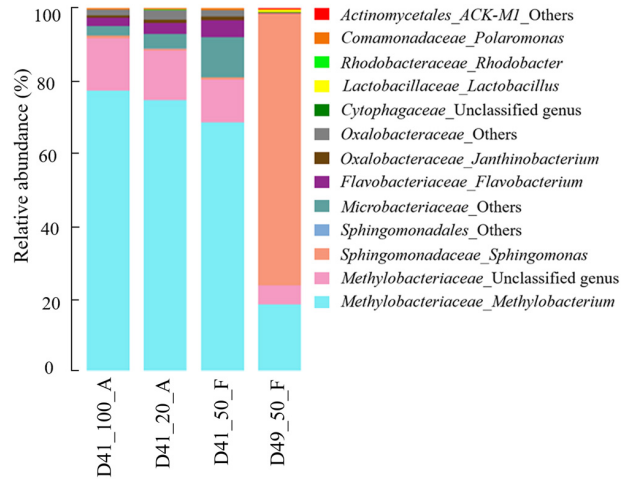


FIGURE 4 | Microbial community structure of the 13 most abundant genera in the GS3 ice core samples. The "others" represent unclassified sequences and could not be assigned to a single genus.

sample, comprised >98.5% of each decontaminated community. These genera were selected to illustrate the microbial community structures of the four ice samples (**Figure 4**). We also pre-amplified the targeted region (V4) of prokaryotic 16S rRNA genes by reconditioning PCR (Thompson et al., 2002) in all four of the ice samples before standard amplicon sequencing, as conducted for the "background" controls. The community compositions in each pair of original and pre-amplified ice samples were indistinguishable (Supplementary Figure S3). These results indicate that reliable microbial profile values were captured for both original and pre-amplified ice samples. The relative abundances of the microbial community among the three D41 samples (i.e., D41_100_A, D41_20_A, and D41_50_F) showed no significant difference from one another based on the results from the two-tailed paired *t*-test (*p*-values were 0.85–0.99, **Figure 4**). In addition, the results from ANOSIM analysis (together with their pre-amplified samples D41_100_A_28+8, D41_20_A_28+8, and D49_50_F_28+8 shown in Supplementary Figure S3) confirmed that the microbial communities in group samples (e.g., D41_100_A and D41_100_A_28+8 were a group) were not significantly different from one another (*p*-values were 0.66–0.99, n = 999). These results indicate that similar microbial profiles were captured from the ice samples regardless of the differences in sample volume and concentrating methods used for collecting cells and DNA extraction.

The genus *Methylobacterium* within the family *Methylobacteriaceae* was the most abundant taxon in the three D41 samples and had a relative abundance of 67.3–76.6%. An unclassified genus belonging to the same family, *Methylobacteriaceae*, was also abundant (relative abundance 11.7–14.3%) in these three samples (**Figure 4**). Members belonging to the genus *Methylobacterium* were also reported to dominate the microbial community in ancient ice cores from many previous studies (Miteva, 2008; Segawa et al., 2010; Antony et al., 2012; Miteva et al., 2015, 2016), including several microbial investigations of the Guliya ice cap ice cores using culture-dependent methods (Christner et al., 2000, 2001; Christner, 2003). Five other genera with relative abundances of 0.1–4.5%, which had recognized names, were also previously reported to be abundant in glacier ice cores, including the genera *Flavobacterium* (Liu et al., 2015; Chen et al., 2016), *Janthinobacterium* (Christner, 2003; Miteva, 2008), *Polaromonas* (Liu et al., 2009; An et al., 2010; Chen et al., 2016), *Sphingomonas* (An et al., 2010; Miteva et al., 2016), and *Rhodobacter* (Liu et al., 2015). The detection of bacterial sequences belonging to similar genera in ice core samples from different glaciers located around the world can be explained by the ubiquitous distribution of certain species in geographically distant environments (Baas Becking, 1934; Martiny et al., 2006). Furthermore, many *Methylobacterium* and *Sphingomonas* members are commonly found in tropospheric clouds and concentrated in cloud water (Amato et al., 2007; DeLeon-Rodriguez et al., 2013), which would allow them to be deposited onto the glaciers with falling snow.

For sample D49_50_F, the genus *Methylobacterium* and the unclassified genus (same as in D41 samples) within the family *Methylobacteriaceae* were also abundant making up 18.3 and 5.2%, respectively, of the total microbial population (**Figure 4**). The most abundant genus in this sample, however, was *Sphingomonas* with a relative abundance of 75.2%. Three other genera, including *Lactobacillus* and two unclassified genera in the phyla *Bacteroidetes* and *Actinobacteria*, accounted for 0.1–0.7% of the sequences (**Figure 4**). Thus there is a notable difference in the microbial profiles between samples D41 and D49. The results from ANOSIM analysis confirmed that the microbial communities were significantly different between samples from D41 and D49 (p = 0.04, n = 999).

Previous studies have often reported different microbial community structures in ice samples collected from different depths of the same ice core, and this probably reflects differences in the environmental conditions among ice samples (Priscu et al., 2007; Miteva et al., 2015; Liu et al., 2016). The D41 and D49 samples were obtained from depths of 41.10–41.84 and 49.51–49.90 m of the GS3 ice core, respectively (**Figure 1B** and Supplementary Table S4). These samples are approximately 20,000 and 30,000 years old, respectively (Supplementary Table S4), as determined by preliminary matching of the GS3 stable oxygen isotopes with those in a 1992 Guliya ice cap ice core (Thompson et al., 1997). The concentrations of nitrogen-related ions NO_3^- and NH_4^+ in D49 were lower than those in D41 while higher concentrations of dust and all other tested ions including Cl^-, SO_4^{2-}, Na^+, K^+, Mg^{2+}, and Ca^{2+} occurred in D49 (Supplementary Table S4). Variations in dust and ion concentrations are commonly found at different depths of an ice core (Thompson et al., 1997; Miteva et al., 2015) and they probably contribute to the differences in their microbial communities. For example, a study of microorganisms in a high Arctic glacier revealed sulfate-reducing bacteria from the basal ice-containing sulfate (Skidmore et al., 2000). Calcium concentrations positively correlated with bacterial abundance in an ice core retrieved from Mount Geladaindong on the Tibetan Plateau (Yao et al., 2008). Dust particle concentrations were reported to correlate with microbial concentrations in ice cores

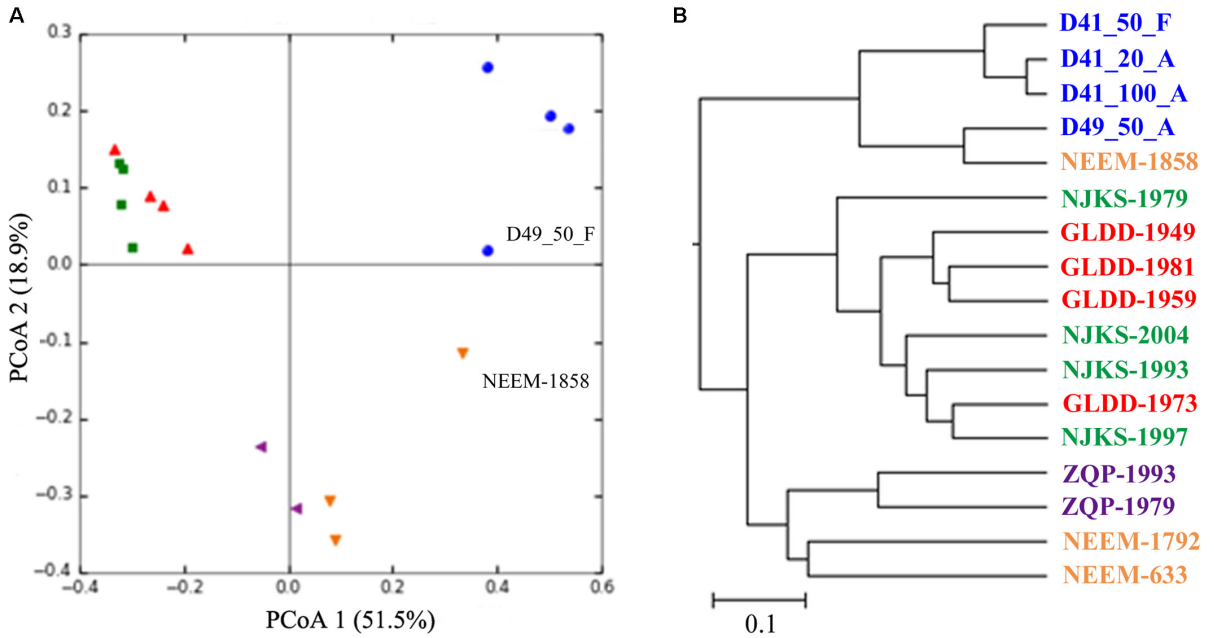


FIGURE 5 | Relationships between individual samples illustrated by PCoA plots **(A)** and UniFrac tree (UPGMA, **B**). Both analyses were performed on the basis of the weighted UniFrac metric. Symbols of the same color indicate samples from the same glacier/ice core: blue color, GS3; red, Geladangdong (GLDD) Glacier; green, Noijinkangsang (NJKS) Glacier; purple, Zuoqiupu (ZQP) Glacier; orange, North Greenland Eemian Ice Drilling (NEEM) ice core.

in many studies (e.g., Abyzov et al., 1998; Miteva et al., 2009; Segawa et al., 2010). Our results suggest that the differences in the microbial communities between samples D41 and D49 probably reflect the difference in the concentrations of dust and many ions in these samples, and that the GS3 ice core contains valuable information about changes in microbial communities over the past ∼30,000 years.

Microbial Community Clusters by Glacier

As noted above the Guliya ice cap is the highest, largest, thickest, and coldest ice cap among all the ice caps in middle–low latitude regions (Yao et al., 1992; Thompson et al., 1995). Considering the distinct characteristics of the Guliya ice cap, we next compared the microbial communities of the Guliya ice cap samples with those from four other glaciers. These samples were chosen because the microbial communities of all these glaciers were investigated with the overlapped region (V4) of 16S rRNA genes using a next-generation sequencing strategy. PCoA of the microbial communities of Guliya and four other ice cap samples indicated that the communities varied among the glaciers and that the communities could be clustered by the ice cap (**Figure 5A**). The first and second dimensions of PCoA showed that the distribution of all samples accounted for 51.5 and 18.9% of community variability, respectively. The weighted UniFrac tree (UPGMA) also showed that most of the samples from a given ice cap formed a lineage (**Figure 5B**). Samples from GLDD and NJKS glaciers clustered together, indicating a closer relationship of their microbial communities. This finding agrees with the original report and might be attributed to the fact that both NJKS and GLDD glaciers are strongly influenced by the same westerly jet stream (Liu et al., 2016). Samples from the Guliya ice cap formed a separate and distant cluster outside the other samples, indicating that the Guliya ice cap might contain more distinct microbial community relative to the other glaciers in the Tibetan Plateau and Greenland. Interestingly, sample NEEM-1858 clustered with sample D49_50_F from the Guliya ice cap but not with other NEEM samples from Greenland (**Figure 5B**). This result can be attributed to the fact that both samples were dominated by the genus *Sphingomonas* with a relative abundance of 94.4 and 75.5% for NEEM-1858 and D49, respectively (data not shown).

The Guliya ice cap also shared some bacterial groups with other more distant glaciers, which supports the perspective that microorganisms are distributed everywhere in the world (Baas Becking, 1934). The two-tailed Mantel test indicated that microbial community compositions correlated significantly (p = 0.04) with the age of ice samples, suggesting that the variation in microbial community composition among these ice samples probably reflects unique climate conditions at the time they were deposited. Unfortunately, the relationships between microbial community composition and environmental parameters were not investigated in this study, because of the absence of relevant data in the other studies. In addition, the other ice samples were collected and analyzed in three different projects and laboratories. Although the microbial communities of all samples were analyzed with the overlapped gene region using a next-generation sequencing strategy, the difference in other experimental steps and/or methods (e.g., ice core drilling, DNA isolation, and investigators) likely also influences the microbial communities reported. Conflicting results of microbial content

were reported in several papers investigating microorganisms in glacier ice (Willerslev et al., 2004). For example, the genus *Aquabacterium* was detected using 16S rRNA gene amplification and sequencing in the Lake Vostok ice samples (Christner et al., 2001), However, *Aquabacterium* was considered to be a contaminant because it was present in both the Lake Vostok ice sample and its negative control from another study (Priscu et al., 1999). As methodologies improve and cooperation increases among research groups around the world, it will be easier to compare ice core microbial communities generated from different laboratories and better understand the ecological implications of the ice microbial communities.

CONCLUSION

Microbial communities in glacier ice with low biomass have been studied previously. However, as a laboratory new to this science, we sought to establish robust "background" controls and *in silico* "contaminant" removal protocols for our work with low-biomass ice samples. Our effort expands prior work that established *in silico* contaminant removal procedures (Ram, 2009; Knowlton et al., 2013; Miteva et al., 2015; Cameron et al., 2016); however, our study also expands the number of control samples (i.e., a co-processed sterile water artificial ice core, air samples collected from the ice processing laboratories, and a blank, sterile water sample) to generate "clean" datasets for further analysis. We used this method to investigate the microbial communities in ice from two depths in a GS3 ice core and found that significantly different microbial profiles were archived. Unfortunately, glaciers around the world are rapidly shrinking primarily due to the warming of the atmosphere in response to increasing concentrations of greenhouse gases released during the burning of fossil fuels (Zemp et al., 2015; Burkhart et al., 2017). This will lead to a gradual loss of the microbial information archived in glacier ice from which past climate and environmental changes may be reconstructed. The "clean" protocol procedures introduced in this study can now be used to help investigate low-biomass microbial communities preserved in Earth's glaciers and ice caps. In addition, with further advancement of methods and technologies, such as metagenomics (Petrosino et al., 2009) and single-cell sequencing (Lasken, 2012), we will be able to better address microbial ecological questions for low-biomass, cold glacier ice, and bring microbial profiles into predictive ecological models of past climate changes in "frozen archive" environments.

AUTHOR CONTRIBUTIONS

Z-PZ, NS, MG, DK, EM-T, VR, JVE, LT, and MS conceived and designed the research, analyzed the data, and critically reviewed the manuscript. Z-PZ, NS, MG, and DK performed the laboratory experiments. Z-PZ, EM-T, VR, JVE, LT, and MS wrote the manuscript.

FUNDING

This study was supported by a Byrd Polar and Climate Research Center Postdoctoral Fellowship to Z-PZ, and by funding from NSF Paleoclimate Program award (No. 1502919) and the Chinese Academy of Sciences to LT, and a Gordon and Betty Moore Foundation Investigator Award (No. 3790) to MS.

ACKNOWLEDGMENTS

The authors greatly appreciate Dr. Karen Dannemiller and Mr. Quentin Platt for their help on air sampling; Dr. Mary Davis for providing information on environmental parameters of ice; Dr. Yueh-Fen Li for helpful suggestions on qPCR analysis; and the Sullivan, Thompsons, and Rich laboratories for critical review and comments through the years.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01094/full#supplementary-material

REFERENCES

Abyzov, S. S. (1993). "Microorganisms in the Antarctic ice," in *Antarctic Microbiology*, ed. E. I. Friedman (New York, NY: Wiley-Liss Press), 265–295.

Abyzov, S. S., Barkov, N. I., Bobin, N. E., Koudryashov, B. B., Lipenkov, V. Y., Mitskevich, I. N., et al. (1998). The ice sheet of central Antarctica as an object of study of past ecological events on the earth. *Izv. Akad. Nauk SSSR Biol.* 5, 610–616.

Abyzov, S. S., Bobin, N. E., and Koudryashov, B. B. (1982). Quantitative assessment of microorganisms in microbiological studies of Antarctic glaciers. *Biol. Bull. Acad. Sci. USSR* 9, 558–564.

Abyzov, S. S., Hoover, R. B., Imura, S., Mitskevich, I. N., Naganuma, T., Poglazova, M. N., et al. (2004). Use of different methods for discovery of ice-entrapped microorganisms in ancient layers of the Antarctic glacier. *Adv. Space Res.* 33, 1222–1230. doi: 10.1016/j.asr.2003.08.033

Amato, P., Parazols, M., Sancelme, M., Laj, P., Mailhot, G., and Delort, A.-M. (2007). Microorganisms isolated from the water phase of tropospheric clouds at the Puy de Dôme: major groups and growth abilities at low temperatures. *FEMS Microbiol. Ecol.* 59, 242–254. doi: 10.1111/j.1574-6941.2006.00199.x

An, L. Z., Chen, Y., Xiang, S. R., Shang, T. C., and Tian, L. D. (2010). Differences in community composition of bacteria in four glaciers in western China. *Biogeosciences* 7, 1937–1952. doi: 10.5194/bg-7-1937-2010

Antony, R., Krishnan, K. P., Laluraj, C. M., Thamban, M., Dhakephalkar, P. K., Engineer, A. S., et al. (2012). Diversity and physiology of culturable bacteria associated with a coastal Antarctic ice core. *Microbiol. Res.* 167, 372–380. doi: 10.1016/j.micres.2012.03.003

Baas Becking, L. G. M. (1934). *Geobiologie of Inleiding tot de Milieukunde*. The Hague: Van Stockum & Zoon.

Ball, M. M., Gomez, W., Magallanes, X., Rosales, R., Melfo, A., and Yarzabal, L. A. (2014). Bacteria recovered from a high-altitude, tropical glacier in Venezuelan Andes. *World J. Microbiol. Biotechnol.* 30, 931–941. doi: 10.1007/s11274-013-1511-1

Barton, H. A., Taylor, N. M., Lubbers, B. R., and Pemberton, A. C. (2006). DNA extraction from low-biomass carbonate rock: an improved method

with reduced contamination and the low-biomass contaminant database. *J. Microbiol. Methods* 66, 21–31. doi: 10.1016/j.mimet.2005.10.005

Biesbroek, G., Sanders, E. A., Roeselers, G., Wang, X., Caspers, M. P., Trzcinski, K., et al. (2012). Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS One* 7:e32942. doi: 10.1371/journal.pone.0032942

Burkhart, P. A., Alley, R. B., Thompson, L. G., Balog, J. D., Baldauf, P. E., and Baker, G. S. (2017). Savor the cryosphere. *GSA Today* 27, 4–11. doi: 10.1130/GSATG293A.1

Cameron, K. A., Stibal, M., Zarsky, J. D., Gozdereliler, E., Schostag, M., and Jacobsen, C. S. (2016). Supraglacial bacterial community structures vary across the Greenland ice sheet. *FEMS Microbiol. Ecol.* 92:fiv164. doi: 10.1093/femsec/fiv164

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/Ismej.2012.8

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Chen, Y., Li, X. K., Si, J., Wu, G. J., Tian, L. D., and Xiang, S. R. (2016). Changes of the bacterial abundance and communities in shallow ice cores from Dunde and Muztagata glaciers, Western China. *Front. Microbiol.* 7:1716. doi: 10.3389/fmicb.2016.01716

Christner, B. C. (2003). *Detection, Recovery, Isolation, and Characterization of Bacteria in Glacial Ice and Lake Vostok Accretion Ice*. Ph.D. thesis, Ohio State University, Columbus, OH.

Christner, B. C., Mikucki, J. A., Foreman, C. M., Denson, J., and Priscu, J. C. (2005). Glacial ice cores: a model system for developing extraterrestrial decontamination protocols. *Icarus* 174, 572–584. doi: 10.1016/j.icarus.2004.10.027

Christner, B. C., Mosley-Thompson, E., Thompson, L. G., and Reeve, J. N. (2001). Isolation of bacteria and 16S rDNAs from Lake Vostok accretion ice. *Environ. Microbiol.* 3, 570–577. doi: 10.1046/j.1462-2920.2001.00226.x

Christner, B. C., Mosley-Thompson, E., Thompson, L. G., and Reeve, J. N. (2003). Bacterial recovery from ancient glacial ice. *Environ. Microbiol.* 5, 433–436. doi: 10.1046/j.1462-2920.2003.00422.x

Christner, B. C., Mosley-Thompson, E., Thompson, L. G., Zagorodnov, V., Sandman, K., and Reeve, J. N. (2000). Recovery and identification of viable bacteria immured in glacial ice. *Icarus* 144, 479–485. doi: 10.1006/icar.1999.6288

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

Corless, C. E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarski, E. B., and Fox, A. J. (2000). Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *J. Clin. Microbiol.* 38, 1747–1752.

Cui, X. F., and Graf, H. F. (2009). Recent land cover changes on the Tibetan Plateau: a review. *Clim. Change* 94, 47–61. doi: 10.1007/s10584-009-9556-8

Dang, V. T., Howard-Varona, C., Schwenck, S., and Sullivan, M. B. (2015). Variably lytic infection dynamics of large *Bacteroidetes* podovirus phi38:1 against two *Cellulophaga baltica* host strains. *Environ. Microbiol.* 17, 4659–4671. doi: 10.1111/1462-2920.13009

Darling, C. A., and Siple, P. A. (1941). Bacteria of Antarctica. *J. Bacteriol.* 42, 83–98.

