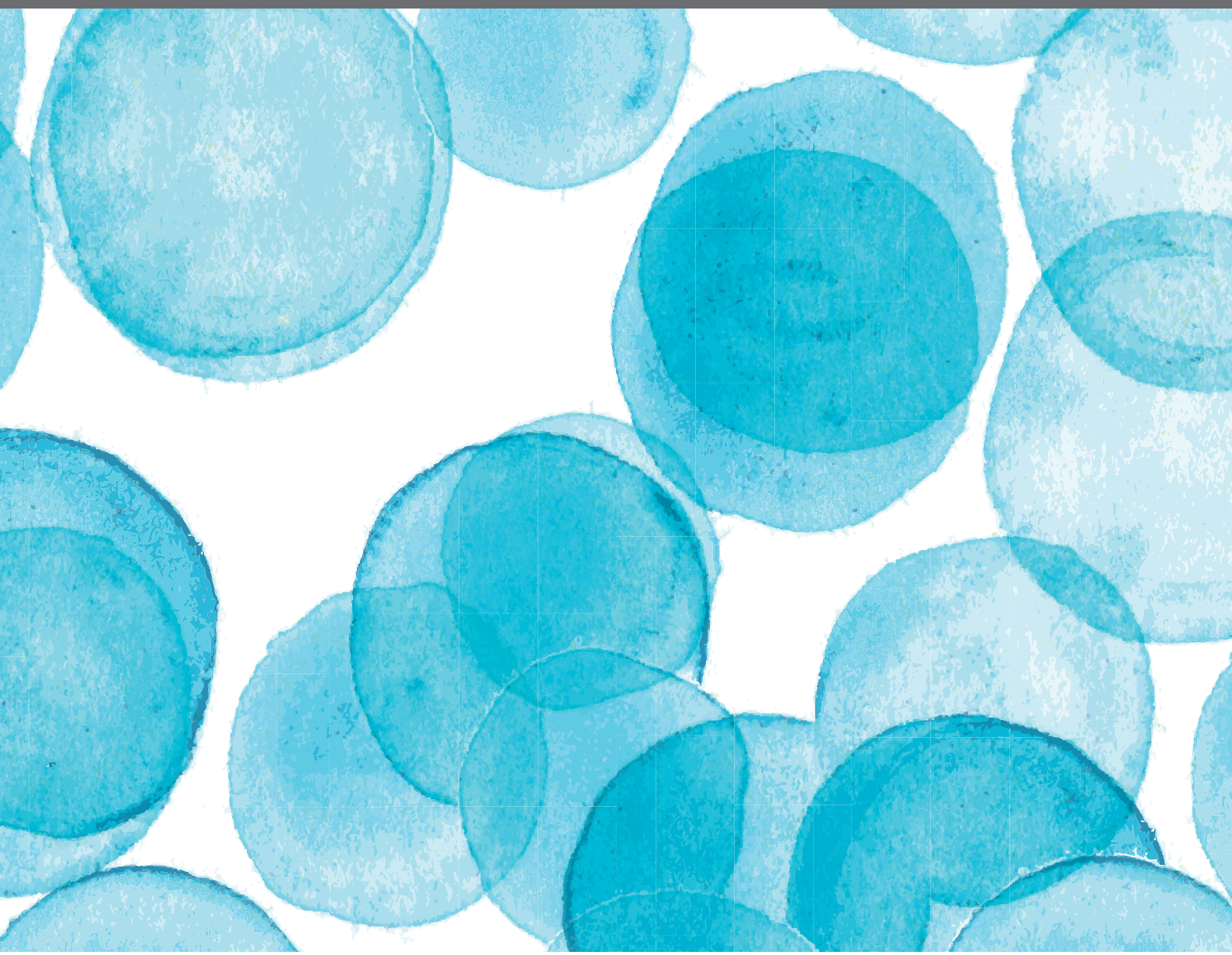
MICROBIAL TAXONOMY, PHYLOGENY AND BIODIVERSITY

EDITED BY: Jesús L. Romalde, Sabela Balboa and Antonio Ventosa
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88963-050-9
DOI 10.3389/978-2-88963-050-9

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

MICROBIAL TAXONOMY, PHYLOGENY AND BIODIVERSITY

Topic Editors:
Jesús L. Romalde, Universidade de Santiago de Compostela, Spain
Sabela Balboa, Universidade de Santiago de Compostela, Spain
Antonio Ventosa, Universidad de Sevilla Spain

The great diversity of microbial life is the remaining major reservoir of unknown biological diversity on earth. To understand this vast, but largely unperceived diversity with its untapped genetic, enzymatic and industrial potential, microbial systematics is undergoing a revolutionary change in its approach to describe novel taxa based on genomic/envirogenomic information.

The characterization of an organism is no longer bounded by methodological barriers, and it is now possible to fully sequence the whole genome of a strain to study individual genes, or to examine the genetic information by using different techniques. In fact, application of genomics is helping not only to provide a better understanding of the boundaries of genera and higher levels of classification, but also to refine our definition of the species concept. In addition, increased understanding of phylogeny is allowing to predict the genetic potential of microorganisms for biotechnological applications and adaptation to environmental changes.

The present Research Topic on "Microbial Taxonomy, Phylogeny and Biodiversity" compiles a collection of papers covering the use of genomic sequence data in microbial taxonomy and systematics, including evolutionary relatedness of microorganisms; application of comparative genomics in systematic studies; or metagenomic approaches for biodiversity studies. We hope that this eBook incentives and encourages researchers for future discussions on microbial taxonomy and phylogenetics.

Citation: Romalde, J. L., Balboa, S., Ventosa, A., eds. (2019). Microbial Taxonomy, Phylogeny and Biodiversity. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88963-050-9

Table of Contents

06 ***Editorial: Microbial Taxonomy, Phylogeny and Biodiversity***
Jesús L. Romalde, Sabela Balboa and Antonio Ventosa

08 ***Genome-Based Characterization of Biological Processes That Differentiate Closely Related Bacteria***
Marike Palmer, Emma T. Steenkamp, Martin P. A. Coetzee, Jochen Blom and Stephanus N. Venter

31 ***GET_PHYLOMARKERS, a Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus* Stenotrophomonas**
Pablo Vinuesa, Luz E. Ochoa-Sánchez and Bruno Contreras-Moreira

53 ***Study of Bacterial Community Composition and Correlation of Environmental Variables in Rambla Salada, a Hypersaline Environment in South-Eastern Spain***
Nahid Oueriaghli, David J. Castro, Inmaculada Llamas, Victoria Béjar and Fernando Martínez-Checa

70 ***Biogeographic Differences in the Microbiome and Pathobiome of the Coral* Cladocora caespitosa *in the Western Mediterranean Sea***
Esther Rubio-Portillo, Diego K. Kersting, Cristina Linares, Alfonso A. Ramos-Esplá and Josefa Antón

81 ***Functional Stability and Community Dynamics During Spring and Autumn Seasons Over 3 Years in Camargue Microbial Mats***
Mercedes Berlanga, Montserrat Palau and Ricardo Guerrero

94 ***Metagenomic Analysis of Cecal Microbiome Identified Microbiota and Functional Capacities Associated With Feed Efficiency in Landrace Finishing Pigs***
Zhen Tan, Ting Yang, Yuan Wang, Kai Xing, Fengxia Zhang, Xitong Zhao, Hong Ao, Shaokang Chen, Jianfeng Liu and Chuduan Wang

105 **Echinococcus granulosus *Infection Results in an Increase in* Eisenbergiella *and* Parabacteroides *Genera in the Gut of Mice***
Jianling Bao, Huajun Zheng, Yuezhu Wang, Xueting Zheng, Li He, Wenjing Qi, Tian Wang, Baoping Guo, Gang Guo, Zhaoxia Zhang, Wenbao Zhang, Jun Li and Donald P. McManus

117 ***Genome-Scale Data Call for a Taxonomic Rearrangement of* Geodermatophilaceae**
Maria del Carmen Montero-Calasanz, Jan P. Meier-Kolthoff, Dao-Feng Zhang, Adnan Yaramis, Manfred Rohde, Tanja Woyke, Nikos C. Kyrpides, Peter Schumann, Wen-Jun Li and Markus Göker

132 ***Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus* Mycobacterium *Into an Emended Genus* Mycobacterium *and Four Novel Genera***
Radhey S. Gupta, Brian Lo and Jeen Son

173 *Corrigendum: Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus* Mycobacterium *Into an Emended Genus* Mycobacterium *and Four Novel Genera*
Radhey S. Gupta, Brian Lo and Jeen Son

175 *Revisiting the Taxonomy of the Genus* Arcobacter*: Getting Order From the Chaos*
Alba Pérez-Cataluña, Nuria Salas-Massó, Ana L. Diéguez, Sabela Balboa, Alberto Lema, Jesús L. Romalde and Maria J. Figueras

194 *Corrigendum: Revisiting the Taxonomy of the Genus Arcobacter: Getting Order From the Caos*
Alba Pérez-Cataluña, Nuria Salas-Massó, Ana L. Diéguez, Sabela Balboa, Alberto Lema, Jesús L. Romalde and Maria José Figueras

196 *Corrigendum (2): Revisiting the Taxonomy of the Genus* Arcobacter*: Getting Order From the Chaos*
Alba Pérez-Cataluña, Nuria Salas-Massó, Ana L. Diéguez, Sabela Balboa, Alberto Lema, Jesús L. Romalde and María J. Figueras