Davis, M. E., and Thompson, L. G. (2006). Andean ice core record of a Middle Holocene Mega-drought in North Africa and the Middle East. *Ann. Glaciol.* 43, 34–41. doi: 10.3189/172756406781812456

DeLeon-Rodriguez, N., Lathem, T. L., Rodriguez, R. L., Barazesh, J. M., Anderson, B. E., Beyersdorf, A. J., et al. (2013). Microbiome of the upper troposphere: species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2575–2580. doi: 10.1073/pnas.1212089110

D'Elia, T., Veerapaneni, R., and Rogers, S. O. (2008). Isolation of microbes from Lake Vostok accretion ice. *Appl. Environ. Microbiol.* 74, 4962–4965. doi: 10.1128/AEM.02501-07

Deng, L., Ignacio-Espinoza, J. C., Gregory, A. C., Poulos, B. T., Weitz, J. S., Hugenholtz, P., et al. (2014). Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* 513, 242–245. doi: 10.1038/nature13459

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x

Duhaime, M. B., Deng, L., Poulos, B. T., and Sullivan, M. B. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* 14, 2526–2537. doi: 10.1111/j.1462-2920.2012.02791.x

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Fountain, A. G., Campbell, J. L., Schuur, E. A. G., Stammerjohn, S. E., Williams, M. W., and Ducklow, H. W. (2012). The Disappearing cryosphere: impacts and ecosystem responses to rapid cryosphere loss. *Bioscience* 62, 405–415. doi: 10.1525/j.bio.2012.62.4.11

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., and Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* 8:24. doi: 10.1186/s13099-016-0103-7

Gonzalez-Toril, E., Santofimia, E., Blanco, Y., Lopez-Pamo, E., Gomez, M. J., Bobadilla, M., et al. (2015). Pyrosequencing-based assessment of the microbial community structure of Pastoruri Glacier area (Huascarán National Park, Peru), a natural extreme acidic environment. *Microb. Ecol.* 70, 936–947. doi: 10.1007/s00248-015-0634-3

Grahn, N., Olofsson, M., Ellnebo-Svedlund, K., Monstein, H. J., and Jonasson, J. (2003). Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiol. Lett.* 219, 87–91. doi: 10.1016/S0378-1097(02)01190-4

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi: 10.1101/gr.112730.110

Hebsgaard, M. B., Phillips, M. J., and Willerslev, E. (2005). Geologically ancient DNA: fact or artefact? *Trends Microbiol.* 13, 212–220. doi: 10.1016/j.tim.2005.03.010

Immerzeel, W. W., van Beek, L. P., and Bierkens, M. F. (2010). Climate change will affect the Asian water towers. *Science* 328, 1382–1385. doi: 10.1126/science.1183188

Karl, D. M., Bird, D. F., Bjorkman, K., Houlihan, T., Shackelford, R., and Tupas, L. (1999). Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* 286, 2144–2147. doi: 10.1126/science.286.5447.2144

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., et al. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763. doi: 10.1038/nmeth.1650

Knowlton, C., Veerapaneni, R., D'Elia, T., and Rogers, S. O. (2013). Microbial analyses of ancient ice core sections from Greenland and Antarctica. *Biology* 2, 206–232. doi: 10.3390/biology2010206

Kolde, R. (2015). *pheatmap: Pretty Heatmaps. R Package Version 1.0.8*. Available at: http://cran.r-project.org/web/packages/pheatmap/index.html

Kwok, S., and Higuchi, R. (1989). Avoiding false positives with PCR. *Nature* 339, 237–238. doi: 10.1038/339237a0

Lasken, R. S. (2012). Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* 10, 631–640. doi: 10.1038/nrmicro2857

Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., et al. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 4:29. doi: 10.1186/s40168-016-0172-3

Laurence, M., Hatzis, C., and Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One* 9:e97876. doi: 10.1371/journal.pone.0097876

Lazarevic, V., Gaia, N., Emonet, S., Girard, M., Renzi, G., Despres, L., et al. (2014). Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front. Cell. Infect. Microbiol.* 4:65. doi: 10.3389/fcimb.2014.00065

Lazarevic, V., Gaia, N., Girard, M., and Schrenzel, J. (2016). Decontamination of 16S rRNA gene amplicon sequence datasets based on bacterial load assessment by qPCR. *BMC Microbiol.* 16:73. doi: 10.1186/s12866-016-0689-4

Lenz, T. L., and Becker, S. (2008). Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci–implications for evolutionary analysis. *Gene* 427, 117–123. doi: 10.1016/j.gene.2008.09.013

Liu, Q., Zhou, Y. G., and Xin, Y. H. (2015). High diversity and distinctive community structure of bacteria on glaciers in China revealed by 454 pyrosequencing. *Syst. Appl. Microbiol.* 38, 578–585. doi: 10.1016/j.syapm.2015.09.005

Liu, Y., Priscu, J. C., Yao, T., Vick-Majors, T. J., Xu, B., Jiao, N., et al. (2016). Bacterial responses to environmental change on the Tibetan Plateau over the past half century. *Environ. Microbiol.* 18, 1930–1941. doi: 10.1111/1462-2920.13115

Liu, Y., Yao, T., Jiao, N., Kang, S., Xu, B., Zeng, Y., et al. (2009). Bacterial diversity in the snow over Tibetan Plateau Glaciers. *Extremophiles* 13, 411–423. doi: 10.1007/s00792-009-0227-5

Maccario, L., Sanguino, L., Vogel, T. M., and Larose, C. (2015). Snow and ice ecosystems: not so extreme. *Res. Microbiol.* 166, 782–795. doi: 10.1016/j.resmic.2015.09.002

Martiny, J. B., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341

McLean, A. L. (1919). Bacteria of ice and snow in Antarctica. *Nature* 102, 35–39. doi: 10.1038/102035a0

Miteva, V. (2008). "Bacteria in snow and glacier ice," in *Psychrophiles: From Biodiversity to Biotechnology*, eds R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday (Berlin: Springer), 31–50. doi: 10.1007/978-3-540-74335-4_3

Miteva, V., Rinehold, K., Sowers, T., Sebastian, A., and Brenchley, J. (2015). Abundance, viability and diversity of the indigenous microbial populations at different depths of the NEEM Greenland ice core. *Polar Res.* 34:25057. doi: 10.3402/polar.v34.25057

Miteva, V., Sowers, T., Schupbach, S., Fischer, H., and Brenchley, J. (2016). Geochemical and microbiological studies of nitrous oxide variations within the new NEEM Greenland ice core during the last glacial period. *Geomicrobiol. J.* 33, 647–660. doi: 10.1080/01490451.2015.1074321

Miteva, V., Teacher, C., Sowers, T., and Brenchley, J. (2009). Comparison of the microbial diversity at different depths of the GISP2 Greenland ice core in relationship to deposition climates. *Environ. Microbiol.* 11, 640–656. doi: 10.1111/j.1462-2920.2008.01835.x

Miteva, V. I., Sheridan, P. P., and Brenchley, J. E. (2004). Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Appl. Environ. Microbiol.* 70, 202–213. doi: 10.1128/AEM.70.1.202-213.2004

Morgan, J. L., Darling, A. E., and Eisen, J. A. (2010). Metagenomic sequencing of an *in vitro*-simulated microbial community. *PLoS One* 5:e10209. doi: 10.1371/journal.pone.0010209

Mumy, K. L., and Findlay, R. H. (2004). Convenient determination of DNA extraction efficiency using an external DNA recovery standard and quantitative-competitive PCR. *J. Microbiol. Methods* 57, 259–268. doi: 10.1016/j.mimet.2004.01.013

Noble, R. T., and Fuhrman, J. A. (1998). Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* 14, 113–118. doi: 10.3354/ame014113

Othman, A. (2015). Isolation and microbiological identification of bacterial contaminants in food and household surfaces: how to deal safely. *Egypt. Pharm. J.* 14, 50–55. doi: 10.4103/1687-4315.154720

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., and Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clin. Chem.* 55, 856–866. doi: 10.1373/clinchem.2008.107565

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Priscu, J. C., Adams, E. E., Lyons, W. B., Voytek, M. A., Mogk, D. W., Brown, R. L., et al. (1999). Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science* 286, 2141–2144. doi: 10.1126/science.286.5447.2141

Priscu, J. C., Christner, B. C., Foreman, C. M., and Royston-Bishop, G. (2007). "Biological material in ice cores," in *Encyclopedia of Quaternary Sciences*, 2nd Edn, ed. S. A. Elias (Amsterdam: Elsevier), 1156–1166.

Priscu, J. C., Fritsen, C. H., Adams, E. E., Giovannoni, S. J., Paerl, H. W., McKay, C. P., et al. (1998). Perennial Antarctic lake ice: an oasis for life in a polar desert. *Science* 280, 2095–2098. doi: 10.1126/science.280.5372.2095

Qiu, J. (2008). China: the third pole. *Nature* 454, 393–396. doi: 10.1038/454393a

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ram, V. (2009). *Analysis and Characterization of Microbes from Ancient Glacial Ice*. Ph.D. thesis, Bowling Green State University, Bowling Green, OH.

Rogers, S. O., Theraisnathan, V., Ma, L. J., Zhao, Y., Zhang, G., Shin, S. G., et al. (2004). Comparisons of protocols for decontamination of environmental ice samples for biological and molecular examinations. *Appl. Environ. Microbiol.* 70, 2540–2544. doi: 10.1128/Aem.70.4.2540-2544.2004

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12:87. doi: 10.1186/s12915-014-0087-z

Santibanez-Avila, P. A. (2016). *Factors Influencing the Abundance of Microorganisms in Icy Environments*. Ph.D. thesis, Montana State University, Bozeman, MT.

Segawa, T., Ushida, K., Narita, H., Kanda, H., and Kohshima, S. (2010). Bacterial communities in two Antarctic ice cores analyzed by 16S rRNA gene sequencing analysis. *Polar Sci.* 4, 215–227. doi: 10.1016/j.polar.2010.05.003

Sheridan, P. P., Miteva, V. I., and Brenchley, J. E. (2003). Phylogenetic analysis of anaerobic psychrophilic enrichment cultures obtained from a greenland glacier ice core. *Appl. Environ. Microbiol.* 69, 2153–2160. doi: 10.1128/AEM.69.4.2153-2160.2003

Skidmore, M. L., Foght, J. M., and Sharp, M. J. (2000). Microbial life beneath a high arctic glacier. *Appl. Environ. Microbiol.* 66, 3214–3220. doi: 10.1128/AEM.66.8.3214-3220.2000

Solonenko, S. A., Ignacio-Espinoza, J. C., Alberti, A., Cruaud, C., Hallam, S., Konstantinidis, K., et al. (2013). Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* 14:320. doi: 10.1186/1471-2164-14-320

Solonenko, S. A., and Sullivan, M. B. (2013). Preparation of metagenomic libraries from naturally occurring marine viruses. *Methods Enzymol.* 531, 143–165. doi: 10.1016/B978-0-12-407863-5.00008-3

Steven, B., Pollard, W. H., Greer, C. W., and Whyte, L. G. (2008). Microbial diversity and activity through a permafrost/ground ice core profile from the Canadian high Arctic. *Environ. Microbiol.* 10, 3388–3403. doi: 10.1111/j.1462-2920.2008.01746.x

Tanner, M. A., Goebel, B. M., Dojka, M. A., and Pace, N. R. (1998). Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* 64, 3110–3113.

Thompson, J. R., Marcelino, L. A., and Polz, M. F. (2002). Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* 30, 2083–2088. doi: 10.1093/nar/30.9.2083

Thompson, L. G., Mosley-Thompson, E., Davis, M. E., Lin, P. N., Dai, J., Bolzan, J. F., et al. (1995). A 1000 year ice core climate record from the Guliya ice cap, China: its relationship to global climate variability. *Ann. Glaciol.* 21, 175–181. doi: 10.1017/S0260305500015780

Thompson, L. G., Yao, T., Davis, M. E., Henderson, K. A., Mosley-Thompson, E., Lin, P. N., et al. (1997). Tropical climate instability: the last glacial cycle from a Qinghai-Tibetan ice core. *Science* 276, 1821–1825. doi: 10.1126/science.276.5320.1821

Thompson, L. G., Yao, T., Mosley-Thompson, E., Davis, M. E., Henderson, K. A., and Lin, P. (2000). A high-resolution millennial record of the south Asian monsoon from Himalayan ice cores. *Science* 289, 1916–1920. doi: 10.1126/science.289.5486.1916

Tung, H. C., Bramall, N. E., and Price, P. B. (2005). Microbial origin of excess methane in glacial ice and implications for life on Mars. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18292–18296. doi: 10.1073/pnas.0507601102

Tung, H. C., Price, P. B., Bramall, N. E., and Vrdoljak, G. (2006). Microorganisms metabolizing on clay grains in 3-km-deep Greenland basal ice. *Astrobiology* 6, 69–86. doi: 10.1089/ast.2006.6.69

Vanwonterghem, I., Jensen, P. D., Dennis, P. G., Hugenholtz, P., Rabaey, K., and Tyson, G. W. (2014). Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *ISME J.* 8, 2015–2028. doi: 10.1038/ismej.2014.50

Wang, C. H., Chen, B. T., Han, B. C., Liu, A. C., Hung, P. C., Chen, C. Y., et al. (2015). Field evaluation of personal sampling methods for multiple bioaerosols. *PLoS One* 10:e0120308. doi: 10.1371/journal.pone.0120308

Wang, N. L., Yao, T. D., Thompson, L. G., Henderson, K. A., and Davis, M. E. (2002). Evidence for cold events in the early Holocene from the Guliya ice core, Tibetan Plateau, China. *Chin. Sci. Bull.* 47, 1422–1427. doi: 10.1360/02tb9313

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/Aem.00062-07

Willerslev, E., and Cooper, A. (2005). Ancient DNA. *Proc. R. Soc. B Biol. Sci.* 272, 3–16. doi: 10.1098/rspb.2004.2813

Willerslev, E., Hansen, A. J., and Poinar, H. N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 19, 141–147. doi: 10.1016/j.tree.2003.11.010

Willner, D., Daly, J., Whiley, D., Grimwood, K., Wainwright, C. E., and Hugenholtz, P. (2012). Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One* 7:e34605. doi: 10.1371/journal.pone.0034605

Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., et al. (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS One* 6:e26161. doi: 10.1371/journal.pone.0026161

Wu, G. J., Yao, T. D., Thompson, L. G., and Li, Z. Q. (2004). Microparticle record in the Guliya ice core and its comparison with polar records since the last interglacial. *Chin. Sci. Bull.* 49, 607–611. doi: 10.1360/03wd0419

Xiang, S. R., Shang, T. C., Chen, Y., and Yao, T. D. (2009). Deposition and postdeposition mechanisms as possible drivers of microbial population variability in glacier ice. *FEMS Microbiol. Ecol.* 70, 165–176. doi: 10.1111/j.1574-6941.2009.00759.x

Yao, T. D., Jiao, K. Q., Zhang, X. P., Yang, Z. H., and Thompson, L. G. (1992). Glaciologic studies on Guliya ice cap. *J. Glaciol. Geocryol.* 14, 233–241.

Yao, T. D., Liu, Y. Q., Kang, S. C., Jiao, N. Z., Zeng, Y. H., Liu, X. B., et al. (2008). Bacteria variabilities in a Tibetan ice core and their relations with climate change. *Glob. Biogeochem. Cycles* 22:Gb4017. doi: 10.1029/2007GB003140

Yao, T. D., Wu, G. J., Pu, J. C., Jiao, K. Q., and Huang, C. L. (2004). Relationship between calcium and atmospheric dust recorded in Guliya ice core. *Chin. Sci. Bull.* 49, 706–710. doi: 10.1360/03wd0417

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., and Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7:e33865. doi: 10.1371/journal.pone.0033865

Zemp, M., Frey, H., Gärtner-Roer, I., Nussbaumer, S. U., Hoelzle, M., Paul, F., et al. (2015). Historically unprecedented global glacier decline in the early 21st century. *J. Glaciol.* 61, 745–762. doi: 10.3189/2015JoG15J017

Zhang, X. F., Yao, T. D., Tian, L. D., Xu, S. J., and An, L. Z. (2008). Phylogenetic and physiological diversity of bacteria isolated from Puruogangri ice core. *Microb. Ecol.* 55, 476–488. doi: 10.1007/s00248-007-9293-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhong, Solonenko, Gazitúa, Kenny, Mosley-Thompson, Rich, Van Etten, Thompson and Sullivan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

METHODS
published: 05 June 2018
doi: 10.3389/fmicb.2018.01172

viGEN: An Open Source Pipeline for the Detection and Quantification of Viral RNA in Human Tumors

Krithika Bhuvaneshwar, Lei Song, Subha Madhavan and Yuriy Gusev**

Innovation Center for Biomedical Informatics, Georgetown University, Washington, DC, United States

An estimated 17% of cancers worldwide are associated with infectious causes. The extent and biological significance of viral presence/infection in actual tumor samples is generally unknown but could be measured using human transcriptome (RNA-seq) data from tumor samples. We present an open source bioinformatics pipeline viGEN, which allows for not only the detection and quantification of viral RNA, but also variants in the viral transcripts. The pipeline includes 4 major modules: The first module aligns and filter out human RNA sequences; the second module maps and count (remaining un-aligned) reads against reference genomes of all known and sequenced human viruses; the third module quantifies read counts at the individual viral-gene level thus allowing for downstream differential expression analysis of viral genes between case and controls groups. The fourth module calls variants in these viruses. To the best of our knowledge, there are no publicly available pipelines or packages that would provide this type of complete analysis in one open source package. In this paper, we applied the viGEN pipeline to two case studies. We first demonstrate the working of our pipeline on a large public dataset, the TCGA cervical cancer cohort. In the second case study, we performed an in-depth analysis on a small focused study of TCGA liver cancer patients. In the latter cohort, we performed viral-gene quantification, viral-variant extraction and survival analysis. This allowed us to find differentially expressed viral-transcripts and viral-variants between the groups of patients, and connect them to clinical outcome. From our analyses, we show that we were able to successfully detect the human papilloma virus among the TCGA cervical cancer patients. We compared the viGEN pipeline with two metagenomics tools and demonstrate similar sensitivity/specificity. We were also able to quantify viral-transcripts and extract viral-variants using the liver cancer dataset. The results presented corresponded with published literature in terms of rate of detection, and impact of several known variants of HBV genome. This pipeline is generalizable, and can be used to provide novel biological insights into microbial infections in complex diseases and tumorigeneses. Our viral pipeline could be used in conjunction with additional type of immuno-oncology analysis based on RNA-seq data of host RNA for cancer immunology applications. The source code, with example data and tutorial is available at: https://github.com/ICBI/viGEN/.

Keywords: RNA-seq, viral detection, liver cancer, TCGA, variant analysis, next-generation sequencing, cancer immunology

OPEN ACCESS

Edited by:
Diana Elizabeth Marco, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Reviewed by:
João Marcelo Pereira Alves, Universidade de São Paulo, Brazil
Hetron Mweemba Munang'andu, Norwegian University of Life Sciences, Norway

****Correspondence:***
Krithika Bhuvaneshwar
kb472@georgetown.edu
Yuriy Gusev
yg63@georgetown.edu

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *26 January 2018*
Accepted: *15 May 2018*
Published: *05 June 2018*

Citation:
Bhuvaneshwar K, Song L, Madhavan S and Gusev Y (2018) viGEN: An Open Source Pipeline for the Detection and Quantification of Viral RNA in Human Tumors. Front. Microbiol. 9:1172. doi: 10.3389/fmicb.2018.01172

INTRODUCTION

An estimated 17% of cancers worldwide are associated with infectious causes. These infectious agents include viruses, bacteria, parasites and other microbes. Examples of viruses include human papilloma viruses (HPVs) in cervical cancer, epstein-Barr virus (EBV) in nasopharyngeal cancer, hepatitis B and C in liver cancer (HBV and HCV), human herpes virus 8 (HHV-8) in Kaposi sarcoma (KS); human T-lymphotrophic virus-1 (HTLV-1) in adult T cell lymphocytic leukemia (ATL) and non-Hodgkin lymphoma; merkel cell polyomavirus (MCV) in Merkel cell carcinoma (ACS, 2007). Bacteria such as *Helicobacter pylori* have been implicated in stomach cancer. Parasites have also been associated with cancer, examples are *Opisthorchis viverrini* and *Clonorchis sinensis* in bile duct cancer and *Schistosoma haematobium* in bladder cancer (ACS, 2007). Detection and characterization of these infectious agents in tumor samples can give us better insights into disease mechanisms and their treatment (Hausen, 2000).

Vaccines have been developed to help protect against infection from the many cancers. But these vaccines can only be used to help prevent infection and cannot treat existing infections (ACS, 2007). There are several screening methods widely used to detect viral infections, especially for blood borne viruses including HBV, HCV, HIV and HTLV. These include the enzyme linked immunosorbent assay (ELISA or EIA) (Yoshihara, 1995), chemluminescent immunoassay (ChLIA), Indirect fluorescent antibody (IFA), Western blot (WB), Polymerase Chain Reaction (PCR), and Rapid immunoassays[1]. ELISA and WB test detects and measures antibodies in serum taken from the patient's blood, and are typically prescribed after certain symptoms are observed in the patient.

There are several challenges in detection of viruses in tumors including loss of viral information in progressed tumors and limited or latent replication resulting in low transcription of tumors (Schelhorn et al., 2013). The extent and biological significance of viral presence/infection in actual tumor samples is generally unknown but could be measured using human transcriptome data from tumor samples.

The popularity of next-generation sequencing (NGS) technology has exploded in the last decade. NGS technologies are able to perform rapid sequencing, and in a massively parallel fashion (Datta et al., 2015). In recent years, applications of NGS technologies in clinical diagnostics have been on the rise (Barzon et al., 2011; Byron et al., 2016). Amongst the various NGS technologies, whole-transcriptome sequencing, also called RNA-seq, has been very popular with methods and tools being actively developed. Exploring the genome using RNA-seq gives a different insight than looking at the DNA since the RNA-seq would have captured actively transcribed regions. Every aspect of data output from this technology is now being used for research, including detection of viruses and bacteria (Khoury et al., 2013; Salyakina and Tsinoremas, 2013; Wang et al., 2016). They are also independent of prior sequence information, and require less starting material compared to conventional cloning based methods, making them powerful and exciting new technologies in virology (Datta et al., 2015). These high throughput technologies give us direct evidence of infection in the tissue as compared to ELISA-based assays, which only proves presence of infection somewhere in the human body. RNA-seq technology has hence enabled the exploration and detection of viral infections in human tumor samples. This technology also enables detection of variants in viral genome, which have been connected to clinical outcome (Moyes et al., 2005; Downey et al., 2015).

In recent years, US regulators approved a viral based cancer therapy (Ledford, 2015), proving that the study of viruses in the human transcriptome has biomedical interest, and is paving the way for promising research and new opportunities.

In this paper, we present our pipeline viGEN to not only detect and quantify read counts at the individual viral-gene level, but also detect viral variants from human RNA-seq data. The characterization of viral variants helps enable better epidemiological analysis. The input file to our pipeline is a fastq (Wikipedia, 2009) file, so our viGEN pipeline can be extended to work with genomic data from any NGS technology. Our pipeline can also be used to detect and explore not only viruses, but other microbes as well, as long as the sequence information is available in NCBI[2].

We applied our viGEN pipeline to two case studies as a proof of concept - a dataset of 304 cervical cancer patients, and a set of 50 liver cancer patients, both from the TCGA collection. We first applied the pipeline to the transcriptome of cervical cancer patients to see if we are able to detect the human papilloma viruses. We also performed additional in-depth analyses on a small focused study of liver cancer patients. In this cohort, we performed viral-gene quantification, viral-variant extraction and survival analysis.

From our analyses, we show that we were able to successfully detect the human papilloma virus among the TCGA cervical cancer patients. We compared the viGEN pipeline with two metagenomics tools and demonstrate similar sensitivity/specificity. We were also able to quantify viral-transcripts and extract viral-variants using the liver cancer dataset. This enabled us to perform downstream analysis to give us new insights into disease mechanisms.

In addition to the two case studies, we have made available an end-to-end tutorial demonstrated on a publicly available

[1]FDA *Complete List of Donor Screening Assays for Infectious Agents and HIV Diagnostic Assays* (Accessed March 05, 2016). Available online at: https://www.fda.gov/biologicsbloodvaccines/bloodbloodproducts/approvedproducts/licensedproductsblas/blooddonorscreening/infectiousdisease/ucm080466.htm

Abbreviations: HBV, Hepatitis B virus; HCV, Hepatitis C Virus; HERV K113, Human Endogenous Retrovirus K113; TCGA, The Cancer Genome Atlas; HCC, Hepatocellular carcinoma; NAFLD, nonalcoholic fatty liver disease; Hep B, Hepatitis B; Hep C, Hepatitis C; HepB + HepC, coinfected with both Hepatitis B and C virus; HBsAg, Hepatitis B surface antigen; HBeAg, Hepatitis B type e antigen; NGS, next-generation sequencing; RNA-seq, whole transcriptome sequencing; BAM, Binary version of Sequence alignment/map format; CDS, coding sequence; Cox PH, Cox Proportional Hazard; HBx, viral gene X; STS, Sequence-tagged sites; NCBI, National Center for Biotechnology Information; GFF, general-feature-format.

[2]NCBI (2009). *NCBI FTP Site for Viruses.* Available online at: ftp://ftp.ncbi.nlm.nih.gov/genomes/Viruses/

TABLE 1 | Comparison of existing pipelines that detect viruses from human transcriptome data.

Tool name	Detect viruses from Human RNA-seq data	Perform quantification at viral-gene/CDS level	Works on DNAseq, RNAseq or both	Variant calling at viral-variant level	Discover viral integration sites	Other comments
Virana (Schelhorn et al., 2013)	Yes	Identifies microbial transcripts, does not quantify	Both	No	Yes	Also offers analysis of homologs
VirusSeq (Chen et al., 2013)	Yes	No	Both	No	Yes	Pre-designed to work on a select set of 18 viruses
Viral Fusion Seq (Li et al., 2013)	Yes	No	Both	No	Yes	Can also detect fusion events
Virus Finder (Wang et al., 2013)	Yes	No	Both	No	Yes	Can be applied to samples infected with undiagnosed viruses
PathSeq (Kostic et al., 2011)	Yes	No	Both	No	No	
RINS (Bhaduri et al., 2012)	Yes	No	RNA-seq	No	No	Generates contigs with these non-human sequences
Kraken (Wood and Salzberg, 2014)	Yes	No	No	No	No	Metagenomic analysis tool
Cetrifuge (Kim et al., 2016)	Yes	No	No	No	No	Metagenomic analysis tool
viGEN (our pipeline)	Yes	Yes	Both	Yes	No	

RNA-seq sample from an HBV liver cancer patient from NCBI SRA (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA279878). We also provided step-by-step instructions on how to run our viGEN pipeline on this sample data, along with the code at https://github.com/ICBI/viGEN/ and demonstrated the detection of HBV transcripts in this sample. This allows other users to apply this pipeline to explore viruses in their data and disease of interest. We are currently implementing the viGEN pipeline in the Seven Bridges Cancer Genomics Cloud[3].

There are a number of existing pipelines that detect viruses from human transcriptome data. Of these, very few pipelines offer quantification at the gene expression level. A comprehensive comparison of these pipelines is provided in **Table 1**. Our goal was not to compete with these other tools, but to offer a convenient and complete end–to-end publicly available pipeline to the bioinformatics community. To the best of our knowledge there are no publicly available pipelines or packages that would provide this type of complete analysis in one package. Customized solutions have been reported in the literature however were not made public.

In the future, our plan is to package this pipeline and make it available to users through Bioconductor (Lawrence et al., 2013), allowing users to perform analysis on either their local computer or the cloud.

MATERIALS AND METHODS

In this paper, we applied our viGEN pipeline to two case studies as a proof of concept - a dataset of 304 cervical cancer patients, and a set of 50 liver cancer patients, both from the TCGA collection (NCI, 2011). We first applied the pipeline to the transcriptome of cervical cancer patients to see if we are able to detect the human papilloma viruses. We also performed additional in-depth analyses on a small focused study of liver cancer patients afflicted with Hepatitis B virus. In this cohort, we perform viral-gene quantification, viral-variant extraction and survival analysis. The results from these analyses allowed us to compare experimental and control groups using viral-gene expression data and viral-variant data, and give us insights into their impacts on the tumor, and disease mechanisms.

In the following sections, we describe the viGEN pipeline, and the two case studies.

The viGEN Pipeline

The viGEN pipeline includes 4 major modules. **Figure 1** shows an image of our viGEN pipeline.

Module 1: Viral Genome Level Analysis (Filtered Human Sample Input)

In Module 1 (labeled as "filtered human sample input"), the human RNA sequences were aligned to the human-reference genome using the RSEM (Li and Dewey, 2011) tool. One of the outputs of RSEM includes sequences that did not align to the human genome (hence the name "filtered human sample input"). These un-aligned sequences were taken and aligned to the viral reference file using popular alignment tools BWA (Li and Durbin, 2009) and Bowtie2 (Langmead and Salzberg, 2012).

Module 2: Viral Genome Level Analysis (Unfiltered Human Sample Input)

In Module 2 (labeled as "unfiltered human sample input"), the RNA seq sequences were directly aligned to the viral reference using Bowtie2 without any filtering.

[3] *Seven Bridges Cancer Genomics Cloud.* Available online at: https://cgc.sbgenomics.com/

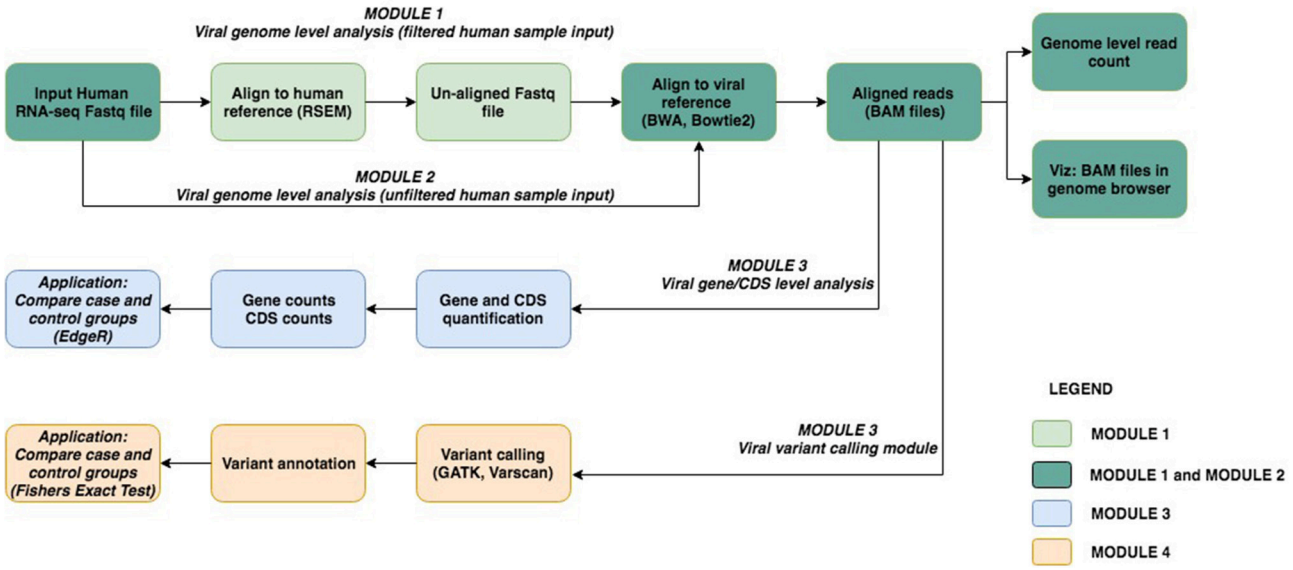


FIGURE 1 | viGEN pipeline. Each module has a color, shown in the legend.

The reason for using two methods to obtain the viral genomes in human RNA-seq data (Module 1 and Module 2) was to allow us to be as comprehensive as possible in viral detection.

The aligned reads from Module 1 and 2 were in the form of BAM files (Center-for-Statistical-Genetics, 2013), from which read counts were obtained for each viral genome species (referred to as "genome level counts") using Samtools idxstats (Li et al., 2009) or Picard BAMIndexStats[4] tools. Using the genome level counts, we estimated the number of reads that covered the genome, a form of viral copy number. Viral copy number was defined as in equation below:

$$Viral\ copy\ number\ = \frac{Number\ of\ mapped\ reads\ \times\ Read\ length}{Genome\ length}$$

Only those viral species with copy number more than a threshold are selected for the next module.

Module 3: Viral Gene Expression Analysis

The BAM files from Module 1 and 2 (from Bowtie2 and BWA) were input into Module 3 (referred to as "viral gene expression level analysis"), which calculated quantitate read counts at the individual viral-gene level. We found existing RNA-seq quantification tools to be not sensitive enough for viruses, and hence developed our own algorithm for this module. Our in-house algorithm used region-based information from the general-feature-format (GFF) files[5] of each viral genome, and the reads from the BAM file. It created a summary file, which had a total count of reads within or on the boundary of each region in the GFF file. This is repeated for each sample and for each viral GFF file. At the end, a matrix is obtained where the features (rows) are regions from the GFF file, and the columns are samples. The read count output from Module 3 (viral gene expression module) allowed for downstream differential expression analysis of viral genes between case and controls groups. The source code for our in-house algorithm, written using the R programming language (R Core Team, 2014), has been made public at available at github.com/ICBI/viGEN.

Module 4: Viral RNA Variant Calling Module

The BAM files from Module 1 and 2 (from Bowtie2) were also input to Module 4 to detect mutations in the transcripts from these viruses (referred to as "viral RNA variant calling module"). The BAM files were first sorted coordinate-wise using Samtools (Li et al., 2009); PCR duplicates were removed using tool Picard[4], then the chromosomes in the BAM file were ordered in the same way as the reference file using Picard. The Viral reference file was created from combining all known and sequenced human viruses obtained from NCBI[2]. Because viral variants are known to be low frequency, we have selected a variant calling tool Varscan2 (Koboldt et al., 2012), which allows detection of low-frequency variants (Spencer et al., 2014). Low quality and low depth variants were flagged, but not filtered out, in case these low values were due to low viral load. Once the variants were obtained, they were merged to form a multi-sample VCF file. Only variants that had a variant in two or more samples were retained. PLINK (Chang et al., 2015) was used to perform case-control association test (Fishers Exact Test) to compare groups.

Tutorial in Github

The viGEN pipeline is easy to implement because our pipeline incorporates existing best practices and tools available. For

[4]BroadInstitute *Picard*. Available online at: http://broadinstitute.github.io/picard

[5]Ensembl *GFF/GTF File Format - Definition and Supported Options 2016*, (Accessed July, 2016). Available online at: http://useast.ensembl.org/info/website/upload/gff.html

Module 3, we developed our own algorithm for viral-gene quantification. The major motivation for this paper was to build on existing viral detection tools, and to build a quantification tool in order to quantify, explore and analyse the genes detected in viruses. The source code for the in-house algorithm, along with a tutorial on how to execute the code on sample data has been made public at https://github.com/ICBI/viGEN/.

Since access to TCGA raw data is controlled access, we could not use this dataset to create a publicly available tutorial. So we used a publicly available RNA-seq dataset to demonstrate our pipeline with an end-to-end workflow. We chose one sample (SRR1946637) from publicly available HBV liver cancer RNA-seq dataset from NCBI SRA (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA279878). This dataset is also available through EBI SRA (http://www.ebi.ac.uk/ena/data/view/SRR1946637). The dataset consisted of 50 HBV Liver cancer patients, and 5 adjacent normal liver tissues. We downloaded the raw reads for one sample, and applied our viGEN pipeline to it and were able to successfully detect HBV transcripts in this sample. A step-by-step workflow that includes – description of tools, code, intermediate and final analysis results are provided in Github: https://github.com/ICBI/viGEN/. This tutorial has also been provided as **Additional File 1**.

Custom Reference Index

We were interested in exploring all viruses existing in humans. So we first obtained reference genomes of all known and sequenced human viruses obtained from NCBI[2] (745 viruses) and merged them into one file (referred to as the "viral reference file") in fasta file format (Wikipedia, 2004). This file has been shared in our Github page.

Case Studies

Cervical Cancer Dataset

Cervical cancer is caused by the Human Papilloma Virus (HPV). This dataset consisted of 304 cervical cancer patients in the TCGA data collection. These samples were primary tumors from either Cervical Squamous Cell Carcinoma or Endocervical Adenocarcinoma where RNA-seq data was available.

We applied our viGEN pipeline on these samples using the Seven Bridges platform (https://cgc.sbgenomics.com). Among the 304 cervical cancer patients, 22 patients had virus detection confirmed by PCR or other lab methods and made available through the clinical data. So we used this information from the 22 patients to estimate the sensitivity and specificity of our viGEN pipeline.

Liver Cancer Dataset

This dataset consisted of 50 liver cancer patients in the TCGA data collection. 25 of these patients were afflicted with Hepatitis B virus (labeled "HepB"), while the rest of the 25 patients had a co-infection of both Hepatitis B and C viruses (labeled "HepB+C"). Information about viral presence was obtained from "Viral Hepatitis Serology" attribute from the clinical information.

We first applied the viGEN pipeline on the 50 samples, using the Globus Genomics platform (Bhuvaneshwar et al., 2015). Once the viral genomes were detected, we then chose only the high abundance viral species for the gene quantification step and viral variant detection steps (Module 3 and 4 respectively).

We then performed a focused analysis on this dataset. We used the viral-gene expression read counts, to examine the differences between "Dead" and "Alive" samples. The Dead/Alive status of the samples was obtained from the clinical data and refers to patients in the cohort that died or not from cancer. We performed this analysis on the 25 patients in the HepB only group to prevent any confounding with the HepB+HepC group. Out of 25 HepB patients, 16 were alive (baseline group), and 9 dead (comparison group) as per the clinical data. The analysis was performed using a Bioconductor software package called EdgeR (Robinson et al., 2010) in the R programming language (http://www.R-project.org).

Cox proportional hazards (Cox PH) regression model (Cox and Oakes, 2000) was then applied to look at the association of viral-gene expression data with overall survival. Thie Cox model was applied on all 50 samples in the cohort (i.e., 25 Hep B and 25 HepB+HepC) samples to maximize power.

We also compared the dead and alive samples at the viral RNA variant level in the HepB group using a tool called PLINK to see if it can add valuable information to the tumor landscape in humans.

RESULTS

Detection of HPV in Cervical Cancer Patients

We used our viGEN pipeline to detect viruses in the RNA of human cervical tissue and obtained viral copy number for each species. We used a threshold copy number of 10 as a "positive" viral detection for both HPV-16, HPV-18 and HPV-26 viruses. Based on this criterion, HPV-16 was detected in 53% of the samples, HPV-18 in 13% of the samples and HPV-26 in 0.3 % of the samples (**Figure 2**). The threshold copy number limit that defines a "positive" detection is one of the parameters of the software which could be set by the user depending on the specifics of the experiment.

We obtained the clinical data for this TCGA cervical cancer cohort from the cBio portal (Cerami et al., 2012). Among the 304 patients, 22 patients had virus detection confirmed by PCR

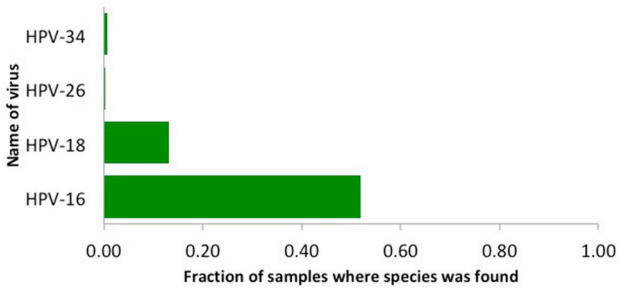


FIGURE 2 | The HPV viruses detected in cervical cancer patients using the viGEN pipeline.

or other lab methods and made available through the clinical data. Out of the 22 patients, 12 patients had the HPV-16 virus, 4 patients had HPV-18, and the rest had other HPV viruses. So we used this information from the clinical data to estimate the sensitivity and specificity of our viGEN pipeline. We got a sensitivity of 83% and specificity of 60% for HPV-16 detection (**Table 2A**); and a sensitivity of 75% and specificity of 94% for HPV-18 detection (**Table 2B**).

Table 2A | Estimation of sensitivity and specificity for HPV-16 detection in TCGA cervical cancer samples using the viGEN pipeline.

		PCR or other lab techniques for screening viruses (obtained from clinical data)		
		Positive	**Negative**	**Total**
RNA-seq from tumor tissue (output of algorithm)	Positive	10	4	14
	Negative	2	6	8
	Total	12	10	22

Table 2B | Estimation of sensitivity and specificity for HPV-18 detection in TCGA cervical cancer samples using the viGEN pipeline.

		PCR or other lab techniques for screening viruses (obtained from clinical data)		
		Positive	**Negative**	**Total**
RNA-seq from tumor tissue (output of algorithm)	Positive	3	1	4
	Negative	1	17	18
		4	18	22

Table 3 | Virus species detected in at-least 5 samples.

Name of Virus
Hepatitis C virus
Tick-borne encephalitis virus
Hepatitis C virus genotype 2
Cutthroat trout virus
Human endogenous retrovirus K113
Lunk virus NKS-1
Hepatitis C virus genotype 6
Hepatitis B virus

Additional Analysis in Liver Cancer Patients

Detection of Hepatitis B Virus at the Genome Level

We applied our viGEN pipeline (modules 1 and 2) on the RNAseq data from the TCGA liver cancer tumors, and obtained genome-level read counts for each viral species. We used a threshold copy number of 10 to define a positive detection of the Hepatitis B virus.

Once the viral genomes were detected, we short-listed the high abundance viral species for the viral-gene quantification step and viral-variant detection steps (Module 3 and 4 respectively). High abundance was defined as those virus species that were detected in at-least 5 samples. In addition to Hepatitis B and C viruses, several other viruses came up in this short list including Human endogenous retrovirus K113 (HERV K113) and others. A complete list is provided in **Table 3**.

Comparing Dead and Alive Samples in the Using Viral Gene Expression Data

To get a more detailed overview of the viral landscape, we applied Module 3 of the viGEN pipeline to the liver cancer dataset. This allowed us to quantify viral-gene expression regions in the RNA of liver tumor tissues. We then used those results to examine the differences between dead and alive samples.

It is known that these patients were afflicted with the Hepatitis B virus and hence many of the differentially expressed regions were from this viral genome. But as we know, other viruses also coexist in humans. This was confirmed by the presence of differentially expressed viral-regions from other viruses.

The differentially expressed regions that were significant among the results are shown in **Tables 4A,B**. **Table 4A** lists only the differentially expressed regions from Hepatitis B virus and **Table 4B** shows the differentially expressed regions from other viruses.

From the differential expression analyses, the two most informative results were (1) a region of the Hepatitis B genome that produced the HBeAg and HBcAg proteins were overexpressed in the dead patients and (2) another region of the Hepatitis B genome that produced HBsAg protein was overexpressed in the alive patients.

In detail, we saw several important findings as described below:

(a) Region NC_003977.1_CDS_1814_2452 of the Hepatitis B genome was 2.18 times overexpressed (log fold change = +1.128) in dead patients. This region contains Gene C that produces pre-code protein external core antigen; HBeAg. HBeAg is produced by proteolytic processing of the pre-core protein

(b) Region NC_003977.1_CDS_1901_2452 which was 1.74 times overexpressed (log fold change = +0.8, FDR = 0.053) in dead patients contains Gene C as above, but encodes a different core antigen HBcAg

(c) Region NC_003977.1_CDS_2848_4050 of the Hepatitis B genome was 6.73 times over expressed (log fold change = −2.7) in the alive patients of compared to the dead'patients. This region encodes Gene S that produces a large surface protein/L glycoprotein/L-HBsAG

(d) We also found several regions of the Human endogenous retrovirus K113 (HERV K113) viral genome (NC_022518.1_region_1112_6746, NC_022518.1_STS_5100_5381 and NC_022518.1_STS_7174_7323) to be about 2 times

Table 4 | Differential expression analysis of transcript level read counts Liver cancer dataset comparing Dead and Alive samples.

Name of region (Name of virus_region_start position of region_end position of region)	Log Fold change (logFC)	Log counts per million (logCPM)	*P*-Value	*Q*-Value (FDR)	Name of virus	Region annotation
A						
NC_003977.1_gene_1814_2452	1.128	13.449	1.71*E* − 06	0.000243	Hepatitis B virus	Contains Gene C that produces pre-code protein external core antigen; HBeAg. HBeAg is produced by proteolytic processing of the pre-core protein
NC_003977.1_CDS_1814_2452	1.128	13.449	1.71*E* − 06	0.000243	Hepatitis B virus	Contains Gene C that produces pre-code protein external core antigen; HBeAg.
NC_003977.1_CDS_1901_2452	0.828	12.42	0.000507	0.053928	Hepatitis B virus	Contains Gene C, encodes core antigen HBcAg
NC_003977.1_CDS_155_835	−2.121	16.335	1.67*E* − 15	1.11*E* − 12	Hepatitis B virus	Encodes Gene S that produces small envelope protein, S protein; S glycoprotein; S-HBsAg,
NC_003977.1_gene_2307_4838	−2.133	12.655	2.61*E* − 15	1.11*E* − 12	Hepatitis B virus	Gene P, encodes protein P
NC_003977.1_CDS_2307_4838	−2.133	12.655	2.61*E* − 15	1.11*E* − 12	Hepatitis B virus	Gene P, encodes protein P
NC_003977.1_CDS_3205_4050	−2.352	8.67	1.93*E* − 12	6.84*E* − 10	Hepatitis B virus	Gene S, encodes middle envelope protein pre-S2/S
NC_003977.1_gene_2848_4050	−2.75	11.741	5.84*E* − 22	6.20*E* − 19	Hepatitis B virus	Encodes Gene S that produces a large surface protein/L glycoprotein/L-HBsAG
NC_003977.1_CDS_2848_4050	−2.75	11.741	5.84*E* − 22	6.20*E* − 19	Hepatitis B virus	Encodes Gene S that produces a large surface protein/L glycoprotein/L-HBsAG
NC_003977.1_CDS_155_835	−2.121	16.335	1.67*E* − 15	1.11*E* − 12	Hepatitis B virus	Encodes Gene S that produces small envelope protein, S protein; S glycoprotein; S-HBsAg,
B						
NC_001405.1_intron_9724_12307	2.527	6.463	4.02*E* − 08	1.22*E* − 05	Human mastadenovirus C	Gene=L1, locus_tag=HAdVC_gp10, note=precedes capsid protein precursor pIIIa CDS
NC_001405.1_intron_9724_11039	2.5	6.451	6.65*E* − 08	1.28*E* − 05	Human mastadenovirus C	Gene=L1, locus_tag=HAdVC_gp10, note=precedes encapsidation protein 52K CDS
NC_001405.1_gene_10866_11023	2.5	6.451	6.65*E* − 08	1.28*E* − 05	Human mastadenovirus C	Gene=VAII, locus_tag=HAdVC_gs02, GeneID:2653002
NC_001405.1_transcript_10866_11023	2.5	6.451	6.65*E* − 08	1.28*E* − 05	Human mastadenovirus C	Gene=VAII, locus_tag=HAdVC_gs02, GeneID:2653002
NC_001405.1_exon_10866_11023	2.5	6.451	6.65*E* − 08	1.28*E* − 05	Human mastadenovirus C	Gene=VAII, locus_tag=HAdVC_gs02, GeneID:2653002
NC_001405.1_intron_10580_14015	2.428	6.475	9.24*E* − 08	1.64*E* − 05	Human mastadenovirus C	Gene=E2B, locus_tag=HAdVC_gp04
NC_022518.1_STS_7174_7323	−0.992	9.122	0.000309	0.034527	Human endogenous retrovirus K113	Sequence-tagged site (STS), locus_tag = Q779_gp1, standard_name=D6S2277, UniSTS:59918
NC_022518.1_STS_5100_5381	−1.051	9.532	0.000118	0.0139	Human endogenous retrovirus K113	Sequence-tagged site (STS), standard_name= D22S1651, UniSTS: 474031
NC_022518.1_region_1112_6746	−1.186	13.022	3.49*E* − 06	0.000463	Human endogenous retrovirus K113	gag-pro-pol; two−1 frameshifts predicted to occur to produce a fusion protein; the location of frameshifts has not been determined
NC_018464.1_region_1_927	−1.288	12.784	5.78*E* − 07	9.45*E* − 05	Shamonda virus	mol_type=genomic RNA, isolate=Ib An 5550, taxon:159150, segment=S
NC_002645.1_gene_293_20568	−2.598	6.126	3.93*E* − 05	0.004911	Human coronavirus 229E	locus_tag=HCoV229Egp1, GeneID: 918764, replicase polyprotein 1ab

These results shown used the viral-gene data obtained from Module 1 (using alignment tool Bowtie2) + Module 3. The table shows results with q value < 0.06 and sorted based on LogFC in the descending order. Table 4A shows transcript level read counts in the Hepatitis B virus while Table 4B shows transcript level read counts in other species.

overexpressed on average in alive patients (log fold change = −1.186, −1.051, −0.992).