198 *Species Delimitation, Phylogenetic Relationships, and Temporal Divergence Model in the Genus* Aeromonas
J. G. Lorén, Maribel Farfán and M. C. Fusté

211 *Genome Data Provides High Support for Generic Boundaries in* Burkholderia *Sensu Lato*
Chrizelle W. Beukes, Marike Palmer, Puseletso Manyaka, Wai Y. Chan, Juanita R. Avontuur, Elritha van Zyl, Marcel Huntemann, Alicia Clum, Manoj Pillay, Krishnaveni Palaniappan, Neha Varghese, Natalia Mikhailova, Dimitrios Stamatis, T. B. K. Reddy, Chris Daum, Nicole Shapiro, Victor Markowitz, Natalia Ivanova, Nikos Kyrpides, Tanja Woyke, Jochen Blom, William B. Whitman, Stephanus N. Venter and Emma T. Steenkamp

223 *Corrigendum: Genome Data Provides High Support for Generic Boundaries in* Burkholderia *Sensu Lato*
Chrizelle W. Beukes, Marike Palmer, Puseletso Manyaka, Wai Y. Chan, Juanita R. Avontuur, Elritha van Zyl, Marcel Huntemann, Alicia Clum, Manoj Pillay, Krishnaveni Palaniappan, Neha Varghese, Natalia Mikhailova, Dimitrios Stamatis, T. B. K. Reddy, Chris Daum, Nicole Shapiro, Victor Markowitz, Natalia Ivanova, Nikos Kyrpides, Tanja Woyke, Jochen Blom, William B. Whitman, Stephanus N. Venter and Emma T. Steenkamp

226 In-silico *Taxonomic Classification of 373 Genomes Reveals Species Misidentification and New Genospecies Within the Genus* Pseudomonas
Phuong N. Tran, Michael A. Savka and Han Ming Gan

233 *Genotypic and Lipid Analyses of Strains From the Archaeal Genus* Halorubrum *Reveal Insights Into Their Taxonomy, Divergence, and Population Structure*
Rafael R. de la Haba, Paulina Corral, Cristina Sánchez-Porro, Carmen Infante-Domínguez, Andrea M. Makkay, Mohammad A. Amoozegar, Antonio Ventosa and R. Thane Papke

249 *Comparative Genomics of* Thalassobius *Including the Description of* Thalassobius activus *sp. nov., and* Thalassobius autumnalis *sp. nov.*
María J. Pujalte, Teresa Lucena, Lidia Rodrigo-Torres and David R. Arahal

264 ***Clarification of Taxonomic Status Within the* Pseudomonas Syringae *Species Group Based on a Phylogenomic Analysis***
Margarita Gomila, Antonio Busquets, Magdalena Mulet, Elena García-Valdés and Jorge Lalucat

277 ***Defining the Species* Micromonospora saelicesensis *and* Micromonospora noduli *Under the Framework of Genomics***
Raúl Riesco, Lorena Carro, Brenda Román-Ponce, Carlos Prieto, Jochen Blom, Hans-Peter Klenk, Philippe Normand and Martha E. Trujillo

293 ***Discovery of Phloeophagus Beetles as a Source of* Pseudomonas *Strains That Produce Potentially New Bioactive Substances and Description of* Pseudomonas bohemica *sp. nov.***
Zaki Saati-Santamaría, Rubén López-Mondéjar, Alejandro Jiménez-Gómez, Alexandra Díez-Méndez, Tomáš Větrovský, José M. Igual, Encarna Velázquez, Miroslav Kolarik, Raúl Rivas and Paula García-Fraile

308 ***Comparative Genomics and Biosynthetic Potential Analysis of Two Lichen-Isolated* Amycolatopsis *Strains***
Marina Sánchez-Hidalgo, Ignacio González, Cristian Díaz-Muñoz, Germán Martínez and Olga Genilloud

324 ***Phylogeny of* Vibrio vulnificus *From the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy***
Francisco J. Roig, Fernando González-Candelas, Eva Sanjuán, Belén Fouz, Edward J. Feil, Carlos Llorens, Craig Baker-Austin, James D. Oliver, Yael Danin-Poleg, Cynthia J. Gibas, Yechezkel Kashi, Paul A. Gulig, Shatavia S. Morrison and Carmen Amaro

337 ***Corrigendum: Phylogeny of* Vibrio vulnificus *From the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy***
Francisco J. Roig, Fernando González-Candelas, Eva Sanjuán, Belén Fouz, Edward J. Feil, Carlos Llorens, Craig Baker-Austin, James D. Oliver, Yael Danin-Poleg, Cynthia J. Gibas, Yechezkel Kashi, Paul A. Gulig, Shatavia S. Morrison and Carmen Amaro

338 ***Strong Genomic and Phenotypic Heterogeneity in the* Aeromonas sobria *Species Complex***
Jeff Gauthier, Antony T. Vincent, Steve J. Charette and Nicolas Derome

352 ***A Polyphasic and Taxogenomic Evaluation Uncovers* Arcobacter cryaerophilus *as a Species Complex That Embraces Four Genomovars***
Alba Pérez-Cataluña, Luis Collado, Oscar Salgado, Violeta Lefiñanco and María J. Figueras

366 ***Culture Strategies for Isolation of Fastidious* Leptospira *Serovar Hardjo and Molecular Differentiation of Genotypes Hardjobovis and Hardjoprajitno***
Roberta T. Chideroli, Daniela D. Gonçalves, Suelen A. Suphoronski, Alice F. Alfieri, Amauri A. Alfieri, Admilton G. de Oliveira, Julio C. de Freitas and Ulisses de Padua Pereira

frontiers in Microbiology

EDITORIAL
published: 19 June 2019
doi: 10.3389/fmicb.2019.01324

Editorial: Microbial Taxonomy, Phylogeny and Biodiversity

Jesús L. Romalde[1]*, *Sabela Balboa*[1] and *Antonio Ventosa*[2]

[1] Department of Microbiology and Parasitology, CIBUS-Faculty of Biology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain, [2] Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain

Keywords: microbial systematics, taxonomy, phylogeny, diversity, genomics

Editorial on the Research Topic

Microbial Taxonomy, Phylogeny and Biodiversity

The great diversity of microbial life is the remaining major reservoir of unknown biological diversity on Earth. To understand this vast, but largely unperceived diversity with its untapped genetic, enzymatic and industrial potential, microbial systematics is undergoing a revolutionary change in its approach to describe novel taxa based on genomic/envirogenomic information (Rosselló-Móra and Whitman, 2019).

The characterization of an organism is no longer bounded by methodological barriers, and it is now possible to fully sequence the whole genome of a strain to study individual genes, or to examine the genetic information by using different techniques (Mahato et al., 2017). In fact, the application of genomics is helping not only to provide a better understanding of the boundaries of genera and higher levels of classification, but also to refine our definition of the species concept. In addition, increased understanding of phylogeny is allowing to predict the genetic potential of microorganisms for biotechnological applications and adaptation to environmental changes.

The present Research Topic on "Microbial Taxonomy, Phylogeny and Biodiversity" compiles 23 articles covering the use of genomic sequence data in microbial taxonomy and systematics, including evolutionary relatedness of microorganisms; application of comparative genomics in systematic studies; or metagenomic, metatranscriptomic or metaproteomic approaches for biodiversity studies, among others.

Two of the articles introduce novel methodological proposals as a genome-inferred biology or a software to infer pangenomic phylogenies using selected orthologous genes. Thus, Palmer et al. propose to utilize a holistic biological perspective by identifying genome-based characteristics in metabolic networks for taxonomic levels higher than species, in order to link taxonomy and evolution to ecology and appearance of relevant differences during speciation. On the other hand, Vinuesa et al. develope the software package GET_PHYLOMARKERS, a pipeline to identify and select optimal orthologous genes as markers to infer pangenome phylogenies, with the aim of avoid loci with undesirable properties for phylogenetic reconstruction. The genus *Stenotrophomonas* was employed by these authors to validate the protocol identifying also several misclassified RefSeq genome sequences.

A group of 5 manuscripts constitute a section dedicated to the study of the microbial diversity in different environments using metagenomic approaches. Oueriaghli et al. analyze the changes in the community composition associated to physico-chemical variations in a hypersaline environment, demonstrating that salinity and oxygen are key paramters for the dominance of some phyla in the community. Rubio-Portillo et al. study the differences in the microbiome and pathobiome associated to *Cladocora caespitosa*, a coral species abundant in the Mediterranean Sea and heavily affected by the global warming, and Berlanga et al. analyze the functional stability and dynamics

OPEN ACCESS

Edited by:
Haiwei Luo,
The Chinese University of Hong Kong, China

Reviewed by:
Zong-Jun Du,
Shandong University, China
Wen-Jun Li,
Sun Yat-sen University, China

***Correspondence:**
Jesús L. Romalde
jesus.romalde@usc.es

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

Received: *02 April 2019*
Accepted: *28 May 2019*
Published: *19 June 2019*

Citation:
Romalde JL, Balboa S and Ventosa A (2019) Editorial: Microbial Taxonomy, Phylogeny and Biodiversity. Front. Microbiol. 10:1324. doi: 10.3389/fmicb.2019.01324

of bacterial populations in microbial mats from the Camargue wetlands of Rhone delta (France). Two of the articles investigate the gut microbiome of pigs and mice to determine its role in the feed efficiency or the effects of chronic infectious, respectively (Tan et al.; Bao et al.).