Survival Analysis (Cox Regression) Using Viral Gene Expression Data

Based on the results from previous section, we selected two most informative regions from the Hepatitis B genome (log counts per million from NC_003977.1_CDS_2848_4050, NC_003977.1_CDS_1814_2452) for a Cox Proportional Hazard (Cox PH) model to look at association with overall survival event and time. The result from this model (**Table 5**), are consistent with the results from differential expression analysis:

(a) The Cox PH model shows that assuming other covariant to be constant, unit increase in expression of this region NC_003977.1_CDS_1814_2452, increases the hazard of event (death) by 70%.
(b) On the other hand, that assuming other covariant to be constant, unit increase in expression of this region NC_003977.1_CDS_2848_4050, decreases the hazard of event (death) by 43%.
(c) The overall model is significant with p-value < 0.05 from the Log rank test (also called Score test).

Comparing Dead and Alive Samples Using Viral-Variant Data

We performed variant calling (Module 4) on the liver cancer patients to see if it can add valuable information to the tumor landscape in humans. We then compared the dead and alive samples at the viral-variant level on the 25 HepB patients. For this analysis, the outputs from both Module 1 and 2 were fed into Module 4.

Most of the top variants from filtered human sample (Module 1 + Module 4) and unfiltered human sample (Module 2 + Module 4), were the same. We collated the significant common results (p-value ≤ 0.05) in **Tables 6A,B**. Among these results, we saw several missense and frameshift variants in Gene X of the Hepatitis B genome (nucleotide 1479), Gene P (2573, 2651, 2813), and a region that overlaps Gene P and PreS1 (nucleotides 2990, 2997, 3105, 3156). All these variants were found mutated more in the cases than controls. Other significant common results included variants in Gene C (nucleotide 1979, 2396) and variants in PreS2 region (nucleotide positions 115, 126 and 148) (**Table 6A**).

In addition, there were two missense variants that were common among the top results, but not significant (p-value = 0.06). They were variants in the X gene of the Hepatitis B genome (nucleotides 1762 and 1764) (**Table 6A**).

Among the significant common results to both, were a few variants of the Human endogenous retrovirus K113 complete genome (HERV K113). These include nucleotide positions 7476, 7426, and 8086. These map to frameshift and missense mutations in the putative envelope protein of this virus (Q779_gp1, also called "env") (**Table 6B**).

Table 5 | Cox proportional hazard survival analysis (across 25 HepB samples and 25 HepB + HepC Samples).

Formula:

coxph(formula = survObject ~ NC_003977.1_CDS_2848_4050 + NC_003977.1_CDS_1814_2452)

Results from the model:

$n = 37$, number of events = 5

(13 observations deleted due to missingness)

Covariate	coef	exp(coef)	se(coef)	Z	Pr(>|z|)
NC_003977.1_CDS_2848_4050	−0.5548	0.5742	0.7434	−0.746	0.456
NC_003977.1_CDS_1814_2452	0.5302	1.6993	0.6145	0.863	0.388

Covariate	exp(coef)	exp(coef)	Lower 0.95	Upper 0.95
NC_003977.1_CDS_2848_4050	0.5742	1.7415	0.1337	2.465
NC_003977.1_CDS_1814_2452	1.6993	0.5885	0.5096	5.667

Concordance = 0.654 (se = 0.188)

Rsquare = 0.12 (max possible = 0.329)

Likelihood ratio test = 4.74 on 2 df, $p = 0.09343$

Wald test = 0.75 on 2 df, $p = 0.6856$

Score (logrank) test = 10.58 on 2 df, $p = 0.00503$

These results shown used the viral-gene/CDS data obtained from Module 1 (using alignment tool Bowtie2) + Module 3. Coef: coefficient (Beta) of the model; exp(coef): Hazard Ratio; se(coef): Standard Error; Pr(>|z|): P-value

(a) The Cox PH model shows that assuming other covariant to be constant, unit increase in expression of this region NC_003977.1_CDS_1814_2452, increases the hazard of event (death) by 70%.

(b) On the other hand, that assuming other covariant to be constant, unit increase in expression of this region NC_003977.1_CDS_2848_4050, decreases the hazard of event (death) by 43%.

(c) The overall model is significant with p-value < 0.05 from the Log rank test (also called Score test).

Table 6 | Results of case-control association test applied on the results from viral variant calling.

CHR (Chromosome)	Species (Name of Virus)	BP (Base pair)	A1 (minor allele)	C_A (Number of cases with A1)	C_U (number of controls with A1)	A2 (major allele)	P (*P* value)	Annotation from GFF file
A								
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	1479	C	4	0	A	0.02857	Gene=X, product=X protein, protein_id=NP_647606.1
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	2573	C	0	6	T	0.03571	Gene=P,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	2651	T	4	0	C	0.00476	product=polymerase,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	2813	C	2	0	T	0.03571	protein_id=NP_647604.2
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	2990	T	2	0	A	0.02222	Gene=P, product=polymerase,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	2997	C	2	0	T	0.03571	protein_id=NP_647604.2;
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	3105	C	2	0	A	0.02222	Gene=S, product=large envelope
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	3156	G	4	0	A	0.00476	protein, protein_id=YP_355333.1
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	1979	G	2	0	A	0.03571	Gene=C, product=pre-capsid protein,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	2396	0	4	0	CG	0.01499	protein_id=YP_355335.1, NP_647607.1
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	115	C	2	0	A	0.02222	Pre S2 region, ID=id0, Dbxref=taxon: 10407,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	126	C	2	0	T	0.02222	Is_circular=true, gbkey=Src, genome=genomic,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	148	G	2	0	A	0.02222	mol_type=genomic DNA, strain=ayr
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	1762	T	0	4	A	0.06061	Gene=X, Name=NP_647606.1, product=X protein,
gi|21326584|ref|NC_003977.1|	Hepatitis B virus	1764	A	0	4	G	0.06061	protein_id=NP_647606.1
B								
gi|548558394|ref|NC_022518.1|	Human endogenous retrovirus K113	7476	0	10	14	TACTG	0.00600	ID=gene0, Name=Q779_gp1; ID=cds0, Name=YP_008603282.1,
gi|548558394|ref|NC_022518.1|	Human endogenous retrovirus K113	7426	G	3	0	A	0.00714	product=putative env,
gi|548558394|ref|NC_022518.1|	Human endogenous retrovirus K113	8086	T	3	0	C	0.00714	protein_id=YP_008603282.1

The table is sorted based on Annotation. Annotation includes gene name, protein name, etc., separated by commas, multiple annotations separated by semi-colon.
Table 6A shows variants in the Hepatitis B virus only while Table 6B shows variants in other species. (Shows only common results between two possible analysis steps).

DISCUSSION

Detection of HPV in Cervical Cancer Patients

The Seven Bridges team used two metagenomic tools,Centrifuge (Kim et al., 2016) and Kraken (Wood and Salzberg, 2014), to detect HPV viruses on the same cohort of TCGA patients (Bridges, 2017; Malhotra et al., 2017), and shared the results with us. They used an abundance of 0.02 as a positive viral detection (Bridges, 2017; Malhotra et al., 2017). We compared viGEN with Kraken and Centrifuge in terms of the percentage of samples where the species was detected (**Table 7**). We can see that the results are in the same range for all three tools.

We also estimated the sensitivity and specificity of these tools using the same 22 patients and compared with that of the viGEN pipeline. The Centrifuge tool had a sensitivity of 83% and specificity of 60% for HPV-16 detection; and a sensitivity of 75% and specificity of 94% for HPV-18 detection. The Kraken tool had a sensitivity of 83% and specificity of 20% for HPV-16 detection; and a sensitivity of 75% and specificity of 17% for HPV-18 detection (detailed in **Additional File 2**). It shows that our viGEN pipeline was able to match the sensitivity and specificity of Centrifuge tool and surpassed that of Kraken (detailed in **Additional Files 2**, **3**).

Table 7 | Comparing the viral detection ability of viGEN with other tools.

Virus Name	Tool Name	% of samples where the species was found
HPV 16 (Alphapapillomavirus 9)	Kraken	58
HPV 16	Centrifuge	55
HPV 16	viGEN	53
HPV 18	Kraken	2
HPV 18 (Alphapapillomavirus 7)	Centrifuge	15
HPV 18	viGEN	13
HPV 26 (Alphapapillomavirus 5)	Kraken	1
HPV 26	Centrifuge	0.3
HPV 26	viGEN	0.3

Additional Analysis on Liver Cancer Patients

We used our viGEN pipeline to get genome-level read counts obtained from viruses detected in the RNA of human liver tissue. In our results, HBV was detected in 20% of the samples. This is similar to earlier analyses of TCGA liver cancer cohort study (Khoury et al., 2013; Tang et al., 2013; The Cancer Genome Atlas Research Network, 2017), which detected the HBV virus in 23 and 32% (with typically low counts range) of cases respectively.

It has also been reported that the viral gene X (HBx) was the most predominately expressed viral gene in liver cancer samples (Tang et al., 2013) which is in concordance with our findings where the peak number of reads were observed for gene X region of the HBV genome.

Comparing Dead and Alive Samples in the Liver Cancer Cohort Using Viral Gene Expression Data

To get a more detailed overview of the viral landscape, we examined the human RNA-seq data to detect and quantify viral gene expression regions. We then examined the differences between dead and alive samples at the viral-transcript level on the Hepatitis B sub-group (**Tables 4A,B**).

From the differential expression analyses, the two most informative results were (1) a region of the Hepatitis B genome that produced the HBeAg protein was overexpressed in the dead patients and (2) another region of the Hepatitis B genome that produced HBsAg protein was overexpressed in the alive patients.

Presence of HBeAg or HBcAg is an indicator of active viral replication; this means the person infected with Hepatitis B can likely transmit the virus on to another person. Typically, loss of HBeAg is an indicator of recovery from acute Hepatitis B infection. Active viral replication could allow the virus to persist in infected cells, and increase the risk of disease (Tage-Jensen et al., 1985; Liang, 2009). So our results, showing that antigens HBeAg and HBcAg were overexpressed in dead patients compared to alive patients makes sense, indicating that these patients never recovered from acute infection.

The results also indicate a higher level of HBsAg in the alive patients compared to the dead patients. The highest levels of HBsAg in the virus are known to occur in the "immunotolerant phase." This pattern is seen in patients who are inactive carriers of the virus i.e., they have the wild type DNA, and the virus has been in the host for so long, that the host does not see the virus as a foreign protein in the body, and hence there's no immune reaction against the virus. In this phase, there is known to be minimal liver inflammation and low risk of disease progression (Park, 2004; Tran, 2011; Locarnini and Bowden, 2012). This could explain why we saw higher level of HBsAg in the alive patients compared to the dead patients.

Also among the significant results were three regions from the Human endogenous retrovirus K113 (HERV K113) genome (with negative log fold change) that were overexpressed in the alive patients. Two of these regions were Sequence-tagged sites (STS) and the third region was in the gag-pro-pol region that has frameshifts. HERV could protect the host from invasion from related viral agents through either retroviral receptor blockade or immune response to the undesirable agent (Nelson et al., 2003).

Overall, we found that our results from viral-gene expression level make biological sense, with much of the results validated through published literature.

Comparing Dead and Alive Samples in the Liver Cancer Cohort Using Viral-Variant Data

We performed variant calling on the viral data to see if it can add valuable information to the tumor landscape in humans. We then compared the dead and alive samples at the viral-variant level on the 25 patients in the Hepatitis B sub-group.

Among the significant results (**Tables 6A,B**) included variants in Gene C (nucleotide 1979, 2396) and variants in PreS2 region (nucleotide positions 115, 126 and 148). The Gene C region creates the pre-capsid protein, which plays a role in regulating genome replication (Tan et al., 2015). The mutation in the 2396 position lies in a known CpG island (ranging from 2215 to 2490), whose methylation level is significantly correlated with hepatocarcinogenesis (Jain et al., 2015). Mutations in PreS2 are associated with persistent HBV infection, and emerge in chronic infections. The PreS1 and PreS2 regions are known to play an essential role in the interaction with immune responses because they contain several epitopes for T or B cells (Cao, 2009).

Mutations in the 1762/1764 positions of the X gene are known to be associated with greater risk of HCC (Cao, 2009; Wang et al., 2014), and is independent of serum HBV DNA level (Wang et al., 2014). This mutation combination is also known to be associated with hepatitis B related acute-on-chronic liver failure (Xiao et al., 2011). It is predicted that mutations associated with HCC variants are likely generated during HBV-induced pathogenesis. The A1762T/G1764A combined mutations was shown to be a valuable biomarker in the predicting the risk of HCC (Cao, 2009; Wang et al., 2014); and are often detected about 10 years before the diagnosis of HCC (Cao, 2009).

Among the significant common results to both, were a few variants of the Human endogenous retrovirus K113 complete genome (HERV K113). These variants map to frameshift and missense mutations in the putative envelope protein of this virus (Q779_gp1, also called "env"). Studies have shown that this envelope protein mediates infections of cells (Robinson and Whelan, 2016). HERV K113 is a provirus and is capable of producing intact viral particles (Boller et al., 2008). Studies have shown a strong association between HERV-K antibodies and clinical manifestation of disease and therapeutic response (Moyes et al., 2005; Downey et al., 2015). It is hypothesized that retroviral gene products can be "reawakened" when genetic damage occurs through mutations, frameshifts and chromosome breaks. Even though the direct oncogenic effects of HERVs in cancer are yet to be completely understood, it has shown potential as diagnostic or prognostic biomarkers and for immunotherapeutic purposes including vaccines (Downey et al., 2015).

We compared various viral detection pipeline using the several criteria (**Table 1**). Our pipeline provides similar functionality as the tools listed in **Table 1** for the detection of viruses from human RNAseq data; but also has an advantage of enabling gene-level expression analysis and quantification, as well as variant analysis of viral genomes in a single open source publicly available package.

Limitations

One limitation of our viGEN pipeline is that it is dependent on sequence information from reference genome. This makes it challenging to detect viral strains where reference sequence information is not known. In the future, we plan to explore de novo assembly incorporating more sophisticated methods like Hidden Markov Models (HMM) (Alves et al., 2016). This would enable us to provide in-depth analysis of strain pathogenicity in the context of clinical outcome.

Biological Significance

In recent years, US regulators approved a viral based cancer therapy (Ledford, 2015), proving that the study of viruses in the human transcriptome has biomedical interest, and is paving the way for promising research and new opportunities.

We show that our viGEN pipeline can thus be used on cancer and non-cancer human NGS data to provide additional insights into the biological significance of viral and other types of infection in complex diseases, and tumorigeneses. Our viral pipeline could be used in conjunction with additional type of immuno-oncology analysis based on RNA-seq data of host RNA for cancer immunology applications. Detection and characterization of these infectious agents in tumor samples can give us better insights into disease mechanisms and their treatment (Hausen, 2000).

CONCLUSION

With the decreasing costs of NGS analysis, our results show that it is possible to detect viral sequences from whole-transcriptome (RNA-seq) data in humans. Our analysis shows that it is not easy to detect DNA and RNA viruses from tumor tissue, but certainly possible. We were able to not only quantify them at a viral-gene expression level, but also extract variants. Our goal is to facilitate better understanding and gain new insights in the biology of viral presence/infection in actual tumor samples. The results presented in this paper on two case studies are in correspondence with published literature and are a proof of concept of our pipeline.

This pipeline is generalizable, and can be used to examine viruses present in genomic data from other next generation sequencing (NGS) technologies. It can also be used to detect and explore other types of microbes in humans, as long as the sequence information is available from the National Center for Biotechnology Information (NCBI) resources.

This pipeline can thus be used on cancer and non-cancer human NGS data to provide additional insights into the biological significance of viral and other types of infection in complex diseases, and tumorigeneses. We are planning to package this pipeline and make it open source to the bioinformatics community through Bioconductor.

AVAILABILITY OF DATA AND MATERIAL

The TCGA liver cancer dataset was used in the analysis and writing of this manuscript. The data can be obtained from https://cancergenome.nih.gov/.

Since access to TCGA raw data is controlled access, we could not use this dataset to create a publicly available tutorial. So we looked for publicly available RNA-seq dataset to demonstrate our pipeline with an end-to-end workflow.

We chose one sample (SRR1946637) from publicly available liver cancer RNA-seq dataset from NCBI SRA (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA279878). This dataset is also available through EBI SRA (http://www.ebi.ac.uk/ena/data/view/SRR1946637). The dataset consists of 50 Liver cancer patients, and 5 adjacent normal liver tissues. We downloaded the raw reads for one sample, and applied our viGEN pipeline to it. A step-by-step workflow that includes – description of tools, code, intermediate and final analysis results are provided in Github: https://github.com/ICBI/viGEN/.

Project details:

Project name: viGEN
Project home page: https://github.com/ICBI/viGEN/
Operating system(s): The R code is platform independent. The shell scripts can run on Unix, Linux, or iOS environment
Programming language: R, bash/shell
Other requirements: N/A
License: N/A
Any restrictions to use by non-academics: N/A

AUTHOR CONTRIBUTIONS

KB and YG designed the pipeline. KB and LS implemented the pipeline. KB and YG wrote the manuscript with editorial comments from SM.

FUNDING

This work was funded by the Lombardi cancer center support grant (P30 CA51008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01172/full#supplementary-material

Additional File 1 | viGEN Github tutorial.

Additional File 2 | Detailed results from analysis of TCGA cervical cancer patients.

Additional File 3 | Output from Kraken and Centrifuge shared by the Seven Bridges Team.

REFERENCES

ACS (2007). *Infections That Can Lead to Cancer 2015*. Available online at: http://www.cancer.org/cancer/cancercauses/othercarcinogens/infectiousagents/infectiousagentsandcancer/infectious-agents-and-cancer-viruses

Alves, J. M., de Oliveira, A. L., Sandberg, T. O., Moreno-Gallego, J. L., de Toledo, M. A., de Moura, E. M., et al. (2016). GenSeed-HMM: a tool for progressive assembly using profile HMMs as seeds and its application in alpavirinae viral discovery from metagenomic data. *Front. Microbiol.* 7:269. doi: 10.3389/fmicb.2016.00269

Barzon, L., Lavezzo, E., Militello, V., Toppo, S., and Palù, G. (2011). Applications of next-generation sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* 12, 7861–7884. doi: 10.3390/ijms12117861

Bhaduri, A., Qu, K., Lee, C. S., Ungewickell, A., and Khavari, P. A. (2012). Rapid identification of non-human sequences in high-throughput sequencing datasets. *Bioinformatics* 28, 1174–1175. doi: 10.1093/bioinformatics/bts100

Bhuvaneshwar, K., Sulakhe, D., Gauba, R., Rodriguez, A., Madduri, R., Dave, U., et al. (2015). A case study for cloud based high throughput analysis of NGS data using the globus genomics system. *Comput. Struct. Biotechnol. J.* 13, 64–74. doi: 10.1016/j.csbj.2014.11.001

Boller, K., Schönfeld, K., Lischer, S., Fischer, N., Hoffmann, A., Kurth, R., et al. (2008). Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *J. Gen. Virol.* 89(Pt 2), 567–572. doi: 10.1099/vir.0.83534-0

Bridges, S. (2017). *Identifying Viral Sequences In TCGA Data Using Kraken And Centrifuge*. (Accessed May 23, 2017). Available online at: https://www.sevenbridges.com/centrifuge/

Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., and Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* 17, 257–271. doi: 10.1038/nrg.2016.10

Cao, G. W. (2009). Clinical relevance and public health significance of hepatitis B virus genomic variations. *World J. Gastroenterol.* 15, 5761–5769. doi: 10.3748/wjg.15.5761

Center-for-Statistical-Genetics (2013). *BAM 2013* (Accessed February 26, 2013). Available online at: http://genome.sph.umich.edu/wiki/BAM

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404. doi: 10.1158/2159-8290.CD-12-0095

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 4:7. doi: 10.1186/s13742-015-0047-8

Chen, Y., Yao, H., Thompson, E. J., Tannir, N. M., Weinstein, J. N., and Su, X. (2013). VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics.* 29, 266–267. doi: 10.1093/bioinformatics/bts665

Cox, D. R., and Oakes, D. (2000). *Analysis of Survival Data.* Boca Raton, FL: Chapman and Hall.

Datta, S., Budhauliya, R., Das, B., Chatterjee, S., Vanlalhmuaka, and Veer, V. (2015). Next-generation sequencing in clinical virology: discovery of new viruses. *World J. Virol.* 4, 265–276. doi: 10.5501/wjv.v4.i3.265

Downey, R. F., Sullivan, F. J., Wang-Johanning, F., Ambs, S., Giles, F. J., and Glynn, S. A. (2015). Human endogenous retrovirus K and cancer: innocent bystander or tumorigenic accomplice? *Int J. Cancer.* 137, 1249–1257. doi: 10.1002/ijc.29003

Hausen, H. Z. (2000). *Infections Causing Human Cancer.* Weinheim: Wiley.

Jain, S., Chang, T. T., Chen, S., Boldbaatar, B., Clemens, A., Lin, S. Y., et al. (2015). Comprehensive DNA methylation analysis of hepatitis B virus genome in infected liver tissues. *Sci. Rep.* 5:10478. doi: 10.1038/srep10478

Khoury, J. D., Tannir, N. M., Williams, M. D., Chen, Y., Yao, H., Zhang, J., et al. (2013). Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *J. Virol.* 87, 8916–8926. doi: 10.1128/JVI.00340-13

Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi: 10.1101/gr.210641.116

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi: 10.1101/gr.129684.111

Kostic, A. D., Ojesina, A. I., Pedamallu, C. S., Jung, J., Verhaak, R. G., Getz, G., et al. (2011). PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat. Biotechnol.* 29, 393–396. doi: 10.1038/nbt.1868

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118

Ledford, H. (2015). Cancer-fighting viruses win approval. *Nature* 526, 622–623. doi: 10.1038/526622a

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformat*. 12:323. doi: 10.1186/1471-2105-12-323

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, J. W., Wan, R., Yu, C. S., Co, N. N., Wong, N., and Chan, T. F. (2013). ViralFusionSeq: accurately discover viral integration events and reconstruct fusion transcripts at single-base resolution. *Bioinformatics* 29, 649–651. doi: 10.1093/bioinformatics/btt011

Liang, T. J. (2009). Hepatitis B: the virus and disease. *Hepatology* 49(5 Suppl), S13–S21. doi: 10.1002/hep.22881

Locarnini, S., and Bowden, S. (2012). Hepatitis B surface antigen quantification: not what it seems on the surface. *Hepatology* 56, 411–414. doi: 10.1002/hep.25732

Malhotra, R., Grady, P., Lehnert, E., Sethi, A., Prakasha, P., and Brandi, N., Davis-Dusenbery (eds.). (2017). "Enabling scalable and rapid metagenomic profiling of the transcriptome with the Seven Bridges Cancer Genomics Cloud," in *BioIT World 2017* (Boston, MA).

Moyes, D. L., Martin, A., Sawcer, S., Temperton, N., Worthington, J., Griffiths, D. J., et al. (2005). The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease. *Genomics* 86, 337–341. doi: 10.1016/j.ygeno.2005.06.004

NCI (2011). *The Cancer Genome Atlas.* Available online at: https://tcga-data.nci.nih.gov/

Nelson, P. N., Carnegie, P. R., Martin, J., Davari Ejtehadi, H., Hooley, P., Roden, D., et al. (2003). Demystified. Human endogenous retroviruses. *Mol. Pathol*. 56, 11–18. doi: 10.1136/mp.56.1.11

Park, J. H. (2004). [Hepatitis B virus surface antigen: a multifaceted protein]. *Korean J. Hepatol*. 10, 248–259.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/

Robinson, L. R., and Whelan, S. P. (2016). Infectious entry pathway mediated by the human endogenous retrovirus K Envelope Protein. *J. Virol*. 90, 3640–3649. doi: 10.1128/JVI.03136-15

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Salyakina, D., and Tsinoremas, N. F. (2013). Viral expression associated with gastrointestinal adenocarcinomas in TCGA high-throughput sequencing data. *Hum. Genomics*. 7:23. doi: 10.1186/1479-7364-7-23

Schelhorn, S. E., Fischer, M., Tolosi, L., Altmüller, J., Nürnberg, P., Pfister, H., et al. (2013). Sensitive detection of viral transcripts in human tumor transcriptomes. *PLoS Comput. Biol*. 9:e1003228. doi: 10.1371/journal.pcbi.1003228

Spencer, D. H., Tyagi, M., Vallania, F., Bredemeyer, A. J., Pfeifer, J. D., Mitra, R. D., et al. (2014). Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. *J. Mol. Diagn*. 16, 75–88. doi: 10.1016/j.jmoldx.2013.09.003

Tage-Jensen, U., Aldershvile, J., and Schlichting, P. (1985). Immunosuppressive treatment of HBsAg-positive chronic liver disease: significance of HBeAg. *Hepatology* 5, 47–49. doi: 10.1002/hep.1840050111

Tan, Z., Pionek, K., Unchwaniwala, N., Maguire, M. L., Loeb, D. D., and Zlotnick, A. (2015). The interface between hepatitis B virus capsid proteins affects self-assembly, pregenomic RNA packaging, and reverse transcription. *J. Virol*. 89, 3275–3284. doi: 10.1128/JVI.03545-14

Tang, K. W., Alaei-Mahabadi, B., Samuelsson, T., Lindh, M., and Larsson, E. (2013). The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat. Commun*. 4:2513. doi: 10.1038/ncomms3513

The Cancer Genome Atlas Research Network (2017). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169, 1327–41.e23. doi: 10.1016/j.cell.2017.05.046

Tran, T. T. (2011). Immune tolerant hepatitis B: a clinical dilemma. *Gastroenterol. Hepatol. (N Y)*. 7, 511–516.

Wang, F., Sun, Y., Ruan, J., Chen, R., Chen, X., Chen, C., et al. (2016). Using small RNA deep sequencing data to detect human viruses. *Biomed. Res. Int*. 2016:2596782. doi: 10.1155/2016/2596782

Wang, Q., Jia, P., and Zhao, Z. (2013). VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS ONE* 8:e64465. doi: 10.1371/journal.pone.0064465

Wang, Y. Z., Zhu, Z., Zhang, H. Y., Zhu, M. Z., Xu, X., Chen, C. H., et al. (2014). Detection of hepatitis B virus A1762T/G1764A mutant by amplification refractory mutation system. *Braz. J. Infect. Dis*. 18, 261–265. doi: 10.1016/j.bjid.2013.09.005

Wikipedia (2004). *FASTA Format*. Available online at: https://en.wikipedia.org/wiki/FASTA_format

Wikipedia (2009). *FASTQ Format*. Available online at: https://en.wikipedia.org/wiki/FASTQ_format

Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. 15:R46. doi: 10.1186/gb-2014-15-3-r46

Xiao, L., Zhou, B., Gao, H., Ma, S., Yang, G., Xu, M., et al. (2011). Hepatitis B virus genotype B with G1896A and A1762T/G1764A mutations is associated with hepatitis B related acute-on-chronic liver failure. *J. Med. Virol*. 83, 1544–1550. doi: 10.1002/jmv.22159

Yoshihara, N. (1995). [ELISA for diagnosis of infections by viruses]. *Nihon Rinsho*. 53, 2277–2282.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bhuvaneshwar, Song, Madhavan and Gusev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

REVIEW
published: 07 April 2017
doi: 10.3389/fmicb.2017.00559

Intriguing Interaction of Bacteriophage-Host Association: An Understanding in the Era of Omics

Krupa M. Parmar[1], Saurabh L. Gaikwad[1], Prashant K. Dhakephalkar[1], Ramesh Kothari[2] and Ravindra Pal Singh[2]*†*

[1] Bioenergy Group, Agharkar Research Institute, Pune, India, [2] Department of Biosciences, Saurashtra University, Rajkot, India

Innovations in next-generation sequencing technology have introduced new avenues in microbial studies through "omics" approaches. This technology has considerably augmented the knowledge of the microbial world without isolation prior to their identification. With an enormous volume of bacterial "omics" data, considerable attempts have been recently invested to improve an insight into virosphere. The interplay between bacteriophages and their host has created a significant influence on the biogeochemical cycles, microbial diversity, and bacterial population regulation. This review highlights various concepts such as genomics, transcriptomics, proteomics, and metabolomics to infer the phylogenetic affiliation and function of bacteriophages and their impact on diverse microbial communities. Omics technologies illuminate the role of bacteriophage in an environment, the influences of phage proteins on the bacterial host and provide information about the genes important for interaction with bacteria. These investigations will reveal some of bio-molecules and biomarkers of the novel phage which demand to be unveiled.

Keywords: bacteriophage, genomics, next-generation sequencing, transcriptomics

OPEN ACCESS

Edited by:
Florence Abram, NUI Galway, Ireland

Reviewed by:
Steven Ripp, University of Tennessee, Knoxville, USA
Sanna Sillankorva, University of Minho, Portugal

****Correspondence:***
Krupa M. Parmar
krupa_11091@yahoo.co.in
Ravindra Pal Singh
ravindrapal.1441@gmail.com

†Present Address:
Ravindra Pal Singh, Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, UK

Specialty section:
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology

Received: *03 January 2017*
Accepted: *16 March 2017*
Published: *07 April 2017*

Citation:
Parmar KM, Gaikwad SL, Dhakephalkar PK, Kothari R and Singh RP (2017) Intriguing Interaction of Bacteriophage-Host Association: An Understanding in the Era of Omics. Front. Microbiol. 8:559. doi: 10.3389/fmicb.2017.00559

INTRODUCTION

Innovations in next-generation sequencing (NGS) technology and the decline in the sequencing cost have triggered a revolution to gain an understanding into the diversity, structure, and function of complex microbial processes (Vlahou and Fountoulakis, 2005). NGS has enhanced our concept of various influences of microbes in maintaining equilibrium in the environment and accentuating the function of diverse hosts including humans (Li et al., 2008). NGS has expedited the interpretation of microbiome using techniques, such as metagenomics, metatranscriptomics, metabolomics, proteomics, and single cell genomics. Apart from sequencing, bioinformatics and statistical tools also represent a significant role in sequence assembly, alignment, binning, and annotation. Software of bioinformatic assists in decoding the identity, abundance profile, genetic composition and functional channels of an organism or for a microbial community (Meyer et al., 2008; Glass et al., 2010). Genomics deals with sequencing of the whole genome of a distinct organism whilst metagenomics study a pool of genomes of a community of different populations (Handelsman, 2004). According to the central dogma, the flow of genetic information in a cell is from DNA (genome) to RNA (transcriptome) and then it is translated into proteins (proteome) (Crick, 1970). Genomics elucidates presence of the gene in an organism, while transcriptomics provides the information about the genes that are actively expressed in an organism and proteomics

study the structure and function of every protein in an organism. A novel technique called single cell genomics takes in record information of the sequence from an individual cell which procures a better degree of accuracy in cellular differences and a finer understanding into the function of a particular cell in an ecosystem (Eberwine et al., 2014). However, metabolomics includes the analysis of metabolites of an organism and these results may vary from genomics and transcriptomics data as they are influenced by surrounding environments (Apel and Hirt, 2004). Apart from sequencing, bioinformatics and statistical tools assist in sequence assembly, alignment, binning and annotation.

Whilst, the information about bacteria present on earth is better understood, data regarding viruses particularly bacteriophages (henceforth called phages) is still in its infancy. Phages are the most abundant and diverse group of viruses found on Earth (Short and Suttle, 2005). Interactions between the bacterial host and phage have significanty played an role in biogeochemical cycles, regulation of the microbial community structures and governing the microbial populations (**Figure 1**). Recent surveys have documented the capacity of phages in maintaining the stability of microflora in the human gut (Minot et al., 2011) and regulation of pathogen and multidrug resistance (MDR) in the environment (Parmar et al., 2017). In bacteria, 16S rRNA genes and several house-keeping genes serve as a biomarker which facilitate their identification, whereas, there is an absence of biomarker gene among phages, which poses as a hindrance for identification of phages and hence the phage database is quite insufficient (Edwards and Rohwer, 2005). Addressing the challenge to design a biomarker for phages may uncover new avenues in better understanding the virosphere. Plaque assay being a culture-dependent technique isolates a specific phage against a bacterial host. Subsequently, their identification can be worked out using phenotypic characteristics and sequencing (Sanchez et al., 2015). In an approach to improve a comprehensive insight into uncultivable phage and their interaction with the bacterial host, this review summarizes different NGS techniques and the bioinformatics tools that are applied to evaluate such data.

BACTERIOPHAGES AND THEIR INTERACTION WITH BACTERIAL HOST

Phages are in 10:1 ratio with bacteria on Earth, though viral DNA corresponds to only 0.1% of total DNA among microbial communities (Qin et al., 2010). Phages are ubiquitous in the environment and are found abundantly where bacterial hosts thrive. Mostly phages flourish in oceans, soil, wastewater treatment plants, hot-water springs and animal gut (Wommack and Colwell, 2000; Prigent et al., 2005; Prestel et al., 2008; Srinivasiah et al., 2008). Phages are classified on the basis of their size, structural composition, genome organization and on the host it infects (Ackermann, 2009). Electron microscopy assists in studying phenotypic characteristics of phage on the basis of the size and shape of head, tail and tail fibers. The most abundant phages in the environment are dsDNA belonging to order *Caudovirales* (Weinbauer and Rassoulzadegan, 2004). *Caudovirales* are furthermore classified into *Podoviridae* having a short tail, *Siphoviridae* with long non-contractile tail and *Myoviridae* possessing a long contractile tail. Every phage is specific toward a particular bacterial host and may have a narrow or broad host range depending on its infection capability. Hosts provide the enzymatic machinery for the phages to replicate and multiply by infecting the most active (exponentially growing) bacteria as implied by "kill the winner" hypothesis (Rodrigue et al., 2009). Phages undergo two types of life cycles, (1) in the lytic lifecycle, phage injects own DNA into a host cell and multiply by manipulating host replicating machinery. After replication, phage cleaves the host bacterial cells releasing progeny virus particles into the environment. Whereas, in (2) lysogeny lifecycle, phage DNA integrates into the bacterial genome, replicates their DNA along the bacterial genome and transfer on to the progeny of host cells (Bertani, 1951).

The interplay between host and phage particle initiates as soon as the phage recognizes specific receptors on the bacterial cell wall. The tail proteins of phage particle recognize the receptor protein(s) of bacteria and inject own DNA into host cytoplasm choosing either lytic or lysogenic lifecycle. Once the phage DNA is inserted into the bacterial cell, the cell is termed as a "virocell" carrying virus auxiliary metabolic genes (vAMGs), which are believed to augment the metabolic potential of the host during infection process as shown in **Figure 1** (Rosenwasser et al., 2016). The phages acquire new genes into their genomes by interactions with the host genome in order to replicates in the host cells. The bacterial genes that attach near to the prophage attachment site, suggests the genes were acquired by inaccurate prophage excision. Some novel genes can similarly be transmitted into the interior part of the genome by some unexplained mechanism (Juhala et al., 2000). However, these genes may be autonomous transcripts or repressed prophages that provide benefit to hosts (Brüssow et al., 2004). Horizontal gene transfer (HGT) by these phage particles from one host to another host genomes, results in an increased microbial diversity (Dutta and Pan, 2002; Weinbauer and Rassoulzadegan, 2004). Thus, the interaction between phage and host chiefly emphasizes the structure of microbial communities (Rohwer and Thurber, 2009). Some genes derived by phage also aid in nutrient cycling and gear up the biogeochemical cycles on Earth. Furthermore, phages have a crucial aspect in host mortality, carbon cycling (Breitbart et al., 2004) and nutrient cycling (Suttle, 2007). Also, microbial lysis by phage infection has significance in bacterial population control and the debris of these dead microbes act as a food source in the food web of the environment (Sime-Ngando and Colombet, 2009), thus involved the cycling nutrients (**Figure 1**). Phages are thus accounted as an application to limit bacterial pathogens and multi-drug resistant organisms in the environment by the mechanism of specifically lysing the bacterial hosts (Parmar et al., 2017). Despite an immense abundance and diversity of phages and their reimbursement in the global webs, molecular knowledge of phage-host interactions is missing. In the era of NGS, employing genomics, single cell genomics, transcriptomics, proteomics, and metabolomics can be a smart attempt to understand the interaction among the phages and their bacterial hosts (**Figure 2**). A review of the literature has been

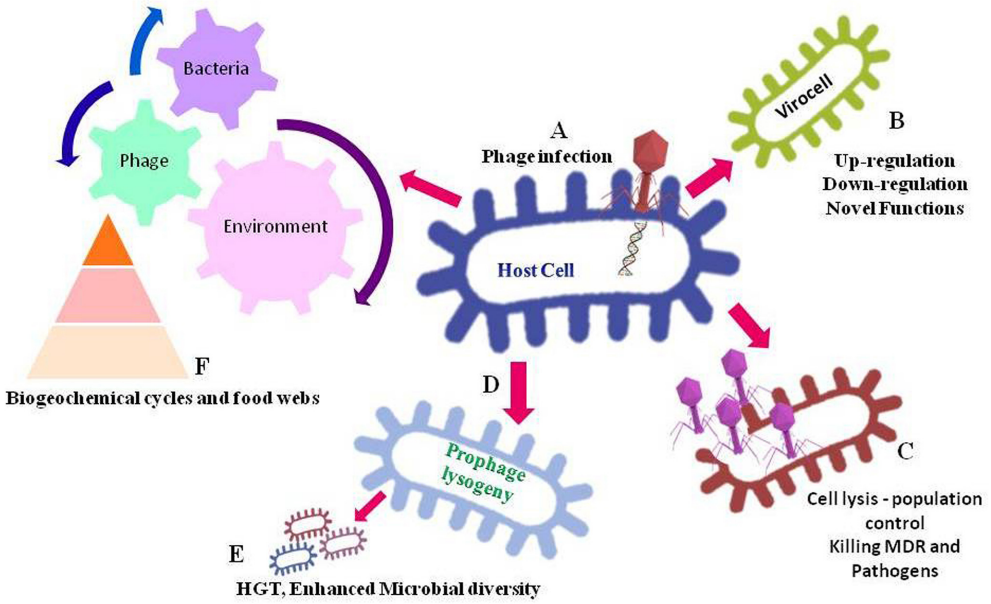


FIGURE 1 | Significance of bacteriophages in regulating environment. In this figure, **(A)** depicts infection and insertion phage DNA into the host cell. After the phage infection, the host cell may get converted into a virocell **(B)**, containing vAMGs that leading to an altered regulation or novel functions in bacterial host cell. The phage infection leading to lytic cycle **(C)** results into lysing host cell hence, controlling cell population. Infected cell leading to a lysogeny cycle **(D)** may contain phage genome into the bacterial genome, which can lead to an increased microbial diversity because of horizontal gene transfers- HGT **(E)**. Also, the dead debris of bacteria as a result of phage lysis enters the food-web and biogeochemical cycles **(F)**, as a result the nutrients get re-circulated in the ecosystem.

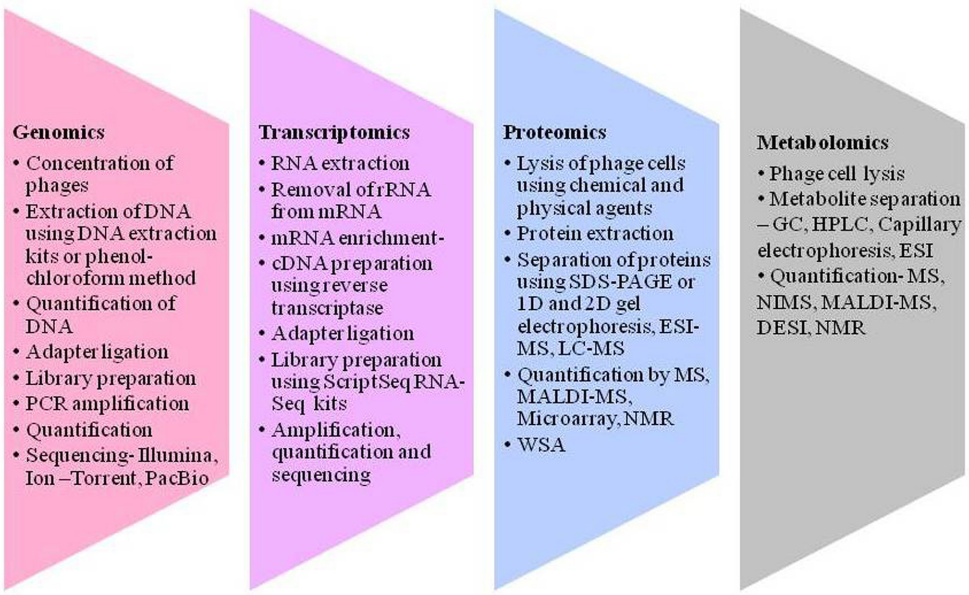


FIGURE 2 | Different techniques to gain an insight into virosphere. Genomics includes concentration of phages, DNA isolation, quantification, and sequencing. Transcriptomics includes processing of RNA converting it to cDNA and sequencing. Proteomics encompasses protein extraction, separation and quantification using several tools like sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), electron spray ionization—mass spectroscopy (ESI-MS), liquid chromatography–mass spectrometry (LC-MS), matrix-assisted laser ionization and deionization (MALDI)-MS and nuclear magnetic resonance (NMR), and Whole phage shotgun analysis (WSA). Metabolomics refers to metabolite extraction separation and quantification in a given time and different metabolites can be analysis using different tools like nanostructure initiator MS (NIMS) and desorption electron spray ionization (DESI) for the understanding of bacteriophage and its interactions.

solicited to confer the claims of omics approach in phage research (**Table 1**).

APPLICATION OF GENOMICS TO REVEAL PHAGE DIVERSITY

Owing to the insufficiency of viral database, there is more than 90% viral dark matter (Hurwitz et al., 2013, 2015). Additionally, the absence of a biomarker gene among phages leads to sequence the whole phage genome for its understanding (Thurber et al., 2009). The genomics of phages would elucidate the genetic composition and putative functional role in the environment. Phage metagenomics would furthermore assist in determining the diversity of phages in a community and reveal novel genes demonstrating phages to be the most diverse beings on the globe (Edwards and Rohwer, 2005). Subsequently, interpretation of functional channels of bacterial viruses would illuminate host-phage interactions (Brum et al., 2015).

A typical genomic experiment begins with isolation of genomic DNA of virus particles (**Figure 3**). The primary step is filtration of the sample through 0.22 μm filters for elimination of bacterial constituents and other contaminations. Samples are then concentrated by ultracentrifugation or polyethylene glycol precipitation (Helms et al., 1985) and subjected to DNase and RNase treatments to exclude residual genomic material from any contaminant bacteria that may pass through 0.22 μm filter. This treated sample would include only virus particles which can be cleaved and their genome can be extracted using kits or standard methods (Adhikary et al., 2014). In order to examine with NGS platforms, DNA is fragmented, ends are repaired and are ligated with adaptors (Holmfeldt et al., 2013). Finally, fragmented DNA library is cleaned and amplified through PCR as well as is quantified and sequenced. Several sequencing platforms are available such as Ion Torrent, Illumina, PacBio which are preferred as per the requirement of their read length, coverage, paired reads, insert size, accuracy, error rates, sequencing yield, run time and sequencing cost (Quail et al., 2012). To reduce chances of bacterial contamination in the library, a section of DNA is PCR amplified for 16S rRNA genes, and if bands are detected, it conveys the presence of host contamination. In the instance of lesser viral DNA yields, amplification using multiple displacement amplification (MDA) can be performed, but it may generate chimeras (Yilmaz et al., 2010). Apparently, amplifications using linkers may depict impartial viromes (Duhaime et al., 2012; Hurwitz et al., 2015). The sequences acquired by sequencing are developed for data filtering and the sequence reads that passes quality check, is mapped to reference genomes or assembled *de novo*. If the sequence of contaminating host is furthermore present in reads even after purifying the sample, it can be distinguished by comparing reads to reference bacterial genome or 16S rRNA database (Hurwitz et al., 2013). For annotation of viral genomes, a database such as NCBI non-redundant nucleotides can be used. ORFs can be determined and annotated using CyVerse (Goff et al., 2011) in the PCPipe application through the iVirus project (Hurwitz et al., 2014).