Seven articles are devoted to the study of taxonomic relationships at genus level in different groups of microorganisms, including Gram-positive and Gram-negative bacteria or archaea. The need of a taxonomic rearrangement of the family *Geodermatophilaceae* is indicated by Montero-Calasanz et al. who describe a novel genus, *Klenkia*, together with the reassignation and emendation of numerous species of this group. Gupta et al. propose the division of the genus *Mycobacterium* in five different genera on the basis of phylogenomic and comparative genomic studies and, similarly, Pérez-Cataluña et al. suggest the division of the genus *Arcobacter* in seven different taxa. On the other hand, Lorén et al. use the generalized mixed Yule coalescent (GMYC) method to determine the species delineation, the phylogenetic relationships as well as to stablish the temporal divergence model in the genus *Aeromonas*, while Beukes et al. evaluate the generic boundaries of the genus *Burkholderia* and related taxa clarifying their evolutionary relationships. The description of new genospecies within the genus *Pseudomonas* through the analysis of more than 370 genomes is performed by Tran et al. who also indicate the misidentification of a large number of genomes mainly from *P. syringae* or *P. fluorescens*. Within archaea, de la Haba et al. perform genomic and polar lipid profiles analyses for the genus *Halorubrum*, being able to establish cutoff values for species delineation using a MLSA approach.

Pujalte et al. employ a polyphasic approach not only to describe two new species within the genus *Thalassobius* but also to carry out a taxogenomic study of the genera *Thalassobius*, *Shimia* and *Thalassococcus*, pointing out the difficulties for classification at the genus level within the *Roseobacter* group. The taxonomic status within the *Pseudomonas syringae* species group is analyzed by Gomila et al. demonstrating not only the misclasification of a high proportion of the strains studied, but also the possibility of 7 new species within this bacterial group. In a study of two species of *Micromonospora*, Riesco et al. define a working framework for defining species within this genus both at genomic and phenotypic levels. On the other hand, Saati-Santamaría et al. describe the new bacterial species *Pseudomonas bohemica* sp. nov. from bark beetles, indicating also its biotechnological potential as producer of bioactive substances with antimicrobial activity, whereas Sánchez-Hidalgo et al. determine the biosynthetic potential of *Amycolatopsis* strains by means of comparative genomics.

A group of articles are focused in the study of the intra-species taxonomy of different bacteria, including *Vibrio vulnificus* (Roig et al.), *Aeromonas sobria* (Gauthier et al.) or *Arcobacter cryaerophilus* (Pérez-Cataluña et al.). In these articles, updated intra-species classifications are proposed including well defined phylogenetic lineages, subclades, or genomovars most of them with epidemiological significance. Finally, the aim of the only article not using a genomic approach is to develop culture strategies for isolation of fastidious *Leptosira* serovar Hardjo and the use of molecular methods for differentiation of its two genotypes, which can be helpful for seroepidemiology and immunoprofilaxis of leptospirosis (Chideroli et al.).

In summary, the collection of papers published in this Research Topic provides an updated picture on the utility of the genomic approach on microbial systematics and biodiversity studies, presents new tools for the *in silico* analysis of phylogenetic relationships, and contributes scientific basis for future standardization of taxa descriptions. If in addition serves as incentive and encouragement to researchers and future discussions on microbial taxonomy and phylogenetics, it would fulfill more than enough its original aims.

AUTHOR CONTRIBUTIONS

JR led writing of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

Mahato, N. K., Gupta, V., Singh, P., Kumari, R., Verma, H., Tripathi, C., et al. (2017). Microbial taxonomy in the era of OMICS: application of DNA sequences, computational tools and techniques. *Antoine van Leeuwenhoek* 110, 1357–1371. doi: 10.1007/s10482-017-0928-1

Rosselló-Móra, R., and Whitman, W. B. (2019). Dialogue on the nomenclatura and clasification of prokaryotes. *Syst. Appl. Microbiol.* 42, 5–14. doi: 10.1016/j.syapm.2018.07.002

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 Romalde, Balboa and Ventosa. 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: 06 February 2018
doi: 10.3389/fmicb.2018.00113

Genome-Based Characterization of Biological Processes That Differentiate Closely Related Bacteria

*Marike Palmer[1], Emma T. Steenkamp[1], Martin P. A. Coetzee[2], Jochen Blom[3] and Stephanus N. Venter[1]**

[1] Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa, [2] Department of Genetic, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa, [3] Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany

Bacteriologists have strived toward attaining a natural classification system based on evolutionary relationships for nearly 100 years. In the early twentieth century it was accepted that a phylogeny-based system would be the most appropriate, but in the absence of molecular data, this approach proved exceedingly difficult. Subsequent technical advances and the increasing availability of genome sequencing have allowed for the generation of robust phylogenies at all taxonomic levels. In this study, we explored the possibility of linking biological characters to higher-level taxonomic groups in bacteria by making use of whole genome sequence information. For this purpose, we specifically targeted the genus *Pantoea* and its four main lineages. The shared gene sets were determined for *Pantoea,* the four lineages within the genus, as well as its sister-genus *Tatumella*. This was followed by functional characterization of the gene sets using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. In comparison to *Tatumella*, various traits involved in nutrient cycling were identified within *Pantoea*, providing evidence for increased efficacy in recycling of metabolites within the genus. Additionally, a number of traits associated with pathogenicity were identified within species often associated with opportunistic infections, with some support for adaptation toward overcoming host defenses. Some traits were also only conserved within specific lineages, potentially acquired in an ancestor to the lineage and subsequently maintained. It was also observed that the species isolated from the most diverse sources were generally the most versatile in their carbon metabolism. By investigating evolution, based on the more variable genomic regions, it may be possible to detect biologically relevant differences associated with the course of evolution and speciation.

Keywords: genome-inferred biology, *Enterobacteriaceae*, phenotype, KEGG, bacterial systematics, *Pantoea*

OPEN ACCESS

Edited by:
Antonio Ventosa, Universidad de Sevilla, Spain

Reviewed by:
Fabiano Thompson, Universidade Federal do Rio de Janeiro, Brazil
Alice Rebecca Wattam, Virginia Tech, United States

***Correspondence:**
Stephanus N. Venter
fanus.venter@up.ac.za

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

Received: *16 August 2017*
Accepted: *17 January 2018*
Published: *06 February 2018*

Citation:
Palmer M, Steenkamp ET, Coetzee MPA, Blom J and Venter SN (2018) Genome-Based Characterization of Biological Processes That Differentiate Closely Related Bacteria. Front. Microbiol. 9:113. doi: 10.3389/fmicb.2018.00113

INTRODUCTION

Since the early twentieth century scientists have recognized the value of phylogenetic inferences in determining natural relationships between taxa, which is essential for both taxonomy and evolutionary studies (Woese, 1987). However, the move toward a more natural classification system by these early bacteriologists, based on phylogenetics, proved exceedingly difficult as traditionally used morphological traits were not variable enough to group taxa reliably (Stanier and Van Niel, 1941; Woese, 1987; Woese et al., 1990). Although this led to the use of physiological

characters, some researchers already argued early on that such data would not be suitable for developing evolutionary hypotheses. They emphasized that physiological traits would generally not be phylogenetically informative as long as there were no clear understanding of their genetic basis and overall biological importance (Stanier and Van Niel, 1941). The consensus view at the time was thus that phylogenetic inferences were definitely needed for elucidating the natural relationships among bacteria, but that this would only be possible with the use of suitably informative characters (Stanier and Van Niel, 1941; Woese, 1987, 1994; Woese et al., 1990). As a result, scholars mostly abandoned the field of bacterial systematics until more reliable characters became available with the advent of nucleic acid-based molecular phylogenetics in the 1970s (Woese, 1987, 1994, 1998; Woese et al., 1990; McInerney et al., 2011).

For studying bacterial systematics, the ubiquitous 16S ribosomal RNA (16S rRNA) gene was initially the marker of choice (Hillis and Dixon, 1991; Woese, 1994, 1998; Garrity et al., 2005; Gevers et al., 2005; Konstantinidis and Tiedje, 2007). Over time, however, as the diversity of examined samples increased, it became apparent that the 16 rRNA gene sequence alone does not provide sufficient phylogenetic resolution. Therefore, more reliable approaches for phylogenetic inference were sought to obtain better resolved trees. This led to the use of multiple locus sequence analyses (MLSA) (Gevers et al., 2005; Konstantinidis and Tiedje, 2007; Glaeser and Kämpfer, 2015), ribosomal MLSA (Bennett et al., 2012) and more recently core genome phylogenies (Bennett et al., 2012; Chan et al., 2012; Rahman et al., 2015; Schwartz et al., 2015; Palmer et al., 2017). These approaches, especially those based on large numbers of core genes, provide robust evolutionary hypotheses that seems to be resilient to most known phylogenetic errors (Beukes et al., 2017; Palmer et al., 2017) and have recently formed the foundation of taxonomic decisions, particularly in problematic taxa (Zhang et al., 2011; Bennett et al., 2012; Chan et al., 2012; Richards et al., 2014; Ormeno-Orrillo et al., 2015; Rahman et al., 2015).