Bioinformatic tools mine enormous volume of sequence data to determine common patterns that govern microbes in an ecosystem. Viral diversity can be estimated using PHACCS toolkit (Angly et al., 2005). To decode the correlation between virus community and environmental factors, an application called Fizkin by CyVerse cyber infrastructure iVirus project selects 300K reads arbitrarily from viromes and examines it using Jellyfish that generates a matrix of shared sequence counts between each virome pair. This matrix uses an input file for Bayesian network analysis resulting in a table of the relevance of environmental factor that determines the diversity of virus and a social network graph (Hurwitz et al., 2014, 2016). This will assist in ecological profiling of viral communities without requiring assembly and annotation. To elucidate sequence matches with a reference database, BLAST is regularly employed along with MG-RAST (Glass et al., 2010), MetaPhyler (Liu et al., 2010) or CARMA (Gerlach et al., 2009). For a taxonomy of viruses, MEGAN (Huson et al., 2007) software can be used whereas Hidden Markov Models, e.g., HMMER (Finn et al., 2011) are applied to match Pfam or KEGG domains. To find specific viral species present in metagenome, k-mer based algorithms such as CLARK (CLAssifier based on Reduced K-mers) (Ounit et al., 2015), USEARCH (Edgar, 2010), KRAKEN (Wood and Salzberg, 2014), and NBC (Naïve Bayes Classifier) (Rosen et al., 2011) have been applied. Sometimes, whole host genome can be observed in viromes when gene transfer agents (GTAs cluster) have filtered along with virus-like particles (Roux et al., 2013b). GTAs and sporadic contaminations can likewise be recognized using software CLARK (Ounit et al., 2015). Alignment of sequences with reference bacteria genome may reveal a prophage viral element using "recruitment plot" in the bacterial genome. Some of the bioinformatics tools adapted for prophage detection include ACLAME, Prophinder, PHAST and PhiSpy which can serve in confirming phage annotation (Akhter et al., 2012).

Along with elucidating diversity and taxonomy of phages, establishing the origin of genes (bacterial or viral) is vague. This ambiguity occur because of vAMGs which incorporates (enhancing cell metabolism in the host) into host cells or some viruses may also uptake some bacterial genes near the prophage excision site. Nevertheless, during integrating into host tRNAs, phages carry an attachment site (attP) which denote a definite match of a host tRNA gene. Example, integrase gene and an attP site (53 bp) of the *Prochlorococcus* phage P-SS2 is a precise analogue of the host tRNA (attB, 36 bp) of *Prochlorococcus* MIT9313 (Sullivan et al., 2009). Such phages that display a putative attP site and an integrase identical to a host tRNA gene fragment are suggestive of a host-phage association (Mizuno et al., 2013). Metagenomics serves to find diversity among phages but knowledge about interaction among phage and host is relatively scanty. By analyzing the spacers in CRISPR to phage metagenomes, the bacterial host of phage (Dutilh et al., 2014) and phage-host interactions can be deduced (Anderson et al., 2011; Berg Miller et al., 2012; Edwards et al., 2016). Characterizing these constraints is requisite to develop our insight about bacterial-phage co-evolution.

TABLE 1 | Applications of omics approach for understanding bacteriophage–bacterial host interaction.

Sr. No.	Bacteriophage	Bacterial Host	Omics	NGS platform and tools	Applications	References
1	*E. coli* bacteriophages lambda and T4	*E. coli*	Single virus genomics	FACS, MDA, and 454 GS FLX	Viral diversity, evolution, adaptation, and ecology	Allen et al., 2011
2	Rumen bacteriophages	Rumen bacteria	Metagenomics	Roche GS FLX	Phage-bacterial relationship	Berg Miller et al., 2012
3	Ø11b	*Flavobacterium*	Genomics and Proteomics	Shotgun sequencing, PAGE and MALDI-TOF-MS	Genome and proteome analysis	Borriss et al., 2007
4	Marine viruses	Marine bacteria	Metaproteomics	HiSeq-2000	Protein profile study	Brum et al., 2016
5	ØvB_CcoM-IBB_35	*Campylobacter coli*	Genomics and Proteomics	Pyrosequencing, SDS-PAGE and MS	Genomics and protein genes	Carvalho et al., 2012
6	Six different phages	*P. aeruginosa*	Metabolomics	TOF-MS	Alterations in host metabolism during infection	De Smet et al., 2016
7	Ø 812	*Staphylococcus*	Proteomics	1-DE, 2-DE, and MALDI-TOF-MS	Proteome analysis	Eyer et al., 2007
8	A4 Mycobacteriophage Kampy	*M. smegmatis*	Transcriptomics	RNA-Seq and MS	Profiling active genes	Halleran et al., 2015
9	Mycobacteriophage	*Mycobacteria*	Metaproteomics	ABI 3730 or ABI3100	Protein family studies	Hatfull et al., 2006
10	Skin virome	–	Metagenomics	Computational approaches	Exploring the viral dark matter	Hurwitz et al., 2016
11	MS2 bacteriophage	*E. coli*	Metabolomics	Flux balance analysis	Host-pathogen interaction	Jain and Srivastava, 2009
12	vB_EcoP_SU10	*E. coli*	Genomics and proteomics	Roche/454 and Nano LC-MS/MS analysis	Genomics and protein cataloging	Khan Mirzaei et al., 2014
13	RNA Bacteriophages		Metagenomics and, transcriptomics	Various based on subjects studied	Study novel RNA phages and their transcripts	Krishnamurthy et al., 2016
14	ØrV5	*E. coli* O157:H7	Genomics and proteomics	Clone library and primer walking and MS/MS	Using phages as analytical tools	Kropinski et al., 2013
15	Marine surface viruses	Marine surface bacterioplankton	Single cell genomics	Various	Host-virus interaction	Labonté et al., 2015
16	ØKMV	*P. aeruginosa*	Proteomics	SDS-PAGE, WSA and LC-MS	Structural proteins	Lavigne et al., 2006
17	Mycoviruses	Plant pathogenic fungi	Metatranscriptomics	RNAseq- Illumina	Diversity of mycoviruses	Marzano et al., 2016
18	ØvB_YecM_ΦR1-37 (ΦR1-37)	*Yersinia enterocolitica*	Transcriptomics	Whole genome transcription	Host-virus interaction	Leskinen et al., 2016
19	Cyanomyo-virus P-SSM2	marine *Cyanobacterium*	Transcriptomics	whole genome transcription	Host transcriptional response on infection	Lin et al., 2016
20	ØDp-1 and ØCp-1	*S. pneumoniae*	Proteomics	AP/MS	Interactions between bacteria and phage	Mariano et al., 2016
21	T5-like Bacteriophage	*E. coli* O157:H7	Genomics, Proteomics	454 Technology by GS FLX, SDS-PAGE, and MALDI TOF MS	Check for potential as biocontrol agent	Niu et al., 2012
22	ΦMSP	*S. aureus*	Proteomics	SDS-PAGE and 2DE	Proteomic characterization	Sangha et al., 2014
23	Halophilic viral communities	Halophiles	Metatranscriptomics	Microarray, DGGE, and ABI PRISM 310 DNA Sequencer	Metatranscriptome analysis of halophilic viruses	Santos et al., 2011
24	ΦCP39O and ΦCP26F	*C. perfringens*	Genomics and proteomics	Pyrosequencing, 2DE, and MALDI-TOF-MS	Prediction of genome and proteome	Seal et al., 2011
25	ΦTMA	*Thermusthermophilus*	Genomics and proteomics	ABI 3700 Sequencer, PFGE, and SDS-PAGE	Genomic and proteomic characterization	Tamakoshi et al., 2011
26	Different phages against *P. aeruginosa*	*P. aeruginosa*	Proteomics	Affinity purification with MS	Identification of hypothetical proteins	Van den Bossche et al., 2014

(Continued)

TABLE 1 | Continued

Sr. No.	Bacteriophage	Bacterial Host	Omics	NGS platform and tools	Applications	References
27	MC1061 (Φ24B)	*E. coli*	Transcriptomics	Whole-genome shotgun pyrosequencing	Altered functions of host after phage infection	Veses-Garcia et al., 2015
28	ØwV8	*E. coli* O157:H7	Genomics and proteomics	Pyrosequencing, MALDI-MS, and QqTOF	Genomic and proteomic analysis	Villegas et al., 2009
29	ØPhV1 type A and ØPhV1 type B	*Cyanobacteria*	Metagenomics and metatranscriptomics	HiSeq-2000	Interactions between bacteria and phage	Voorhies et al., 2016
30	ØvB_BpuM_BpSp	*Bacillus*	Proteomic analysis	HPLC-ESI-MS/MS	Check a particular activity of protein	Yuan and Gao, 2016
31	ØPaP3	*P. aeruginosa*	Transcriptomics	Microarray, RT-qPCR	Interactions among bacteria and phage	Zhao et al., 2016

1-DE, 1 dimensional electrophoresis; DGGE, denaturing gradient gel electrophoresis; FACS, fluorescence-activated cell sorting; LC-MS, Liquid chromatography–mass spectrometry; MDA, multiple displacement amplification; MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy; PAGE, polyacrylamide gel electrophoresis; PFGE, pulsed field gel electrophoresis; RT-qPCR, Real-time quantitative polymerase chain reaction; RNA-Seq, RNA-Sequencing.

Single Cell Genomics

Apart from metagenomics, attempts have been instigated to investigate only an individual isolate in detail reinforcing our perception of the mechanisms of a specific cell rather than the influences of the entire population. Recently, single cell genomics (SCG) has been promoted to infer the phage genome which is present in or on the surface of host cells in a particular niche without culturing (Lasken and McLean, 2014) and facilitates in assuming sole genetic and metabolic profiles of uncultivable microbes. SCG helps in understanding the interplay between the phage and their host and can ascertain phage genome in a bacterial host cell. To isolate a single cell from an environment, techniques such as flow cytometry (Podar et al., 2007) and micromanipulation (Ishøy et al., 2006) have demonstrated to be advantageous. To sort a single cell precisely, a fluorescence-activated cell sorter (FACSAria^{TM}) with a forward scatter photomultiplier tube (PMT) has been adopted to simplify accurate detection and high-resolution entry of single cell (Picot et al., 2012). Confocal laser scanning microscopy has been applied to support a single phage separation stained by fluorescent dyes lodged into agarose (Luef et al., 2009). Multiple displacement amplification (MDA) (Hosono et al., 2003) utilizes an advanced properties of phi29 DNA polymerase which intensifies a microbial genome at million-fold, sufficient for sequencing using any of the available sequencing platforms.

Interpretation of viral diversity has become easier after the expansion of single virus genomics (Allen et al., 2011) while attending one virus at a time. New computational challenges to analyze the outcomes of SCG using bioinformatics tools have emerged, reflecting the vast opportunity to figure out the *in-situ* phage-host communications. Several bioinformatics tools have been in practice for the classification of prophages-pathogenicity islands such as PIPS (Soares et al., 2012) and HGT- using Alien Hunter (Vernikos and Parkhill, 2006), but these tools seem weak when studying novel phages because of a deficiency of genomic sequences in the viral public database. Because of this constraint, semi-continuous and partial SCG sequences in the database do not allow the accurate identity of isolates (Kalisky and Quake, 2011). However, SCG provides cytoplasmic insights during various interactions with phages like lysogeny, lytic infections, chronic infections and unspecific attachments (Allen et al., 2011). For distinguishing between these synergies, sequences have been examined for integration of phage into the host DNA, portion of phage and host DNA was measured for the speed of single cell MDA reactions and comparisons have been made between the coverage depth of phage and host contigs (Labonté et al., 2015). Phages infecting previously unknown hosts have been discovered in the marine environment using this technique (Roux et al., 2013a, 2014; Labonté et al., 2015).

Non-specific amplification or distortions in the single genome may be a reason for the loss of data, but approximately 90% of DNA can be retrieved using SCG (Rodrigue et al., 2009). A newly developed technology called Hi-C sequencing determines closely arranged genome sequences, like virus-bacterial host genomes within an individual cell (Beitel et al., 2014). The concept of this facility includes genome cross-linking using formaldehyde

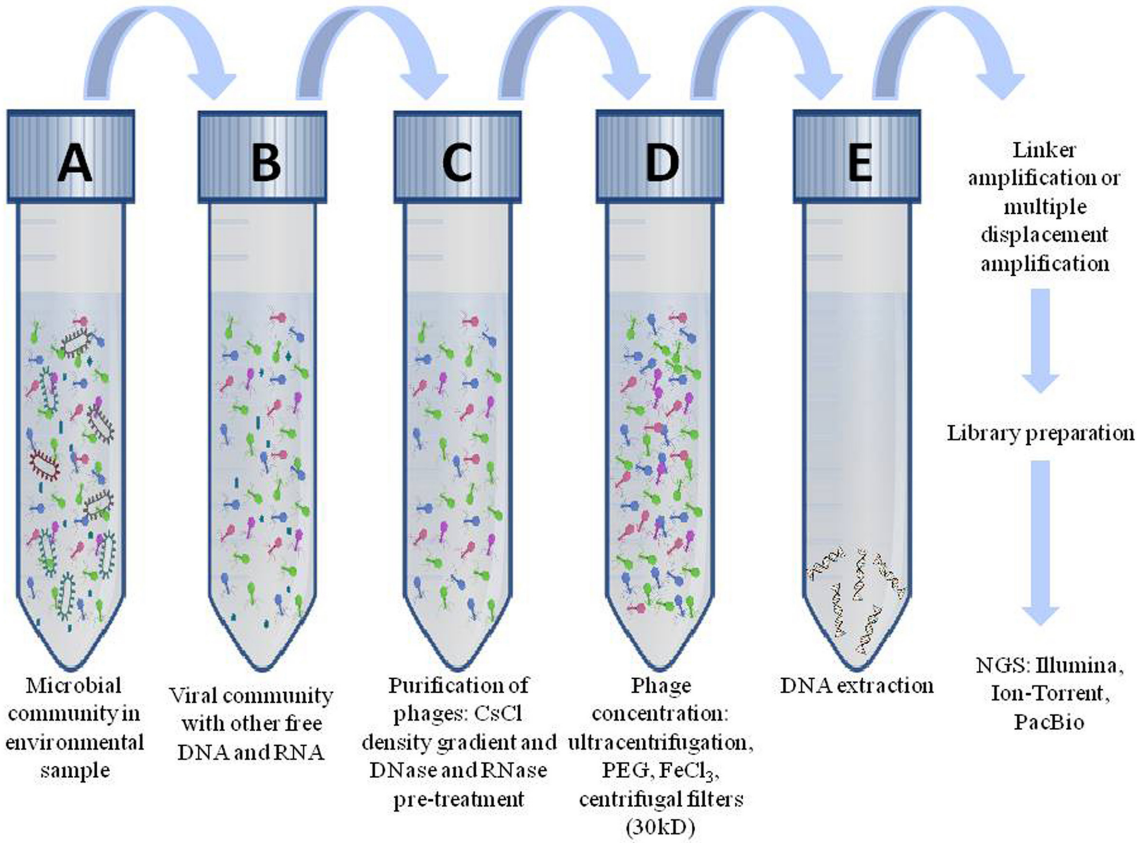


FIGURE 3 | Workflow to study the genomic content of bacteriophages. Genomics of phages initiates by filtering phage particles from microbial community **(A)** through 0.22 μ filter which results in **(B)** virus particles containing residual DNA and RNA of other microbial communities. **(C)** Purification of virus particles from residual DNA and RNA are removed by CsCl density gradient method. **(D)** Phage particles are concentrated using polyethylene glycol or ultra-centrifugation. **(E)** Extraction of viral DNA uses kits or conventional methods. The DNA is amplified and libraries are prepared and subsequently sequenced them.

and a restriction excision, followed by re-ligation of sequences using ligases in a dilute condition that support ligation events between cross-linked DNA fragments, conforming the pairs to each other that were originally in close contiguity (van Berkum et al., 2010). This technique can be adapted to phage-bacterial host communities to figure out close entity while they have been successfully operated for various microbial studies (Beitel et al., 2014; Burton et al., 2014).

Oxford Nanopore sequencing has been utilized where individual DNA molecule is directly sequenced without amplifying or labeling genome with chemical or using visualization tool to recognize the chemical label (Branton et al., 2008). Nanopore sequencing works on the principle that when a voltage is applied to a nanopore imbibed in a conducting liquid, electric current can pass through the nanopore. This electric current is highly responsive to nanopore size and shape such that indeed a single passage of DNA nucleotide pass through nanopore could affect an alteration in the current. The magnitude of current differs based on the type of nucleotides (A, T, G, or C) passing through the nanopore. Thus changing in current corresponds to the precise sequence of a DNA stretch. Viral pathogens have been examined using Nanopore technology (Greninger et al., 2015). Concurrently, MinION sequencer has similarly been recommended which is incredibly rapid, smaller in size, produce 200 kb long reads with high accuracy as well as it has been used to study lambda phage DNA (Mikheyev and Tin, 2014).

Metagenomics and SCG technologies can be strongly adapted to illustrate the exact identity and diversity of phages in an environment which can guide across the dark matter of viral ecosystem (**Figure 4**). Along with investigating the diversity of phages in an environment, to succumb with a coherent outline of their functional aspect in an ecosystem, it becomes imperative to deduce mechanisms underlying transcriptions of phages. Hence, transcriptomic studies provide knowledge about functions of active genes in given condition.

EMPLOYING TRANSCRIPTOMICS TO STUDY ACTIVE PHAGE FUNCTION

Transcriptomics provides a measure to investigate the active microorganisms within a community at a specific time and under a definite array of conditions. Study of the transcriptome

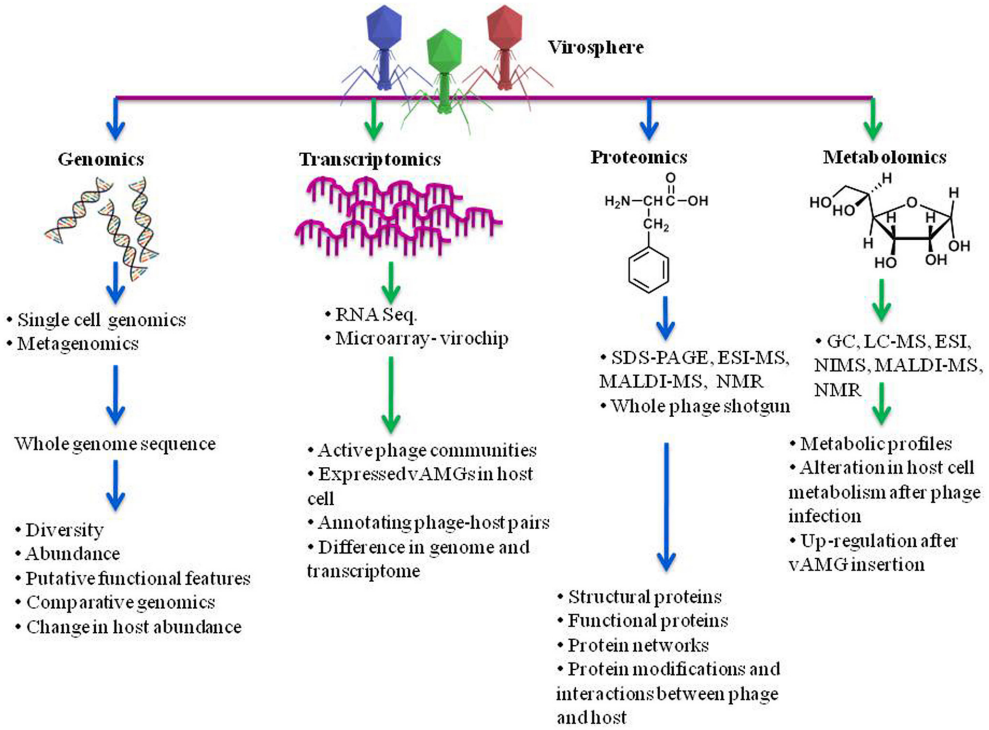


FIGURE 4 | Understanding phage diversity, community interactions and chemical profiles using meta-omics approach. Genomics elucidates the phage diversity, abundance, probable functional features while transcriptomics gives an insight about the actively expressed genes in a community. Proteomics suggests the phage structural proteins, its functions and the proteins responsible for interaction between phage and host. Metabolomics advocates the metabolites produced by hosts in presence and absence of phage infection, alterations in regulation and metabolic profiles after infection. SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; ESI-MS, electron spray ionization–mass spectroscopy; LC-MS, liquid chromatography–mass spectrometry; MALDI-MS, matrix-assisted laser ionization and deionization; NMR, nuclear magnetic resonance; NIMS, nanostructure initiator MS.

is critical to analyze molecular constituent of phages and to understand genome function during a distinct period or situation such as development or infection state. The principle objectives of transcriptomics include recording transcripts of all species including mRNAs, non-coding RNAs and small RNAs, to estimate transcriptional organization of genes in terms of 5′ end and 3′ end, gene splicing and post-transcriptional modification as well as to demonstrate varying activity of each gene under different conditions (Bikel et al., 2015). Several technologies have been developed to inspect and determine transcriptomes, such as DNA hybridization technique, DNA microarray, cDNA-amplified fragment length polymorphism (cDNA-AFLP), expressed sequence tag (EST) sequencing, serial analysis of gene expression (SAGE), massive parallel signature sequencing (MPSS), and RNA-seq (Mutz et al., 2013). DNA hybridization employs fluorescently labeled cDNA to hybridize with DNA templates on microarray chips. However, this tool possesses some reserves as it relies on the hitherto studied genome sequences, high background levels for cross-hybridization and a smaller detection range. Phage meta-transcriptomics may present few difficulties as phages are incredibly diverse and their database is considerably less. Also, the availability of RNA, especially mRNA may be in a rather less volume because of inactive phase when phage is not in association with bacterial hosts. Thus, this may lead to challenges in isolating and enriching mRNA for sequencing.

A typical transcriptomics analysis using NGS includes isolation of the total RNA from the virus particles, depending on the RNA to be sequenced (mRNA, lincRNA or microRNA). Initially, the bacterial fractions are separated and purified from the phage particles. DNase and RNase treatments are administered to filter phage particles from any free DNA or RNA of bacteria. The RNA is thus extracted from virus particle using RNA extraction kits such as RNeasy mini kit. Selective elimination of rRNAs can be achieved using rRNA removing kits or using probes complementary to the rRNA region that is attached to magnetic beads. The mRNA can be enriched by magnetic bead capture method of rRNA, preferential polyadenylation of mRNA or preferential digestion of rRNA through enzymes. The cDNA are synthesized using random hexamers or oligo (dT) primers or priming with poly dT primers after polyadenylation. For amplification, RNA polymerase (Ozsolak and Milos, 2011) or MDA (Gonzalez et al., 2005) or emulsion PCR is/are performed (Tang et al., 2009). The 5′ and/or 3′ ends of the cDNA are then repaired along with adapter ligation, following library cleanup, amplification, quantification and sequencing of the library. Single-end or paired-end libraries can be prepared using kits like ScriptSeq

RNA-Seq library preparation kit (Illumina, San Diego, CA) and can be sequenced on platforms such as Illumina HiSeq2500. Sometimes, conversion of RNA into cDNA introduces bias into the quantification of transcripts, thus a semi-direct sequencing of RNA by-passing the synthesis of cDNA has been established (Hickman et al., 2013).

The bioinformatics analysis of raw data retrieved by transcriptome sequencing uses reference genes and genomes to map against the raw reads or performing a *de novo* assembly for unreported transcriptomes. Mapping of transcripts against reference genome would confer taxonomy and function of active phages. Mapping the active functional pathways would expound the up-regulated, down-regulated or unaffected genes of phage during development or infection cycle. The same can be advised for the bacterial host during infection by a phage. The transcriptome reads which are short can be assembled *de novo* using several softwares such as Trinity. Efficiency and sensitivity of the software are exceptionally promising in procuring full-length transcripts (Ghaffari et al., 2014). The assembled contigs that are obtained by *de novo* or reference-based assembly can be equated with the NCBI viral reference amino acid sequence database using USEARCH (Edgar, 2010). Moreover, the virus annotated hits can be compared with NCBI non-redundant database using BLASTX. Bowtei software can be used to calculate sequence read number and coverage depth (Langmead et al., 2009). Alignment of reads can be prepared using software called MUSCLE (Edgar, 2004) and for constructing neighbor-joining tree, MEGA software can be employed as uses for the bacterial analyses (Tamura et al., 2013).

Several phage meta-transcriptomics studies have been conducted in an attempt to analyze active phage communities. Studies have revealed the effectiveness of phage metagenome for constructing templates in the microarray (Virochip) to annotate and identify the sample (Santos et al., 2011). The total RNA extracted from the sample can be converted to cDNA, labeled and allowed to hybridize with the virochip (Santos et al., 2012). A combination of metagenomics and meta-transcriptomics study would specifically determine active phages in an environment in which phage transcripts may vary as compared to their genomic abundance (Lim et al., 2013). There can be a case when a particular set of family of genes remain less abundant in metagenomic analyses whilst those genes may be remarkably active in meta-transcriptomics dataset and/or vice versa (Franzosa et al., 2014). This insinuates that performing only a metagenomic study may not be a perfect snapshot of functional active genes in a metagenome. To overcome the tedious isolation of viral mRNA from total mRNA, SCG can be conducted along with microarrays to designate the phage-host systems without cultivating them (Santos I. M. et al., 2015). In another study, phage-host pairs have been investigated in which a fosmid viral metagenomic library was constructed and immobilized on microarray "virochip," along with them. The genomes of uncultured bacterial host cells can be sorted by fluorescence-activated cell sorting (FACS) followed by amplification via MDA. Single host cells were hybridized on virochip, and the host cells and immobilized phages with positive results were sequenced (Martínez-Garcia et al., 2014). With this new technique, advancements toward the discovery of phage-host interactions arise in current decade (Santos F. et al., 2015). Moreover, meta-transcriptomics based enzyme discovery from phages can assist in utilizing novel enzymes with specific enzymatic characteristics for the industries and scientific communities (Schoenfeld et al., 2010).

Transcriptomics can be employed to analyze the influencing of a phage on the bacterial host after the phage infection. One such response includes induction of Shiga toxin production and acid resistance in *E. coli* by Shiga toxigenic phages (Veses-Garcia et al., 2015). Studies confirm the fact that host genes get differentially expressed after the phage infection such as a phage "PaP3" had a down regulatory impact on host transcriptional regulators and it proved early genes of phage affected strongly by regulation of hosts (Zhao et al., 2016). This feature of phages can be promoted for formulating a phage therapy. Transcriptomics studies of phage during infection of the host can serve an insight of sequence of transcriptional events, such as initial phase consisting of gene metabolism, DNA synthesis, and regulation genes, is accompanied by a prolonged phase of structural and lysis genes (Halleran et al., 2015). These views can yield information about vAMGs which alters metabolic functions of bacterial host after phage infection. During late phase of phage infection, several up-regulating mechanisms have been observed in bacterial gene expression including stress response and stability of membranes (Leskinen et al., 2016). Additionally, enrichment of ATP synthase and ribosomal protein genes have been revealed during phage infection of phosphorous starved *Cyanobacterium* host (Lin et al., 2016). With further advancement and a few drawbacks, transcriptomics gives an insight of the phage-host interactions and evaluates the regulatory mechanism in bacterial hosts by phage and/or vice versa (effect of host interactions on phage regulation) which are noteworthy for developing phage therapy and comprehend novel phage antimicrobial compounds.

UNDERSTANDING THE PROTEOMIC PROFILE OF PHAGES

A proteome can be represented as a set of all expressed proteins in a cell, tissue or an organism (Theodorescu and Mischak, 2007). Proteomics is a methodology for the characterization of genetic data in a cell or an organism via protein pathways and networks (Petricoin et al., 2002) and for distinguishing the functional implication of proteins (Vlahou and Fountoulakis, 2005). It focuses at cataloging protein expression profiles at a specific period, in a definite location of the cell and as a response to foreign stimulations. It is applied to design a plot of protein networks which can be used to demonstrate interaction among protein in an organism (Corpillo et al., 2004). It provides an estimate of occurrence, quantity and modified state of proteins in an environment in a significant-throughput method.

Genome and transcriptome analysis evaluate the indirect functional profile of a cell or a community whereas proteomic reveal a direct estimate of functional activity of a cell (Schwanhäusser et al., 2011). Abundance profiles of proteins

can be plotted using comparative metaproteomics, while the reduction or increase in the quantity of some proteins may signify a distinct purpose in an organism or during particular situations of phage infection on the bacterial host (Sangha et al., 2014). The post-infection protein expression changes can be classified as (1) function which alters rapidly on phage-infection, but can get reverted back (2) variations that develop gradually and persist consistent or cannot revert back, and (3) alterations that appear abruptly and are maintained for a longer term.

Developments in next generation tools have drastically enhanced quantification and identification of proteins (Schleicher and Wieland, 1978). The proteomic analysis commences with phage concentration accompanied by lysing phage using physical and chemical agents, consequently releasing phage proteins (**Figure 2**). The concentration of proteins can be measured using Bradford's method (Bradford, 1976) or can be denatured using urea (Lavigne et al., 2006) or can be digested by trypsin (Borriss et al., 2007). Several approaches and facilities have been in practice for proteomic studies (Chandramouli and Qian, 2009), however, employing some tools such as a mass spectrometer (MS) and protein-chips (microarray) have significantly contributed in the field (Horgan and Kenny, 2011). Proteins have earlier been detected and quantified using enzyme-linked immunosorbent assay (ELISA) and Western blot where proteins were initially separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Lavigne et al., 2006). Studies have been performed to understand phage proteins using MS after separating by 1D and 2D PAGE (Clement et al., 2013). Additionally, mass-spectrometry-based techniques such as matrix-assisted laser desorption/ionization (MALDI-MS) (Borriss et al., 2007) and electron spray ionization (ESI) (Carvalho et al., 2012) have been established for analyzing various proteins of phages. Recently, fluorescence 2D differential gel electrophoresis has been employed to distinguish between amounts of human lymph and plasma proteins (Clement et al., 2013). Structural proteomics can interpret the structure of proteins thereby determining the functions of novel genes. Lavigne et al. (2006) described structural proteome of phiKMV, a lytic bacteriophage of *Pseudomonas aeruginosa* using SDS-PAGE, LC-ESI-MS/MS, and GC-MS. Nuclear magnetic resonance (NMR) (Horgan and Kenny, 2011) and X-ray crystallography (Drulis-Kawa et al., 2012) can be employed to investigate the interaction between phage-binding protein and receptor site on the bacterial host (Sundell and Ivarsson, 2014).

Protein analysis using MS requires a prior separation of the sample either by 2D-gel electrophoresis (Renesto et al., 2006) or isotope-coded affinity tag (ICAT) labeling (Weston and Hood, 2004), accompanied by digestion into peptides and separating peptides using LC. Microarrays can be applied for assorting protein interaction with DNA, protein or ligands. Protein microarray technique can be exploited in the analytical study to check for presence/absence of a distinct protein in a sample (biomarker detection during phage infections) or for defining function (Uzoma and Zhu, 2013). When phage proteins are immobilized on a microarray chip, it can be applied to probe for complementary bacterial host receptors that bind with phage recognition proteins (Santos F. et al., 2015). Reverse-phase protein microarray can serve as a comparative protein profile in case of phage-infected and uninfected bacterial host (Haider and Pal, 2013). Thus, correlative examination of proteome and genome provide an interpretation of the post-translational modifications.

Functional identification of hypothetical phage proteins is performed using MS analysis after affinity purification of host protein mixtures (Van den Bossche et al., 2014). MS/MS spectra can be interpreted using SEQUEST (http://fields.scripps.edu/sequest/) or Mascot (Matrix Sciences) and classifying using DTASelect and Contrast softwares (Tabb et al., 2002). Proteomic phage display techniques are similarly employed to identify target proteins and consensus motifs (Sundell and Ivarsson, 2014). Whole phage shotgun analysis (WSA) is a recently developed technique for protein analysis using NGS platform. It is a culture-independent technique which offers annotation of proteins associated with phages. WSA combines all structural proteins separated on the basis of mass and charge before identification (Lavigne et al., 2006). After separation, the data can be annotated to open reading frames (ORFs) by aligning with reference protein sequences using BLASTP. HHpred is another tool for assigning the protein structure (Hildebrand et al., 2009). The function and evolution of identified proteins can be determined by program COGnitor (www.ncbi.nlm.nih.gov/COG) and InterProScan to find conserved domains (Eyer et al., 2007). When a predicted protein does not match along known proteins from the database, protein clustering can be developed for the comparative analyses to assess the protein diversity (Hurwitz et al., 2013; Brum et al., 2016). Some software can extract the data from MS and microarray and decipher protein identification using databases such as UniProt (http://www.uniprot.org/), PROSITE (http://prosite.expasy.org/), Pfam, Conserved Domain and PDB databases. Thus, with an advent in high-throughput proteomic technology, analytical tools, bioinformatics software and database, research on proteins have emerged as an easy task to elucidate protein matter in an environment.

CATALOGING THE METABOLOME OF VIROSPHERE

The breakdown products of metabolism or intermediates involved in the process of metabolism are termed as metabolites. Metabolites can be (1) primary- which are precisely involved in process of metabolism or (2) secondary- which may not directly take part in the growth of an organism. The metabolome of an organism corresponds to a set of metabolites including hormones, intermediates, signaling and secondary molecules in a particular cell, tissue, organ or an organism (Griffin and Vidal-Puig, 2008; Jordan et al., 2009). To explain the physiology of a particular cell, the study of metabolites is very substantial as every cell possesses a specific metabolic catalog which can influence the accurate implication of function of a cell or an organism (Nicholson and Wilson, 2003; Zhang et al., 2016). These are results of gene transcriptional and

translational mechanisms which remain exceptionally complex, hence variations in metabolites intensify as compared to variations among transcriptome and proteome.

Various approaches have been established for separation and detection of metabolites, chiefly when metabolites are of higher molecular mass. The segregations of metabolites can be carried out using gas chromatography (GC) and high performance liquid chromatography (HPLC), capillary electrophoresis, electron spray ionization (ESI) accompanied by GC, atmospheric-pressure chemical ionization (APCI) on the ground of characteristics of metabolite to be processed (Alonso et al., 2015). Detection of separated metabolites have furthermore been attainable by using nanostructure-initiator MS (NIMS), MALDI-MS, secondary ion mass spectrometry (SIMS), desorption electron spray ionization (DESI), and NMR (Drexler et al., 2007; Cornett et al., 2008; Wiseman et al., 2008; Greer et al., 2011). Statistical tools are additionally applicable for the evaluation of elicited data such as XCMS (Patti et al., 2013), MZmine (Katajamaa et al., 2006), MetAlign (Lommen, 2009), MathDAMP (Baran et al., 2006), and LCMStats[1] (Gahlaut et al., 2013). The metabolic database is available in form of METLIN (Smith et al., 2005).

Metabolomics would serve in interpreting the significance of active phage community on the environment in real time. Based on distinct phases of the phage infection or metabolic profile of the host infected with a phage, the gene markers can be inferred. The modification in host-cell metabolism by phage-encoded genes (vAMGs) into the host genome, is described as a virocell amendment (Rosenwasser et al., 2016). Studying highly specific metabolic profiles of a virocell can improve in interpreting metabolic profile of vAMGs. Such comparisons were conducted to recognize host-viral interactions (Vardi et al., 2009, 2012; Fulton et al., 2014). Metabolomic analysis of phage interprets the influence of the vAMGs which is responsible for enhancing nucleotide biosynthesis (De Smet et al., 2016) via degrading host macromolecules such as DNA through catabolic pathways. The vAMGs encoded nucleases can generally degrade host DNA and encoded triglyceride lipase can degrade host triacylglycerols which yield energy and ultimately engages in the the formation of virus membrane (Malitsky et al., 2016). Example, ceramidase in *Mimivirus* helps in the catabolism of sphingolipids (Arslan et al., 2011). Thus, the vAMGs develop the metabolic potential of virocell through triggering novel enzymes which were not present in host machinery prior to phage infection (DeAngelis et al., 1997; Graves et al., 1999). The vAMGs can serve as a shunt between phage and their host by imparting several functional genes from one another especially assisting during stress conditions (Rosenwasser et al., 2016). These mechanisms illustrate unique attributes of gene products of phage that can mediate dynamics phage-host interaction, as an effect, shaping the microbial communities in an environment. Thus, biochemical composition and metabolic profile of bacterial hosts are greatly governed by phages and released metabolites in the environment influence the microbial food web (Miki et al., 2008). Cataloging the metabolome of phages can elucidate special phage-derived metabolites which usually act as decision making between lytic or lysogenic lifecycle in virocell. Study of metagenomic and metabolomic profiles can simultaneously determine whether the metabolites are encoded by the phage or the host. Furthermore, metabolic profiles of phage can aid in tracing a novel biomarker to recognize the nutrient source in biogeochemical cycles. Thus, the advents in the omics approaches utilizing NGS techniques bear a tremendous potential in exploring virosphere and thus the microbial world.

[1]Lcmstats: LCMStats: an [R (programming language)] package for detailed analysis of LCMS data. http://sourceforge.net/projects/lcmstats.html.

CONCLUSIONS

Advancements in the field of NGS have facilitated the discoveries on the verge of a revolution in the course of microbial research. There has been a tremendous microbial data generated about the microbes present on Earth and their diversity and functional roles in regulating the ecosystem. Progress in interpreting the phage diversity and functions as well as the interactions among phages and their hosts are promising using the "omics" concepts. This would illuminate the function of phages in regulating microbial diversity by HGTs, governing the biogeochemical cycles, host population controls and determining the novel biomarkers. NGS will also nurture the upcoming phage therapy research for limiting MDR pathogens. With strong prospects in the field and developments in phage database, "omics" approach is witnessing a remarkable motive for a transformation in the yet uncultivable microbial research.

AUTHOR CONTRIBUTIONS

KP, PD, and RS conceived and designed the work. KP, SG, and RS wrote the manuscript. PD, RK, and RS carefully checked the manuscript and corrected. All of the authors contributed to the discussion and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are thankful to Agharkar Research Institute, John Innes Center and Saurashtra University for support and encouragement.

REFERENCES

Ackermann, H. W. (2009). Phage classification and characterization. *Methods Mol. Biol.* 501, 127–140. doi: 10.1007/978-1-60327-164-6_13

Adhikary, A. K., Hanaoka, N., and Fujimoto, T. (2014). Simple and cost-effective restriction endonuclease analysis of human adenoviruses. *Biomed. Res. Int.* 2014:363790. doi: 10.1155/2014/363790

Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 40:e126. doi: 10.1093/nar/gks406

Allen, L. Z., Ishoey, T., Novotny, M. A., McLean, J. S., Lasken, R. S., and Williamson, S. J. (2011). Single virus genomics: a new tool for virus discovery. *PLoS ONE* 6:e17722. doi: 10.1371/journal.pone.0017722

Alonso, A., Marsal, S., and Julia, A. (2015). Analytical methods in untargeted metabolomics: state of the art in 2015. *Front. Bioeng. Biotechnol.* 3:23. doi: 10.3389/fbioe.2015.00023

Anderson, R. E., Brazelton, W. J., and Baross, J. A. (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77, 120–133. doi: 10.1111/j.1574-6941.2011.01090.x

Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., et al. (2005). PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* 6:41. doi: 10.1186/1471-2105-6-41

Apel, K., and Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701

Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and Claverie, J. M. (2011). Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17486–17491. doi: 10.1073/pnas.1110889108

Baran, R., Kochi, H., Saito, N., Suematsu, M., Soga, T., Nishioka, T., et al. (2006). MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics* 7:530. doi: 10.1186/1471-2105-7-530

Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Michelmore, R. W., Eisen, J. A., et al. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* 2:e415. doi: 10.7717/peerj.415

Berg Miller, M. E., Yeoman, C. J., Chia, N., Tringe, S. G., Angly, F. E., Edwards, R. A., et al. (2012). Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome. *Environ. Microbiol.* 14, 207–227. doi: 10.1111/j.1462-2920.2011.02593.x

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62, 293–300.

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberon, X., et al. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Comput. Struct. Biotechnol. J.* 13, 390–401. doi: 10.1016/j.csbj.2015.06.001

Borriss, M., Lombardot, T., Glockner, F. O., Becher, D., Albrecht, D., and Schweder, T. (2007). Genome and proteome characterization of the psychrophilic Flavobacterium bacteriophage 11b. *Extremophiles* 11, 95–104. doi: 10.1007/s00792-006-0014-5

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., et al. (2008). The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* 26, 1146–1153. doi: 10.1038/nbt.1495

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., et al. (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc. Biol. Sci.* 271, 565–574. doi: 10.1098/rspb.2003.2628

Brum, J. R., Ignacio-Espinoza, J. C., Kim, E. H., Trubl, G., Jones, R. M., Roux, S., et al. (2016). Illuminating structural proteins in viral "dark matter" with metaproteomics. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2436–2441. doi: 10.1073/pnas.1525139113

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., et al. (2015). Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. doi: 10.1126/science.1261498

Brüssow, H., Canchaya, C., and Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602. doi: 10.1128/MMBR.68.3.560-602.2004

Burton, J. N., Liachko, I., Dunham, M. J., and Shendure, J. (2014). Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3 (Bethesda)* 4, 1339–1346. doi: 10.1534/g3.114.011825

Carvalho, C. M., Kropinski, A. M., Lingohr, E. J., Santos, S. B., King, J., and Azeredo, J. (2012). The genome and proteome of a *Campylobacter coli* bacteriophage vB_CcoM-IBB_35 reveal unusual features. *Virol. J.* 9:35. doi: 10.1186/1743-422x-9-35

Chandramouli, K., and Qian, P.-Y. (2009). Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum. Genomics Proteomics* 2009:239204. doi: 10.4061/2009/239204

Clement, C. C., Aphkhazava, D., Nieves, E., Callaway, M., Olszewski, W., Rotzschke, O., et al. (2013). Protein expression profiles of human lymph and plasma mapped by 2D-DIGE and 1D SDS-PAGE coupled with nanoLC-ESI-MS/MS bottom-up proteomics. *J. Proteomics* 78, 172–187. doi: 10.1016/j.jprot.2012.11.013

Cornett, D. S., Frappier, S. L., and Caprioli, R. M. (2008). MALDI-FTICR imaging mass spectrometry of drugs and metabolites in tissue. *Anal. Chem.* 80, 5648–5653. doi: 10.1021/ac800617s

Corpillo, D., Gardini, G., Vaira, A. M., Basso, M., Aime, S., Accotto, G. P., et al. (2004). Proteomics as a tool to improve investigation of substantial equivalence in genetically modified organisms: the case of a virus-resistant tomato. *Proteomics* 4, 193–200. doi: 10.1002/pmic.200300540

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0

DeAngelis, P. L., Jing, W., Graves, M. V., Burbank, D. E., and Van Etten, J. L. (1997). Hyaluronan synthase of chlorella virus PBCV-1. *Science* 278, 1800–1803. doi: 10.1126/science.278.5344.1800

De Smet, J., Zimmermann, M., Kogadeeva, M., Ceyssens, P. J., Vermaelen, W., Blasdel, B., et al. (2016). High coverage metabolomics analysis reveals phage-specific alterations to Pseudomonas aeruginosa physiology during infection. *ISME J.* 10, 1823–1835. doi: 10.1038/ismej.2016.3

Drexler, D. M., Garrett, T. J., Cantone, J. L., Diters, R. W., Mitroka, J. G., Prieto Conaway, M. C., et al. (2007). Utility of imaging mass spectrometry (IMS) by matrix-assisted laser desorption ionization (MALDI) on an ion trap mass spectrometer in the analysis of drugs and metabolites in biological tissues. *J. Pharmacol. Toxicol. Methods* 55, 279–288. doi: 10.1016/j.vascn.2006.11.004

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A. S., and Lavigne, R. (2012). Learning from bacteriophages–advantages and limitations of phage and phage-encoded protein applications. *Curr. Protein Pept. Sci.* 13, 699–722. doi: 10.2174/138920312804871193

Duhaime, M. B., Deng, L., Poulos, B. T., and Sullivan, M. B. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* 14, 2526–2537. doi: 10.1111/j.1462-2920.2012.02791.x

Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G., Boling, L., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5:4498. doi: 10.1038/ncomms5498

Dutta, C., and Pan, A. (2002). Horizontal gene transfer and bacterial diversity. *J. Biosci.* 27, 27–33. doi: 10.1007/BF02703681

Eberwine, J., Sul, J. Y., Bartfai, T., and Kim, J. (2014). The promise of single-cell sequencing. *Nat. Methods* 11, 25–27. doi: 10.1038/nmeth.2769

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edwards, R. A., McNair, K., Faust, K., Raes, J., and Dutilh, B. E. (2016). Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* 40, 258–272. doi: 10.1093/femsre/fuv048

Edwards, R. A., and Rohwer, F. (2005). Viral metagenomics. *Nat. Rev. Microbiol.* 3, 504–510. doi: 10.1038/nrmicro1163

Eyer, L., Pantucek, R., Zdrahal, Z., Konecna, H., Kasparek, P., Ruzickova, V., et al. (2007). Structural protein analysis of the polyvalent staphylococcal bacteriophage 812. *Proteomics* 7, 64–72. doi: 10.1002/pmic.200600280

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2329–E2338. doi: 10.1073/pnas.1319284111

Fulton, J. M., Fredricks, H. F., Bidle, K. D., Vardi, A., Kendrick, B. J., Ditullio, G. R., et al. (2014). Novel molecular determinants of viral susceptibility and resistance in the lipidome of *Emiliania huxleyi*. *Environ. Microbiol.* 16, 1137–1149. doi: 10.1111/1462-2920.12358

Gahlaut, A., Vikas., Dahiya, M., Gothwal, A., Kulharia, M., Chhillar, A. K., et al. (2013). Proteomics & metabolomics: mapping biochemical regulations. *Drug Invention Today* 5, 321–326. doi: 10.1016/j.dit.2013.08.007

Gerlach, W., Junemann, S., Tille, F., Goesmann, A., and Stoye, J. (2009). WebCARMA: a web application for the functional and taxonomic classification of unassembled metagenomic reads. *BMC Bioinformatics* 10:430. doi: 10.1186/1471-2105-10-430