The next logical step after having used phylogenetics to identify taxa, particularly those above the species level, would be to assign biological characters to them. For example, if bacterial genera or the lineages within them represent natural clusters, it should be possible to identify properties that they share with one another, but that are different from those occurring in other such clusters. Previously, various standardized sets of physiological tests have been used to study phenotypic cohesion of bacterial taxa (Schubert, 1968; Gavini et al., 1989; Mergaert et al., 1993; Brady et al., 2010a, 2013), but these have been mainly developed from a clinical diagnostics perspective (Konstantinidis and Tiedje, 2005a,b; Sutcliffe, 2015). Accordingly, the characters identified by these tests have limited application outside this environment (Sutcliffe et al., 2012; Sutcliffe, 2015). Other than revealing some basic physiological capabilities, these standard phenotypic tests are incapable of capturing the countless traits encoded on bacterial genomes. In addition, the very limited set of traits analyzed rarely differentiates between taxa, as the members of a taxon can show immense physiological variability. Therefore, characters that are biologically meaningful and that potentially define or distinguish higher-level bacterial groups and taxa would thus have to be sought through other means.

For identifying biological traits that are potentially taxon-defining, whole genome sequences represent invaluable resources. A wealth of traits can be inferred from a bacterium's genome sequence by making use of bioinformatics approaches and databases built from experimental evidence. For example, metabolic and physiological networks or pathways can be inferred from gene sequences by making use of their homology to sequences in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016b). Each sequence in the KEGG database have an associated KEGG Orthology (KO) term, which is in turn coupled to proteins whose functions have been experimentally verified (Kanehisa et al., 2016b). In this workflow for inferring physiological properties, the database of functionally verified protein entries is often regarded as a significant limitation. This is because functional characterization of genes occurs at a much slower pace than gene discovery, thus making it impossible to functionally annotate certain genes (Linghu et al., 2008). As a result, taxa related to the model organisms typically have a higher number genes that can be functionally annotated because of the higher similarity between their genomes (Linghu et al., 2008). Despite this limitation, the current information would still provide valuable biological knowledge, especially as the information in these databases increase.

In this study, we explored the possibility of linking biological characters to higher-level taxonomic groups in bacteria by making use of whole genome sequence information. For this purpose we used the genus *Pantoea* for which genome sequences of 21 species were available (Hong et al., 2012; Kamber et al., 2012; Wan et al., 2015; Palmer et al., 2017). These were chosen to span the diversity of the genus, which includes plant pathogens (*P. agglomerans, P. ananatis, P. stewartii* amongst others; Coutinho and Venter, 2009) and species that affect humans (*P. brenneri, P. conspicua, P. eucrina,* and *P. septica;* Walterson and Stavrinides, 2015), as well as species that have been isolated from insects, fungal fruiting bodies and environmental samples (Walterson and Stavrinides, 2015; Ma et al., 2016; Palmer et al., 2016; Rong et al., 2016). Overall the members of this genus appear to be highly adaptable to changing environments and may act opportunistically when in contact with potential eukaryotic hosts (Coutinho and Venter, 2009; De Maayer et al., 2012b, 2014; Walterson and Stavrinides, 2015). From a phylogenetic perspective, *Pantoea* and its sister taxon, *Tatumella*, are nested within the *Enterobacteriaceae* where they are closely related to *Erwinia* (Glaeser and Kämpfer, 2015; Palmer et al., 2017). *Pantoea* further separates into four well-supported lineages, *viz.* the *P. agglomerans* (containing *P. agglomerans*, *P. anthophila*, *P. brenneri*, *P. conspicua*, *P. deleyi*, *P. eucalypti,* and *P. vagans*), *P. ananatis* (containing *P. allii*, *P. ananatis*, *P. stewartii* ssp. *stewartii* and *P. stewartii* ssp. *indologenes*), *P. rodasii* (containing *P. rodasii*, *P. rwandensis,* and *Pantoea* sp. GM01) and *P. dispersa* (containing *P. dispersa*, *P. eucrina* and *P. wallisii*) lineages (Palmer et al., 2017). Other than a limited set of general biological traits (e.g., colony and cell morphology, respiration

status, growth temperature), characters that potentially define *Pantoea* and its lineages have never been identified.

The overall goal of this study was to link biological properties to the current evolutionary hypothesis of *Pantoea* (Palmer et al., 2017), thus allowing the identification of phenotypic characters that potentially define the genus and its lineages. Our specific aims were three-fold. First, for each of *Pantoea* and *Tatumella*, we functionally compared their shared gene sets (i.e., in terms of the pathways and processes each gene is predicted to be involved in) to evaluate the feasibility of using whole genome sequences for identifying taxon-defining characters at ranks higher than the species level. Secondly, the shared gene sets in each of the four *Pantoea* lineages were functionally compared to identify characters associated with the specific evolutionary path of the lineage and that potentially contributed to its initial emergence or subsequent maintenance. Thirdly, the loci underlying these differential characters were further characterized in order to determine their gene composition and distribution among species and whether their conservation is maintained by purifying selection as suggested before (Fang et al., 2008; Sorrels et al., 2009). Broadly, our strategy (**Figure 1**) involved the identification of shared gene sets, followed by their functional annotation.

MATERIALS AND METHODS

Genomes Analyzed

All genomes analyzed during this study are publicly available and accessible at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). Whole genome sequence data for 21 species of *Pantoea* and three species of *Tatumella* (Tracz et al., 2015) were included in the analyses (**Table 1**). These *Pantoea* species span the current known phylogenetic and phenotypic diversity of the genus, with most representatives of all of the major lineages (Palmer et al., 2017). For the inter-generic comparisons, all 24 genomes were utilized. For the intra-generic comparisons, we included only 17 *Pantoea* genomes. We excluded three of the known lineages of this genus as one contained only two species (i.e., *Pantoea* sp. At-9b and *P. cypripedii* LMG 2655 ^{T}), while the other two are each represented by single species (i.e., *Pantoea* sp. A4 and *P. septica* LMG 5345 ^{T}).

Generation of Shared Gene Sets

The shared gene sets of the two genera, as well as the different lineages within *Pantoea* (see **Figure 1**), were generated with the EDGAR (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios) server (https://edgar.computational.bio.uni-giessen.de; Blom et al., 2016). For each gene set, a representative of the lineage/genus was used for downstream analyses. The representatives used for the different lineages were *P. agglomerans* R190 for the first lineage (encompassing *P. agglomerans*, *P. eucalypti*, *P. vagans*, *P. deleyi*, *P. anthophila*, *P. brenneri,* and *P. conspicua*), *P. ananatis* LMG 2665^{T} for the second lineage (comprising of *P. ananatis*, *P. allii*, *P. stewartii* subsp. *stewartii,* and *P. stewartii* subsp. *indologenes*), *P. dispersa* strain number to EGD-AAK13 for the third lineage (encompassing *P. dispersa*, *P. eucrina,* and *P. wallisii*) and *P. rodasii* LMG 26273^{T} for the fourth lineage

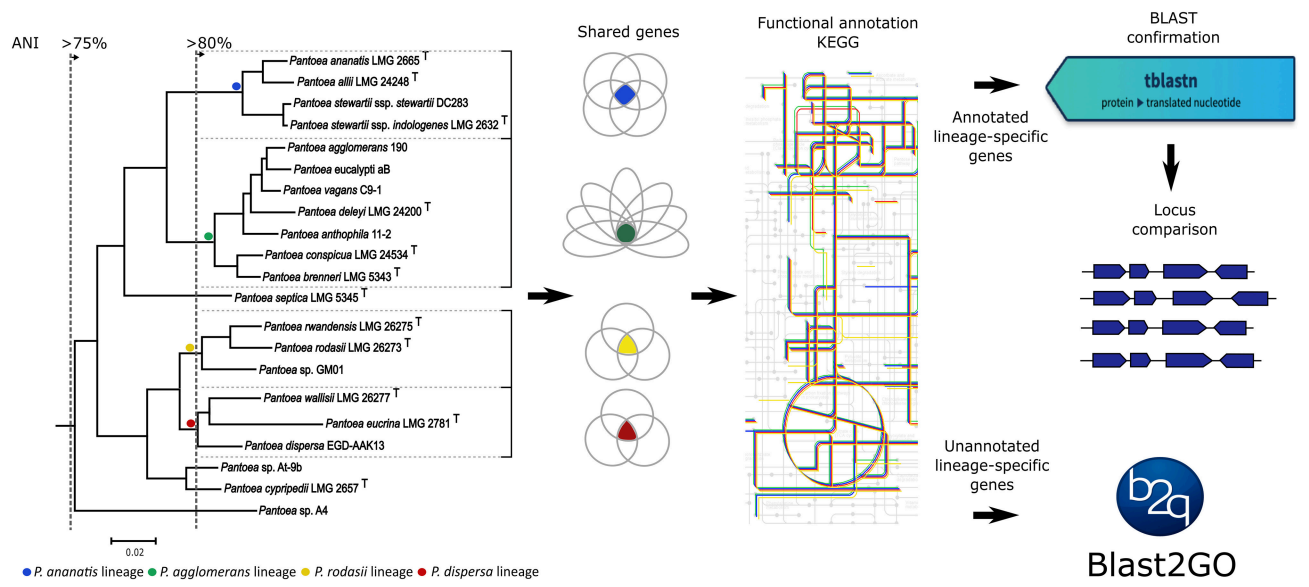


FIGURE 1 | Experimental strategy followed for the lineages within *Pantoea* in the study. Lineages were identified from the subtree of *Pantoea* from the protein sequence maximum-likelihood tree of all the shared genes of Palmer et al. (2017). Average nucleotide identity (ANI) values were used as a measure of relatedness between species of a lineage as obtained from Palmer et al. (2017). Shared gene sets were determined from the genome sequences of species within each lineage. Gene sets were then annotated with the Kyoto Encyclopedia of Genes and Genomes (KEGG), followed by BLAST verification and locus comparisons of characterized genes. Uncharacterized genes were subjected to Blast2GO analyses. A similar strategy was followed for the generic comparisons with the exception of the locus comparisons.

TABLE 1 | Genomes analyzed in the study.

Species	Strain	Origin	Accession number
Pantoea agglomerans	R 190	Apple, Korea	JNGC00000000.1
Pantoea allii	LMG 24248^T	Onion seed, South Africa	MLFE00000000.1
Pantoea ananatis	LMG 2665^T	Pineapple, Brazil	JFZU00000000.1
Pantoea anthophila	11-2	Hypersaline lake, Hawaii	JXXL00000000.1
Pantoea brenneri	LMG 5343^T	Human, USA	MIEI00000000.1
Pantoea calida	LMG 25383^T	Infant formula, -	MLFO00000000.1
Pantoea conspicua	LMG 24534^T	Human, France	MLFN00000000.1
Pantoea cypripedii	LMG 2657^T	Orchid, USA	MLJI00000000.1
Pantoea deleyi	LMG 24200^T	Eucalyptus, Uganda	MIPO00000000.1
Pantoea dispersa	EGD-AAK13	Soil, India	AVSS00000000.1
Pantoea eucalypti	aB	Bark beetle, USA	AEDL00000000.1
Pantoea eucrina	LMG 2781^T	Human, USA	MIPP00000000.1
Pantoea gaviniae	LMG 25382^T	Infant formula, -	MLFQ00000000.1
Pantoea rodasii	LMG 26273^T	Eucalyptus, Colombia	MLFP00000000.1
Pantoea rwandensis	LMG 26275^T	Eucalyptus, Rwanda	MLFR00000000.1
Pantoea septica	LMG 5345^T	Human, USA	MLJJ00000000.1
Pantoea stewartii ssp. *stewartii*	DC 283	Maize, USA	AHIE00000000.1
Pantoea stewartii ssp. *indologenes*	LMG 2632^T	Fox millet, India	JPKO00000000.1
Pantoea vagans	C9-1	Apple, USA	CP001894.1, CP001893.1, CP001894.1
Pantoea wallisii	LMG 26277^T	Eucalyptus, South Africa	MLFS00000000.1
Pantoea sp.	At-9b	Leaf cutter ant, USA	CP002433.1, CP002434.1, CP002435.1, CP002436.1, CP002437.1, CP002438.1
Pantoea sp.	A4	Rafflesia flower, Malaysia	ALXE00000000.1
Pantoea sp.	GM01	Poplar, USA	AKUI00000000.1
Tatumella morbirosei	LMG 23360^T	Pineapple, Philippines	CM003276.1
Tatumella ptyseos	ATCC 33301^T	Human, USA	ATMJ00000000.1
Tatumella saanichensis	NML 06-3099^T	Human, Canada	ATMI00000000.1

(consisting of *P. rodasii*, *P. rwandensis,* and *Pantoea* sp. GM01). For the intergeneric comparisons, *P. agglomerans* R190 was again used as representative of *Pantoea* and *T. ptyseos* ATCC 33301^T as representative of *Tatumella*.

Functional Annotation and Identification of Differentially Present Metabolic Pathways

Functional annotation of the different gene sets were first performed by orthology searches against the KEGG database (Kanehisa et al., 2016b) using GhostKOALA (KEGG Orthology and Links Annotation; Kanehisa et al., 2016a) for all gene sets. Genes with KO terms associated with them could be separated based on the functional role of the pathways to which they could be mapped. Specific pathways where differences were detected in the global maps were also considered for comparative purposes (**Figure 1**).

For the *Pantoea* lineages, genes with no KO associations were then analyzed to assign putative functions using Blast2GO (Conesa et al., 2005; Götz et al., 2008). This was done by subjecting these genes to BLAST analyses for Gene Ontology (GO) associations using Blast2GO implemented in CLC Genomics Workbench (CLC Bio). In these analyses, assignment to more than one GO term per gene was allowed when functional annotation suggested that a gene product is involved in multiple processes. All Blast2GO analyses were initiated by BLAST searches against the RefSeq non-redundant protein database of NCBI followed by InterproScan (Jones et al., 2014) analyses to identify protein domains as a means of identifying putative functions. Genes remaining without annotation was again subjected to BLAST analyses against the non-redundant database on NCBI to determine the distribution of these genes across taxa.

Individual sets of reconstructed metabolic pathways obtained from the KEGG database were compared to identify differences between lineages and genera. This was done by assigning the set of KEGG pathway maps from each genus/lineage a unique color and then overlaying them onto each other for identifying differences (**Figure 1**). From these overview pathway maps, specific metabolic pathways were identified for further investigation in eight functional categories used in KEGG. These were carbohydrate, lipid, nucleotide, amino acid and energy metabolism, as well as genes involved in environmental information processing and the metabolism of co-factors, vitamins, and xenobiotics.

Multi-gene pathways that were differentially present or absent were identified from the full set of differences obtained from the KEGG pathway comparisons. This was done to limit the number of genes potentially identified as absent due to sequencing or

assembly errors and also aided in simplifying the overall analysis. For this purpose multi-gene pathways were defined as processes where more than one gene was required to complete a process. From these pathways, the absence of genes from the genomes included in the respective gene sets were verified using local BLAST (Altschul et al., 1990) analyses (tblastn). The genomic coordinates of these genes were then noted to identify clustered genes. The gene clusters were identified and visualized using Geneious 6.1.6 (Biomatters).

Sequences for complete clusters were subsequently extracted from genomes and aligned using the MAFFT 7.309 (Katoh and Standley, 2013) server. When more than three members of a lineage possessed the gene clusters, their sequence alignments were subjected to codon-based selection analyses in MEGA 6.0.6 (Tamura et al., 2013) using HyPhy (Pond and Muse, 2005), to obtain dN (proportion of non-synonymous substitutions) and dS (proportion of synonymous substitutions) values at all codon positions across the alignments. The normalized dN-dS values were then plotted against codon positions in Microsoft Excel 2013.

RESULTS

Generation of Shared Gene Sets

In this study, we identified a number of biologically informative characters for *Pantoea* and the four lineages examined. For analyses at both the inter- and intra-generic levels, comparable taxon sets were compiled based on phylogenetic relatedness and Average Nucleotide Identity values (ANI Konstantinidis and Tiedje, 2005b; see Palmer et al., 2017). At the intra-generic level, these sets were also comparable in the sizes of the gene sets (**Figure 2**), with the exception of the *P. rodasii* lineage. This larger gene set could be attributed to the large size of the genomes of the three current members in the lineage, in comparison to the members of other lineages. However, a large proportion of the genes (>25%) in the respective genomes were present in both *Pantoea* and *Tatumella*. For the *Pantoea* comparisons more than 30% of the genes in respective genomes were present in all taxa, with 55–75% associated with specific lineages and 25–45% apparently species-specific. Overall, the gene sets for the four lineages consisted of 2844 genes for the *P. agglomerans* lineage, 2924 genes for the *P. ananatis* lineage, 3599 genes for the *P. rodasii* lineage and 2872 genes for the *P. dispersa* lineage (**Figure 3**).

For the inter-generic comparisons, the *Pantoea* gene set (calculated from 21 genomes) consisted of 1862 genes. The *Tatumella* gene set consisted of 2196 genes (calculated from three genomes). This difference in the number of shared genes can most likely be attributed to the number of genomes analyzed in these genera, as the number of genomes available for *Tatumella* is underrepresented.

Functional Annotation of the *Pantoea* and *Tatumella* Gene Sets with KEGG

The number of genes with KO associations for *Pantoea* and *Tatumella* were 1,576 (84.6% of the *Pantoea* gene set) and 1,760 (80.1% of the *Tatumella* gene set), respectively (**Supplementary File S1**). In both cases, the highest number of genes was involved in "Genetic Information Processing", followed by "Environmental Information Processing", with "Unclassified" genes making up the third largest gene group.

The pathways in which we identified differences between *Pantoea* and *Tatumella* were "Metabolic pathways", "Biosynthesis of secondary metabolites", "Microbial metabolism in diverse environments", "Biosynthesis of antibiotics", "Carbon metabolism", "Biosynthesis of amino acids", as well as "2-Oxocarboxylic acid metabolism", and "Fatty acid metabolism". Comparison of the relevant global metabolic maps revealed a higher number of reactions predicted for *Tatumella* (as would be expected due to the higher number of shared genes), except for "Fatty acid metabolism". Closer inspection of the fatty acid metabolism pathways indicated the ability to perform β-oxidation of fatty acids occurred in *Pantoea* but not *Tatumella*.

A total of 124 differences were identified between *Pantoea* and *Tatumella* (**Supplementary File S1**). These consisted of reactions involved in all functional classes, namely "Carbohydrate metabolism" (citrate cycle, pentose phosphate pathway, fructose and mannose metabolism, ascorbate and aldarate metabolism, starch and sucrose metabolism, glyoxylate and dicarboxylate metabolism and inositol phosphate metabolism), "Energy metabolism" (including methane, sulfur and nitrogen metabolism), "Lipid metabolism" (including fatty acid degradation and sphingolipid metabolism), "Nucleotide metabolism" (purine and pyrimidine metabolism), "Amino acid metabolism" (cysteine and methionine metabolism, lysine degradation, arginine and proline metabolism, histidine metabolism and β-alanine metabolism), "Cofactor metabolism" (nicotinate and nicotinamide metabolism), "Xenobiotics metabolism" (benzoate degradation, chloroalkane and chloroalkene degradation) and "Environmental information processing" [ABC transporters, two-component systems, phosphotransferase systems (PTSs) and chemotaxis]. By limiting the pathways investigated to those where two or more genes are required to complete a pathway, 10 pathways (32 differences) were retained and subsequently absence was confirmed with BLAST analyses (**Table 2**, **Supplementary File S1**).

As suggested from the global maps, genes required for β-oxidation of long-chain fatty acids ("Lipid metabolism"—fatty acid degradation) were present in all members of *Pantoea* and absent in all members of *Tatumella*. A number of genes involved in specific pathways in carbohydrate metabolism where detected only in *Pantoea*, namely the "Pentose phosphate pathway" (D-ribose to D-ribose-1-P), "Fructose and mannose metabolism" (D-mannitol to β-D-fructose-6-P), "Starch and sucrose metabolism" (sucrose to ADP-glucose; glycogen to trehalose and amylose, respectively) and "Inositol phosphate metabolism" (myo-inositol to 2-deoxy-5-keto-D-gluconate-6-P). The "Energy metabolism" pathway with differences was sulfur metabolism, where *Pantoea* possessed genes required for the uptake of extracellular taurine and its subsequent conversion to sulfite. *Pantoea* also possessed genes required for the conversion of guanine to (S)-allantoin during "Purine metabolism". Several pathways involved in "Amino acid metabolism" were also present only in *Pantoea*, specifically those involved in arginine and proline metabolism

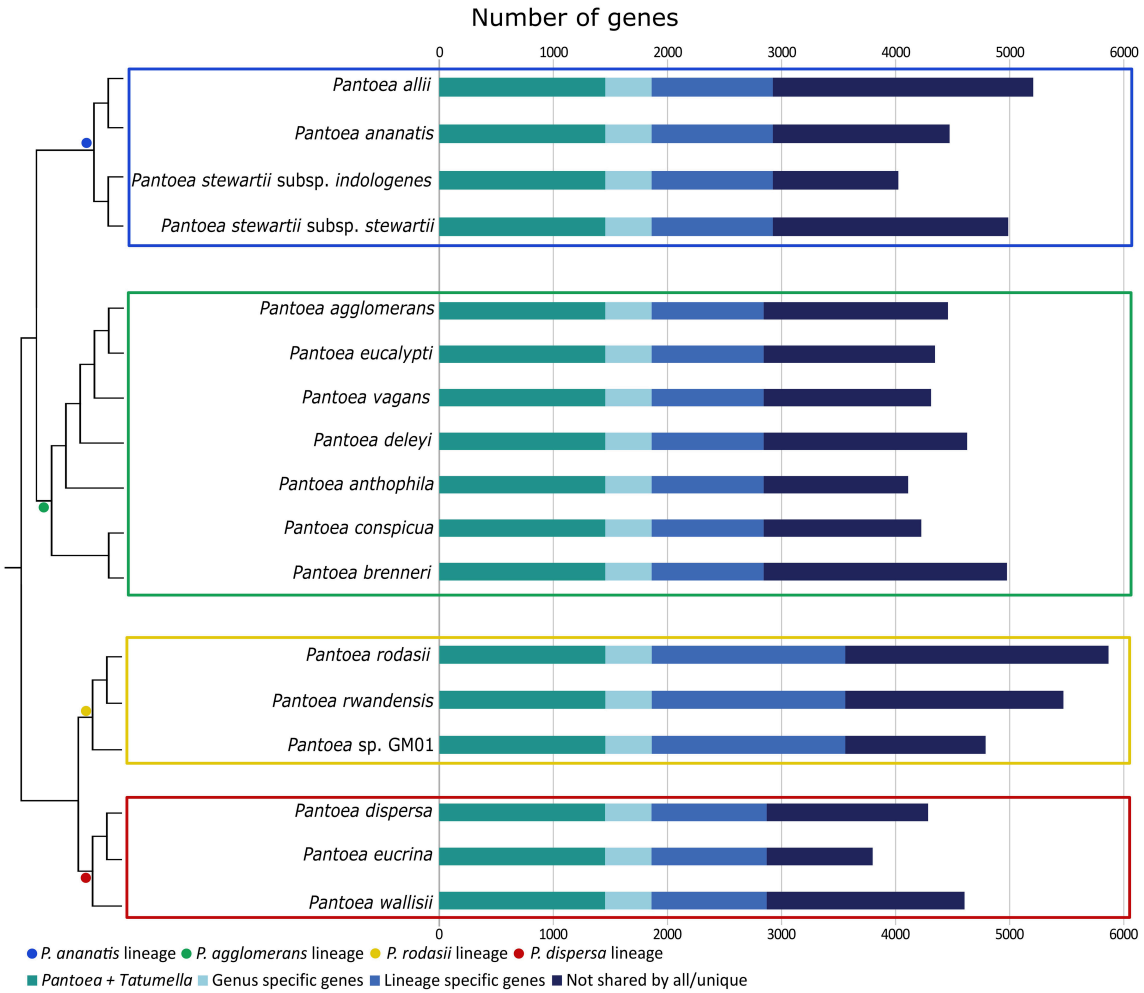


FIGURE 2 | A bar graph indicating the number of genes for each isolate, separated into genes highly conserved in the sister genera (*Pantoea* + *Tatumella*), through to genes not shared by all closely related species or unique genes. The dendrogram was inferred for the different lineages and their relationships to each other from the amino acid based topology of the core genome of Palmer et al. (2017). The length of each bar is indicative of the size of each genome analyzed (in terms of the number of genes). The different lineages are indicated with colored blocks (blue—*P. ananatis* lineage; green—*P. agglomerans* lineage; yellow—*P. rodasii* lineage; red—*P. dispersa* lineage). All genomes analyzed encoded a similar number of genes, with the genome of *P. rodasii* encoding the highest number of genes (~5,800) and *P. eucrina* encoding the least number of genes (~3,800).

(creatine and N-carbamoylsarcosine to sarcosine) and histidine metabolism (L-histidinol to urocanate). In addition, *Pantoea* also possessed genes for the conversion of L-aspartate to nicotinate-D-ribonucleotide during "Cofactor metabolism" (nicotinate and nicotinamide metabolism).

We also observed differences for genes in the category "Environmental Information Processing". Amongst the ABC transporters, those for glutamine, glutathione and glycine betaine/proline transporters were found only in *Tatumella,* while those for osmoprotectant, taurine (also seen in sulfur metabolism), L-arabinose, and microcin C were only present in *Pantoea*. In terms of the phosphotransferase systems, *Pantoea* possessed genes necessary to transport and convert N-acetylmuramic acid to N-acetylmuramic acid-6-P, while *Tatumella* possessed genes required for the transport and conversion of N-acetyl-D-glucosamine to N-acetyl-D-glucosamine-6-P and arbutin/salicin to arbutin-6-P/salicin-6-P.

Functional Annotation of the Genes Shared by Lineages in *Pantoea* with KEGG

For the *P. agglomerans* lineage, the number of shared genes with KO associations resulted in 2148 genes (75.5% of the gene set). A total of 2197 genes (75.1% of the gene set) in the *P. ananatis* lineage could be annotated using KO terms (**Supplementary File S2**). The highest percentage of genes with associated KO terms were 75.9% (2181 genes) for the *P. dispersa* lineage, with the lowest being 71.9% for the *P. rodasii* lineage (2559 genes) (**Supplementary File S2**). In contrast to the inter-generic gene sets, the highest number of genes in all four lineages were involved in "Environmental Information Processing", followed by "Genetic Information Processing",

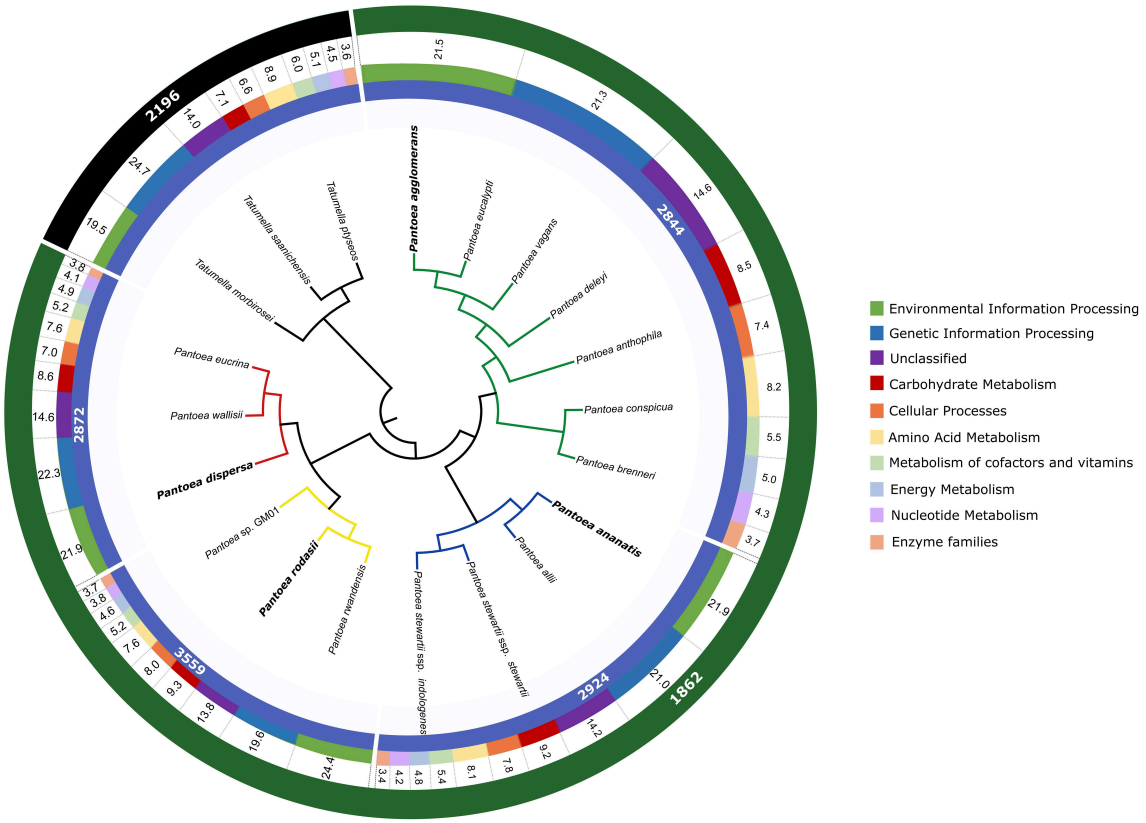


FIGURE 3 | An overview of the gene sets shared between *Pantoea* and *Tatumella* and among the various lineages of *Pantoea*. For perspective, these are indicated relative to the genome-based species tree for *Pantoea* generated by Palmer et al. (2017). The lineages examined in this study are indicated with colored branches (green—*P. agglomerans* lineage; blue—*P. ananatis* lineage; yellow—*P. rodasii* lineage; red—*P. dispersa* lineage) with the representative of each lineage indicated in bold. The inner track indicates the sizes of the gene sets (in number of genes) of the lineages within *Pantoea*. The second track indicates the proportion of the genes annotated with the KEGG database involved in each of the functional classes (see legend). Values in the third track depicts the percentage of annotated genes involved in each functional class. The outer track indicates the size of the shared gene sets for the genera *Pantoea* and *Tatumella*, respectively.

with "Unclassified" genes again being the third most prevalent (**Figure 3**, **Supplementary File S2**).

Comparisons of global maps indicated differences between the four lineages in "Biosynthesis of amino acids", "Biosynthesis of antibiotics", "Biosynthesis of secondary metabolites", "Carbon metabolism", "Overview metabolism" and "Microbial metabolism in diverse environments" (**Supplementary File S2**). Limiting the reactions investigated to two or more genes acting together to complete a pathway, led to the identification of a number of reactions involved in "Polyketide sugar unit biosynthesis", "Biosynthesis of siderophore group non-ribosomal peptides", "Starch and sucrose metabolism", "Riboflavin metabolism", "Fructose and mannose metabolism", "Lysine degradation", "Chloroalkane and chloroalkene degradation", "Benzoate degradation", "Pentose and glucuronate interconversions" and "Cysteine and methionine metabolism" (**Supplementary File S3**). However, we excluded "Starch and sucrose metabolism" and "Riboflavin biosynthesis" after local BLAST analyses showed that homologs of the respective genes were detected in all taxa (**Supplementary File S3**). They were likely not recognized previously in our generation of the shared gene datasets with EDGAR's strict orthology estimation criteria. The genes involved in the remaining nine processes (two of which were involved in siderophore synthesis) were all found to be clustered and allowed comparison of the gene clusters across all taxa containing these genes (**Table 3**).

The two genes identified (*rfbC* and *rfbD*) being involved in "Polyketide sugar unit biosynthesis" were present in all the members of the *P. ananatis* lineage, with various members of the other three lineages lacking the genes (*P. eucalypti*, *P. brenneri*, *Pantoea* sp. GM01 and *P. wallisii*; **Table 3**). Upon examination of the gene cluster containing these genes, two different loci were identified (**Figure 4**). The first locus was observed in nearly all members of the genus that possessed these genes (including *P. dispersa* and *P. stewartii* subsp. *indologenes*), while the second locus (lacking *rfbB*) was found only in *P. ananatis*, *P. stewartii* subsp. *stewartii*, *P. dispersa* and a partial locus in *P. stewartii* subsp. *indologenes*. The first locus was also slightly different in *P. deleyi* and *P. eucrina*, as the position of *rfbC* in *P. deleyi* differed (**Figure 4**—gene indicated in a darker shade) and the locus of *P. eucrina* contained an additional three genes in comparison to the other taxa (**Figure 4**). Furthermore,

TABLE 2 | Intergeneric differences in multi-gene pathways.

Process[a]	Genes	Biological function	Distribution across lineages[b]	References
D-ribose utilization	*rbsK, pgm*	• Converts D-ribose to D-ribose-1-P • Ribose donor during pyrimidine cycling • Involved in co-factor cycling	*Pantoea* (21/21) *Tatumella* (0/3)	Giorgelli et al., 1997; Gossmann et al., 2012; Armenta-Medina et al., 2014
D-mannitol utilization	*mtlAD*	• Converts D-mannitol to β-D-fructose-6-P • Used as carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Berkowitz, 1971; Sprenger, 1993; Boer et al., 1996
Myo-inositol utilization	*iolBCDEG*	• Converts 2-deoxy-5-keto-D-gluconate-6-P to myo-inositol • Used as carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Anderson and Magasanik, 1971; Yoshida et al., 2004
Sucrose utilization	sacA, MGAM, SI	• Degrades sucrose to ADP-glucose and D-fructose • Used as carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Fouet et al., 1986; Van Beers et al., 1995; Sim et al., 2008
Glycogen degradation	*treXYZ*, *glgB*	• Converts glycogen to trehalose or amylose • Used as carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Baecker et al., 1986; Amemura et al., 1988; Chandra et al., 2011
β-oxidation	*fadABE*	• Cleaves fatty acids cyclically into acetyl-CoA (even numbered) or succinyl-CoA (odd-numbered) • Used in citric acid cycle	*Pantoea* (21/21) *Tatumella* (0/3)	Schulz, 1991; Fujita et al., 2007; Liu et al., 2010
Purine metabolism	*guaD, XDH, uaZ, pucM, PRHOXNB*	• Converts guanine to (S)-allantoin • Process involved in nitrogen cycling	*Pantoea* (21/21) *Tatumella* (0/3)	Kimiyoshi et al., 1993; Colloc'h et al., 1997; Nygaard et al., 2000; Moffatt and Ashihara, 2002; Cendron et al., 2007
Creatine degradation	E3.5.3.3, E3.5.1.59	• Converts creatine and N-carbamoylsarcosine to sarcosine	*Pantoea* (21/21) *Tatumella* (0/3)	Yoshimoto et al., 1976; Deeg et al., 1987; Suzuki, 1991
Histidine degradation	*hisD, hutH*	• Converts L-histidinol to urocanate • Use of histidine as sole carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Mehler and Tabor, 1953; Schwede et al., 1999; Yang et al., 2011
L-aspartate degradation	*nadABCX*	• Converts L-aspartate to nicotinate-D-ribonucleotide • Process involved in co-factor cycling	*Pantoea* (21/21) *Tatumella* (0/3)	Yang et al., 2003; Ollagnier-de Choudens et al., 2005; Katoh et al., 2006; Gossmann et al., 2012
Glutamine ABC Transporter	*glnHPQ*	• Transports glutamine from outside environment into cell through the use of ATP	*Pantoea* (0/21) *Tatumella* (3/3)	Nohno et al., 1986; Walshaw et al., 1997; Hosie and Poole, 2001
Glycine/ betaine/proline ABC Transporter	*proVWX*	• Transports glycine, betaine and proline from outside environment into cell through the use of ATP	*Pantoea* (0/21) *Tatumella* (3/3)	Gowrishankar, 1989; Stirling et al., 1989; Kempf and Bremer, 1995; Ko and Smith, 1999
Glutathione ABC Transporter	*gsiABCD (yliABCD)*	• Transports glutathione from outside environment into cell through the use of ATP	*Pantoea* (0/21) *Tatumella* (3/3)	Suzuki et al., 2005
Osmoprotectant ABC Transporter	*opuABCD*	• Transports osmoprotectants from outside environment into cell through the use of ATP	*Pantoea* (21/21) *Tatumella* (0/3)	Kempf and Bremer, 1995; Kappes et al., 1999

(Continued)

TABLE 2 | Continued

Process[a]	Genes	Biological function	Distribution across lineages[b]	References
Taurine ABC Transporter	*tauABC*	• Transports taurine from outside environment into cell through the use of ATP • Allows uptake for subsequent degradation of taurine for sulfur cycling	*Pantoea* (21/21) *Tatumella* (0/3)	Van der Ploeg et al., 1996; Javaux et al., 2007
L-arabinose transporter	*araFGH*	• Transports L-arabinose from outside environment into cell through the use of ATP • Required for utilization of L-arabinose as sole carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Scripture et al., 1987; Kehres and Hogg, 1992; Schneider, 2001
Microcin C ABC Transporter	*yejABEF*	• Transports microcin C from outside environment into cell through the use of ATP • Results in susceptibility to antimicrobial microcin C	*Pantoea* (21/21) *Tatumella* (0/3)	Vanneste et al., 2001; Novikova et al., 2007; Metlitskaya et al., 2009
N-acetylmuramic acid PTS	*murP, crr*	• Transports and phosphorylates N-acetylmuramic acid through the use of ATP • Essential for the use of N-acetylemuramic acid (cell wall component) as carbon source	*Pantoea* (21/21) *Tatumella* (0/3)	Dahl et al., 2004; Uehara et al., 2006,
N-acetyl-D-glucosamine PTS	*nagE*	• Transports and phosphorylates N-acetyl-D-glucosamine through the use of ATP • Essential for the use of N-acetyl-D-glucosamine (cell wall component) as carbon source	*Pantoea* (0/21) *Tatumella* (3/3)	Jaeger and Mayer, 2008; Plumbridge, 2009
Arbutin/salicin PTS	*ascF, crr*	• Transports and phosphorylates arbutin and salicin through the use of ATP • Essential for the use of plant derived glycosides as sole carbon source	*Pantoea* (0/21) *Tatumella* (3/3)	Hall and Xu, 1992; Dahl et al., 2004

[a] *PTS refers to phosphotransferase system.*

[b] *Brackets indicate the number of taxa for which the specific locus is present out of all taxa in the genus.*

selection analyses indicated purifying selection for *rfbA* and *rfbB* and diversifying selection for *rfbC* and *rfbD* in the first locus (**Figure 4**, **Supplementary File S4**). Contrary to this, both *rfbC* and *rfbD* were under purifying selection in the second locus, with *rfbA* being under mainly purifying selection for the first part and diversifying selection for the second part of the gene (**Figure 4**, **Supplementary File S4**).

The genes involved in "Biosynthesis of siderophore group non-ribosomal peptides" and "Lysine degradation" both encoded for different iron acquisitioning molecules (siderophores) (**Table 3**). The genes present in most of the members in the genus ("Biosynthesis of siderophore group non-ribosomal peptides") were identified as being required for the production of enterobactin. The majority of the gene cluster encoding enterobactin appeared to have evolved under purifying selection, with only some regions that evolved mainly under diversifying selection (for example see **Figure 5** *entB*; **Supplementary File S4**). Conversely, the genes involved in "Lysine degradation" in the *P. ananatis* lineage were those required to produce aerobactin from lysine. All members of the *P. ananatis* lineage lacked the genes involved in enterobactin biosynthesis, but contained the genes required for aerobactin synthesis, while all other members of the genus lacked the genes required for aerobactin biosynthesis (**Figure 5**). Selection analyses amongst the genes encoding for aerobactin biosynthesis indicated purifying selection in particular for *iucA* and *iucB*. Both these loci were absent from *P. eucalypti*, *P. deleyi* and *P. eucrina*.

The differentially present genes associated with "Fructose and mannose metabolism" consisted of *rhaA*, *rhaB,* and *rhaD* which convert L-rhamnose to glycerone-P and S-lactaldehyde (*rhaD*) (**Table 3**). This cluster was present in all the members of the *P. agglomerans* and *P. rodasii* lineages, but present only in *P. allii* and *P. ananatis* in the *P. ananatis* lineage, and *P. dispersa* and *P. wallisii* in the *P. dispersa* lineage (**Supplementary File S4**). From the selection analyses of the *P. agglomerans* and *P. rodasii* lineages it was observed that *rhaB* and *rhaD* evolved under purifying selection, with *rhaA* evolving under diversifying selection (**Supplementary File S4**).

Our analysis showed that for "Chloroalkane and chloroalkene degradation", the specific pathway was absent from all the members of the *P. dispersa* lineage (**Table 3**). This pathway

TABLE 3 | Distribution multi-gene pathways among the *Pantoea* lineages.

Processes	Pathway reactions	Genes[a]	*Pantoea* lineages[b]			
			Ag	An	Di	Ro
Polyketide sugar unit biosynthesis	dTDP-4-oxo-6-deoxy-D-glucose <-> dTDP-4-dehydro-beta-L-rhamnose	*rfbC*	v	v	v	v
	dTDP-L-rhamnose <-> dDTP-4-dehydro-beta-L-rhamnose	*rfbD*	v	v	v	v
Biosynthesis of siderophore group non-ribosomal peptides	2,3-Dihydroxybenzoate <-> Enterochelin	*entE & F*	v	–	v	+
	2,3-Dihydroxybenzoate <-> (2S,3S)-2,3-Dihydro-2,3-dihydroxybenzoate	*entA*	v	–	v	+
	(2S,3S)-2,3-Dihydro-2,3-dihydroxybenzoate <-> Isochorismate	*entB*	v	–	v	+
	Isochorismate <-> Chorismate	*entC*	v	–	v	+
Fructose and mannose metabolism	L-Rhamnulose <-> L-Rhamnose	*rhaA*	+	v	v	+
	L-Rhamnulose 1-P <-> L-Rhamnulose	*rhaB*	+	v	v	+
	Glycerone-P + S-Lactaldehyde <-> L-Rhamnulose 1-P	*rhaD*	+	v	v	+
Lysine degradation	L-Lysine <-> N6-Hydroxy-L-Lysine	*iucD*	–	+	–	–
	N6-Hydroxy-L-lysine <-> N6-Acetyl-N6-hydroxy-L-lysine	*iucB*	–	+	–	–
	N6-Acetyl-N6-hydroxy-L-lysine <-> N2-Citryl-N6-acetyl-N6-hydroxy-L-lysine	*iucA*	–	+	–	–
	N2-Citryl-N6-acetyl-N6-hydroxy-L-lysine + N6-Acetyl-N6-hydroxy-L-lysine <-> Aerobactin	*iucC*	–	+	–	–
Chloroalkane degradation	Chloroacetic acid <-> Glycolate + Hydrochloric acid	2-haloacid DH	+	+	–	+
	Chloroacetaldehyde <-> Chloroacetic acid	aldehyde DH	+	+	–	+
Benzoate degradation	Succinyl-CoA + 3-Oxoadipate <-> Succinate + 3-Oxoadipyl-CoA	*pcaI & J*	–	+	–	–
	2-Oxo-2,3-dihydrofuran-5-acetate + H2O <-> 3-Oxoadipate	*pcaL*	–	+	–	–
	3,4-Dihydroxybenzoate + Oxygen <-> 3-Carboxy-cis,cis-muconate	*pcaG & H*	–	+	–	–
Pentose and glucuronate interconVersions	D-Glucuronate <-> D-fructuronate	*uxaC*	+	+	v	+
	D-Galacturonate <-> D-Tagaturonate	*uxaC*	+	+	v	+
	D-Altronate <-> D-Tagaturonate	*uxaB*	+	+	v	+
	D-Altronate <-> 2-Dehydro-3-deoxy-D-gluconate	*uxaA*	+	+	v	+

[a] *The presence or absence of each pathway was verified using BLAST searches with the relevant sequences against the respective genomes.*

[b] *The lineages are indicated as follows: Ag = P. agglomerans lineage, An = P. ananatis lineage, Di = P. dispersa lineage and Ro = P. rodasii lineage. The presence of genes are indicated with "+", their absence with "–", while "v" is used to indicate the presence in some but not all members of a lineage.*

catalyzes the conversion of chloroacetaldehyde to glycolate and hydrochloric acid. All other members of the genus possessed the genes required for this process (**Supplementary File S4**).

The differentially present genes associated with "Benzoate degradation" were involved in the utilization of protochatechuate. They were present only in the *P. ananatis* lineage. However, closer examination revealed that most members of the lineage contained a cluster of 9 genes (*pcaH*, *pcaG*, *pcaQ*, *pcaL*, *pcaB*, KAT, *pcaJ*, *pcaI,* and *pcaR*), but that it contained a deletion in *P. stewartii* subsp. *stewartii* which truncated *pcaL* and removed *pcaB* and KAT from the cluster. Overall, the cluster appeared to be under purifying selection (**Supplementary File S4**).

All members of the genus, except *P. eucrina*, possessed a gene cluster (*uxaA*, *uxaB,* and *uxaC*) involved in "Pentose and glucuronate interconversions" (**Table 3**, **Supplementary File S4**). The products of *uxaA* and *uxaB* catalyze, respectively, the reversible conversion of 2-dehydro-3-deoxy-D-gluconate to D-altronate and D-altronate to D-tagaturonate, while *uxaC* facilitates interconversions between D-tagaturonate and D-glucuronate and between D-fructuronate and D-galacturonate. These three genes were conserved within the *P. agglomerans*, *P. ananatis,* and *P. rodasii* lineages, with only *uxaA* and *uxaB* being present in *P. dispersa* and *P. wallisii*. Overall, it appeared that these genes were evolving under neutral selection (**Supplementary File S4**).