Ghaffari, N., Sanchez-Flores, A., Doan, R., Garcia-Orozco, K. D., Chen, P. L., Ochoa-Leyva, A., et al. (2014). Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Sci. Rep.* 4:7081. doi: 10.1038/srep07081

Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D., and Meyer, F. (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* 2010:pdb.prot5368. doi: 10.1101/pdb.prot5368

Goff, S. A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A. E., Gessler, D., et al. (2011). The iPlant collaborative: cyberinfrastructure for plant biology. *Front. Plant Sci.* 2:34. doi: 10.3389/fpls.2011.00034

Gonzalez, J. M., Portillo, M. C., and Saiz-Jimenez, C. (2005). Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. *Environ. Microbiol.* 7, 1024–1028. doi: 10.1111/j.1462-2920.2005.00779.x

Graves, M. V., Burbank, D. E., Roth, R., Heuser, J., Deangelis, P. L., and Van Etten, J. L. (1999). Hyaluronan synthesis in virus PBCV-1-infected chlorella-like green algae. *Virology* 257, 15–23. doi: 10.1006/viro.1999.9628

Greer, T., Sturm, R., and Li, L. (2011). Mass spectrometry imaging for drugs and metabolites. *J. Proteomics* 74, 2617–2631. doi: 10.1016/j.jprot.2011.03.032

Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., et al. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* 7:99. doi: 10.1186/s13073-015-0220-9

Griffin, J. L., and Vidal-Puig, A. (2008). Current challenges in metabolomics for diabetes research: a vital functional genomic tool or just a ploy for gaining funding? *Physiol. Genomics* 34, 1–5. doi: 10.1152/physiolgenomics.00009.2008

Haider, S., and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* 14, 91–110. doi: 10.2174/1389202911314020003

Halleran, A., Clamons, S., and Saha, M. (2015). Transcriptomic characterization of an infection of *Mycobacterium smegmatis* by the cluster a4 mycobacteriophage kampy. *PLoS ONE* 10:e0141100. doi: 10.1371/journal.pone.0141100

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685. doi: 10.1128/MMBR.68.4.669-685.2004

Hatfull, G. F., Pedulla, M. L., Jacobs-Sera, D., Cichon, P. M., Foley, A., Ford, M. E., et al. (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet.* 2:e92. doi: 10.1371/journal.pgen.0020092

Helms, C., Graham, M. Y., Dutchik, J. E., and Olson, M. V. (1985). A new method for purifying lambda DNA from phage lysates. *DNA* 4, 39–49. doi: 10.1089/dna.1985.4.39

Hickman, S. E., Kingery, N. D., Ohsumi, T. K., Borowsky, M. L., Wang, L. C., Means, T. K., et al. (2013). The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* 16, 1896–1905. doi: 10.1038/nn.3554

Hildebrand, A., Remmert, M., Biegert, A., and Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins* 77 (Suppl. 9), 128–132. doi: 10.1002/prot.22499

Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., Verberkmoes, N. C., et al. (2013). Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12798–12803. doi: 10.1073/pnas.1305956110

Horgan, R. P., and Kenny, L. C. (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *J. Obstet. Gynaecol.* 13, 189–195. doi: 10.1576/toag.13.3.189.27672

Hosono, S., Faruqi, A. F., Dean, F. B., Du, Y., Sun, Z., Wu, X., et al. (2003). Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* 13, 954–964. doi: 10.1101/gr.816903

Hurwitz, B. L., Brum, J. R., and Sullivan, M. B. (2015). Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean virome. *ISME J.* 9, 472–484. doi: 10.1038/ismej.2014.143

Hurwitz, B. L., Hallam, S. J., and Sullivan, M. B. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* 14:R123. doi: 10.1186/gb-2013-14-11-r123

Hurwitz, B. L., U'ren, J. M., and Youens-Clark, K. (2016). Computational prospecting the great viral unknown. *FEMS Microbiol. Lett.* 363:fnw077. doi: 10.1093/femsle/fnw077

Hurwitz, B. L., Westveld, A. H., Brum, J. R., and Sullivan, M. B. (2014). Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc. Natl. Acad. Sci. U.S.A.* 111, 10714–10719. doi: 10.1073/pnas.1319778111

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Ishøy, T., Kvist, T., Westermann, P., and Ahring, B. K. (2006). An improved method for single cell isolation of prokaryotes from meso-, thermo- and hyperthermophilic environments using micromanipulation. *Appl. Microbiol. Biotechnol.* 69, 510–514. doi: 10.1007/s00253-005-0014-x

Jain, R., and Srivastava, R. (2009). Metabolic investigation of host/pathogen interaction using MS2-infected *Escherichia coli*. *BMC Syst. Biol.* 3:121. doi: 10.1186/1752-0509-3-121

Jordan, K. W., Nordenstam, J., Lauwers, G. Y., Rothenberger, D. A., Alavi, K., Garwood, M., et al. (2009). Metabolomic characterization of human rectal adenocarcinoma with intact tissue magnetic resonance spectroscopy. *Dis. Colon Rectum* 52, 520–525. doi: 10.1007/DCR.0b013e31819c9a2c

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 299, 27–51. doi: 10.1006/jmbi.2000.3729

Kalisky, T., and Quake, S. R. (2011). Single-cell genomics. *Nat. Methods* 8, 311–314. doi: 10.1038/nmeth0411-311

Katajamaa, M., Miettinen, J., and Oresic, M. (2006). MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22, 634–636. doi: 10.1093/bioinformatics/btk039

Khan Mirzaei, M., Eriksson, H., Kasuga, K., Haggard-Ljungquist, E., and Nilsson, A. S. (2014). Genomic, proteomic, morphological, and phylogenetic analyses of vB_EcoP_SU10, a podoviridae phage with C3 morphology. *PLoS ONE* 9:e116294. doi: 10.1371/journal.pone.0116294

Krishnamurthy, S. R., Janowski, A. B., Zhao, G., Barouch, D., and Wang, D. (2016). Hyperexpansion of RNA bacteriophage diversity. *PLoS Biol.* 14:e1002409. doi: 10.1371/journal.pbio.1002409

Kropinski, A. M., Waddell, T., Meng, J., Franklin, K., Ackermann, H. W., Ahmed, R., et al. (2013). The host-range, genomics and proteomics of *Escherichia coli* O157:H7 bacteriophage rV5. *Virol. J.* 10:76. doi: 10.1186/1743-422X-10-76

Labonté, J. M., Swan, B. K., Poulos, B., Luo, H., Koren, S., Hallam, S. J., et al. (2015). Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* 9, 2386–2399. doi: 10.1038/ismej.2015.48

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Lasken, R. S., and McLean, J. S. (2014). Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat. Rev. Genet.* 15, 577–584. doi: 10.1038/nrg3785

Lavigne, R., Noben, J. P., Hertveldt, K., Ceyssens, P. J., Briers, Y., Dumont, D., et al. (2006). The structural proteome of *Pseudomonas aeruginosa* bacteriophage phiKMV. *Microbiology* 152, 529–534. doi: 10.1099/mic.0.28431-0

Leskinen, K., Blasdel, B. G., Lavigne, R., and Skurnik, M. (2016). RNA-sequencing reveals the progression of phage-host interactions between phir1-37 and *Yersinia enterocolitica*. *Viruses* 8:111. doi: 10.3390/v8040111

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., et al. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2117–2122. doi: 10.1073/pnas.0712038105

Lim, Y. W., Schmieder, R., Haynes, M., Willner, D., Furlan, M., Youle, M., et al. (2013). Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J. Cyst. Fibros.* 12, 154–164. doi: 10.1016/j.jcf.2012.07.009

Lin, X., Ding, H., and Zeng, Q. (2016). Transcriptomic response during phage infection of a marine cyanobacterium under phosphorus-limited conditions. *Environ. Microbiol.* 18, 450–460. doi: 10.1111/1462-2920.13104

Liu, B., Gibbons, T., Ghodsi, M., and Pop, M. (2010). "MetaPhyler: taxonomic profiling for metagenomic sequences," in *IEEE International Conference on Bioinformatics and Biomedicine* (Hong-Kong: IEEE), 95–100.

Lommen, A. (2009). MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.* 81, 3079–3086. doi: 10.1021/ac900036d

Luef, B., Neu, T. R., and Peduzzi, P. (2009). Imaging and quantifying virus fluorescence signals on aquatic aggregates: a new method and its implication for aquatic microbial ecology. *FEMS Microbiol. Ecol.* 68, 372–380. doi: 10.1111/j.1574-6941.2009.00675.x

Malitsky, S., Ziv, C., Rosenwasser, S., Zheng, S., Schatz, D., Porat, Z., et al. (2016). Viral infection of the marine alga *Emiliania huxleyi* triggers lipidome remodeling and induces the production of highly saturated triacylglycerol. *New Phytol.* 210, 88–96. doi: 10.1111/nph.13852

Mariano, R., Wuchty, S., Vizoso-Pinto, M. G., Hauser, R., and Uetz, P. (2016). The interactome of *Streptococcus pneumoniae* and its bacteriophages show highly specific patterns of interactions among bacteria and their phages. *Sci. Rep.* 6:24597. doi: 10.1038/srep24597

Martínez-Garcia, M., Santos, F., Moreno-Paz, M., Parro, V., and Antón, J. (2014). Unveiling viral-host interactions within the 'microbial dark matter'. *Nat. Commun.* 5:4542. doi: 10.1038/ncomms5542

Marzano, S. Y., Nelson, B. D., Ajayi-Oyetunde, O., Bradley, C. A., Hughes, T. J., Hartman, G. L., et al. (2016). Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. *J. Virol.* 90, 6846–6863. doi: 10.1128/JVI.00357-16

Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Mikheyev, A. S., and Tin, M. M. (2014). A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* 14, 1097–1102. doi: 10.1111/1755-0998.12324

Miki, T., Nakazawa, T., Yokokawa, T., and Nagata, T. (2008). Functional consequences of viral impacts on bacterial communities: a food-web model analysis. *Freshw. Biol.* 53, 1142–1153. doi: 10.1111/j.1365-2427.2007.01934.x

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., et al. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625. doi: 10.1101/gr.122705.111

Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., and Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genet.* 9:e1003987. doi: 10.1371/journal.pgen.1003987

Mutz, K. O., Heilkenbrinker, A., Lonne, M., Walter, J. G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* 24, 22–30. doi: 10.1016/j.copbio.2012.09.004

Nicholson, J. K., and Wilson, I. D. (2003). Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov.* 2, 668–676. doi: 10.1038/nrd1157

Niu, Y. D., Stanford, K., Kropinski, A. M., Ackermann, H. W., Johnson, R. P., She, Y. M., et al. (2012). Genomic, proteomic and physiological characterization of a T5-like bacteriophage for control of Shiga toxin-producing *Escherichia coli* O157:H7. *PLoS ONE* 7:e34585. doi: 10.1371/journal.pone.0034585

Ounit, R., Wanamaker, S., Close, T. J., and Lonardi, S. (2015). CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics* 16:236. doi: 10.1186/s12864-015-1419-2

Ozsolak, F., and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98. doi: 10.1038/nrg2934

Parmar, K. M., Hathi, Z. J., and Dafale, N. A. (2017). Control of multidrug-resistant gene flow in the environment through bacteriophage intervention. *Appl Biochem Biotechnol.* 181, 1007–1029. doi: 10.1007/s12010-016-2265-7

Patti, G. J., Tautenhahn, R., Rinehart, D., Cho, K., Shriver, L. P., Manchester, M., et al. (2013). A view from above: cloud plots to visualize global metabolomic data. *Anal. Chem.* 85, 798–804. doi: 10.1021/ac3029745

Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C., and Liotta, L. A. (2002). Clinical proteomics: translating benchside promise into bedside reality. *Nat. Rev. Drug Discov.* 1, 683–695. doi: 10.1038/nrd891

Picot, J., Guerin, C. L., Le Van Kim, C., and Boulanger, C. M. (2012). Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology* 64, 109–130. doi: 10.1007/s10616-011-9415-0

Podar, M., Abulencia, C. B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J. A., et al. (2007). Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* 73, 3205–3214. doi: 10.1128/AEM.02985-06

Prestel, E., Salamitou, S., and Dubow, M. S. (2008). An examination of the bacteriophages and bacteria of the Namib desert. *J. Microbiol.* 46, 364–372. doi: 10.1007/s12275-008-0007-4

Prigent, M., Leroy, M., Confalonieri, F., Dutertre, M., and Dubow, M. S. (2005). A diversity of bacteriophage forms and genomes can be isolated from the surface sands of the Sahara Desert. *Extremophiles* 9, 289–296. doi: 10.1007/s00792-005-0444-5

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341. doi: 10.1186/1471-2164-13-341

Renesto, P., Abergel, C., Decloquement, P., Moinier, D., Azza, S., Ogata, H., et al. (2006). Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J. Virol.* 80, 11678–11685. doi: 10.1128/JVI.00940-06

Rodrigue, S., Malmstrom, R. R., Berlin, A. M., Birren, B. W., Henn, M. R., and Chisholm, S. W. (2009). Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE* 4:e6864. doi: 10.1371/journal.pone.0006864

Rohwer, F., and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* 459, 207–212. doi: 10.1038/nature08060

Rosen, G. L., Reichenberger, E. R., and Rosenfeld, A. M. (2011). NBC: the Naive Bayes Classification tool webserver for taxonomic classification of metagenomic reads. *Bioinformatics* 27, 127–129. doi: 10.1093/bioinformatics/btq619

Rosenwasser, S., Ziv, C., Creveld, S. G., and Vardi, A. (2016). Virocell metabolism: metabolic innovations during host-virus interactions in the ocean. *Trends Microbiol.* 24, 821–832. doi: 10.1016/j.tim.2016.06.006

Roux, S., Enault, F., Bronner, G., Vaulot, D., Forterre, P., and Krupovic, M. (2013a). Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nat. Commun.* 4:2700. doi: 10.1038/ncomms3700

Roux, S., Hawley, A. K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., et al. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife* 3:e03125. doi: 10.7554/eLife.03125

Roux, S., Krupovic, M., Debroas, D., Forterre, P., and Enault, F. (2013b). Assessment of viral community functional potential from viral metagenomes

may be hampered by contamination with cellular sequences. *Open Biol.* 3:130160. doi: 10.1098/rsob.130160

Sanchez, S. E., Cuevas, D. A., Rostron, J. E., Liang, T. Y., Pivaroff, C. G., Haynes, M. R., et al. (2015). Phage phenomics: physiological approaches to characterize novel viral proteins. *J. Vis. Exp.* e52854. doi: 10.3791/52854

Sangha, K. K., Kumar, B. V., Agrawal, R. K., Deka, D., and Verma, R. (2014). Proteomic characterization of lytic bacteriophages of *Staphylococcus aureus* Isolated from Sewage Affluent of India. *Int. Sch. Res. Notices* 2014:265298. doi: 10.1155/2014/265298

Santos, F., Martínez-García, M., Parro, V., and Antón, J. (2015). Microarray tools to unveil viral-microbe interactions in nature. *Front. Ecol. Evol.* 3:31. doi: 10.3389/fevo.2014.00031

Santos, F., Moreno-Paz, M., Meseguer, I., Lopez, C., Rossello-Mora, R., Parro, V., et al. (2011). Metatranscriptomic analysis of extremely halophilic viral communities. *ISME J.* 5, 1621–1633. doi: 10.1038/ismej.2011.34

Santos, F., Yarza, P., Parro, V., Meseguer, I., Rossello-Mora, R., and Anton, J. (2012). Culture-independent approaches for studying viruses from hypersaline environments. *Appl. Environ. Microbiol.* 78, 1635–1643. doi: 10.1128/AEM.07175-11

Santos, Í. M., Da Rosa, E. A., Graf, T., Ferreira, L. G., Petry, A., Cavalheiro, F., et al. (2015). Analysis of immunological, viral, genetic, and environmental factors that might be associated with decreased susceptibility to HIV infection in serodiscordant couples in Florianopolis, Southern Brazil. *AIDS Res. Hum. Retroviruses* 31, 1116–1125. doi: 10.1089/aid.2015.0168

Schleicher, E., and Wieland, O. H. (1978). Evaluation of the Bradford method for protein determination in body fluids. *J. Clin. Chem. Clin. Biochem.* 16, 533–534.

Schoenfeld, T., Liles, M., Wommack, K. E., Polson, S. W., Godiska, R., and Mead, D. (2010). Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol.* 18, 20–29. doi: 10.1016/j.tim.2009.10.001

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Seal, B. S., Fouts, D. E., Simmons, M., Garrish, J. K., Kuntz, R. L., Woolsey, R., et al. (2011). Clostridium perfringens bacteriophages PhiCP39O and PhiCP26F: genomic organization and proteomic analysis of the virions. *Arch. Virol.* 156, 25–35. doi: 10.1007/s00705-010-0812-z

Short, C. M., and Suttle, C. A. (2005). Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl. Environ. Microbiol.* 71, 480–486. doi: 10.1128/AEM.71.1.480-486.2005

Sime-Ngando, T., and Colombet, J. (2009). [Virus and prophages in aquatic ecosystems]. *Can. J. Microbiol.* 55, 95–109. doi: 10.1139/W08-099

Smith, C. A., O'maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., et al. (2005). METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751. doi: 10.1097/01.ftd.0000179845.53213.39

Soares, S. C., Abreu, V. A., Ramos, R. T., Cerdeira, L., Silva, A., Baumbach, J., et al. (2012). PIPS: pathogenicity island prediction software. *PLoS ONE* 7:e30848. doi: 10.1371/journal.pone.0030848

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K. E. (2008). Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res. Microbiol.* 159, 349–357. doi: 10.1016/j.resmic.2008.04.010

Sullivan, M. B., Krastins, B., Hughes, J. L., Kelly, L., Chase, M., Sarracino, D., et al. (2009). The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ. Microbiol.* 11, 2935–2951. doi: 10.1111/j.1462-2920.2009.02081.x

Sundell, G. N., and Ivarsson, Y. (2014). Interaction analysis through proteomic phage display. *Biomed. Res. Int.* 2014:176172. doi: 10.1155/2014/176172

Suttle, C. A. (2007). Marine viruses–major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750

Tabb, D. L., McDonald, W. H., and Yates, J. R. III. (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* 1, 21–26. doi: 10.1021/pr0 15504q

Tamakoshi, M., Murakami, A., Sugisawa, M., Tsuneizumi, K., Takeda, S., Saheki, T., et al. (2011). Genomic and proteomic characterization of the large Myoviridae bacteriophage varphiTMA of the extreme thermophile *Thermus thermophilus*. *Bacteriophage* 1, 152–164. doi: 10.4161/bact.1.3.16712

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382. doi: 10.1038/nmeth.1315

Theodorescu, D., and Mischak, H. (2007). Mass spectrometry based proteomics in urine biomarker discovery. *World J. Urol.* 25, 435–443. doi: 10.1007/s00345-007-0206-3

Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470–483. doi: 10.1038/nprot.2009.10

Uzoma, I., and Zhu, H. (2013). Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. *Genomics Proteomics Bioinformatics* 11, 18–28. doi: 10.1016/j.gpb.2012.12.005

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* 39:e1869 doi: 10.3791/1869

Van den Bossche, A., Ceyssens, P.-J., De Smet, J., Hendrix, H., Bellon, H., Leimer, N., et al. (2014). Systematic identification of hypothetical bacteriophage proteins targeting key protein complexes of Pseudomonas aeruginosa. *J. Proteome Res.* 13, 4446–4456. doi: 10.1021/pr500796n

Vardi, A., Haramaty, L., Van Mooy, B. A., Fredricks, H. F., Kimmance, S. A., Larsen, A., et al. (2012). Host-virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19327–19332. doi: 10.1073/pnas.1208895109

Vardi, A., Van Mooy, B. A., Fredricks, H. F., Popendorf, K. J., Ossolinski, J. E., Haramaty, L., et al. (2009). Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science* 326, 861–865. doi: 10.1126/science.1177322

Vernikos, G. S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. *Bioinformatics* 22, 2196–2203. doi: 10.1093/bioinformatics/btl369

Veses-Garcia, M., Liu, X., Rigden, D. J., Kenny, J. G., McCarthy, A. J., and Allison, H. E. (2015). Transcriptomic analysis of Shiga-toxigenic bacteriophage carriage reveals a profound regulatory effect on acid resistance in *Escherichia coli*. *Appl. Environ. Microbiol.* 81, 8118–8125. doi: 10.1128/AEM.02034-15

Villegas, A., She, Y. M., Kropinski, A. M., Lingohr, E. J., Mazzocco, A., Ojha, S., et al. (2009). The genome and proteome of a virulent *Escherichia coli* O157:H7 bacteriophage closely resembling Salmonella phage Felix O1. *Virol. J.* 6:41. doi: 10.1186/1743-422X-6-41

Vlahou, A., and Fountoulakis, M. (2005). Proteomic approaches in the search for disease biomarkers. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 814, 11–19. doi: 10.1016/j.jchromb.2004.10.024

Voorhies, A. A., Eisenlord, S. D., Marcus, D. N., Duhaime, M. B., Biddanda, B. A., Cavalcoli, J. D., et al. (2016). Ecological and genetic interactions between cyanobacteria and viruses in a low-oxygen mat community inferred through metagenomics and metatranscriptomics. *Environ. Microbiol.* 18, 358–371. doi: 10.1111/1462-2920.12756

Weinbauer, M. G., and Rassoulzadegan, F. (2004). Are viruses driving microbial diversification and diversity? *Environ. Microbiol.* 6, 1–11. doi: 10.1046/j.1462-2920.2003.00539.x

Weston, A. D., and Hood, L. (2004). Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* 3, 179–196. doi: 10.1021/pr0499693

Wiseman, J. M., Ifa, D. R., Zhu, Y., Kissinger, C. B., Manicke, N. E., Kissinger, P. T., et al. (2008). Desorption electrospray ionization mass spectrometry: imaging drugs and metabolites in tissues. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18120–18125. doi: 10.1073/pnas.0801066105

Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114. doi: 10.1128/MMBR.64.1.69-114.2000

Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15:R46. doi: 10.1186/gb-2014-15-3-r46

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of

metagenomes. *Nat. Methods* 7, 943–944. doi: 10.1038/nmeth1210-943

Yuan, Y., and Gao, M. (2016). Proteomic analysis of a novel bacillus jumbo phage revealing glycoside hydrolase as structural component. *Front. Microbiol.* 7:745. doi: 10.3389/fmicb.2016.00745

Zhang, A., Sun, H., Yan, G., Wang, P., and Wang, X. (2016). Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. *Biomed. Chromatogr.* 30, 7–12. doi: 10.1002/bmc.3453

Zhao, X., Chen, C., Shen, W., Huang, G., Le, S., Lu, S., et al. (2016). Global transcriptomic analysis of interactions between *Pseudomonas aeruginosa* and Bacteriophage PaP3. *Sci. Rep.* 6:19237. doi: 10.1038/srep19237

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Parmar, Gaikwad, Dhakephalkar, Kothari and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership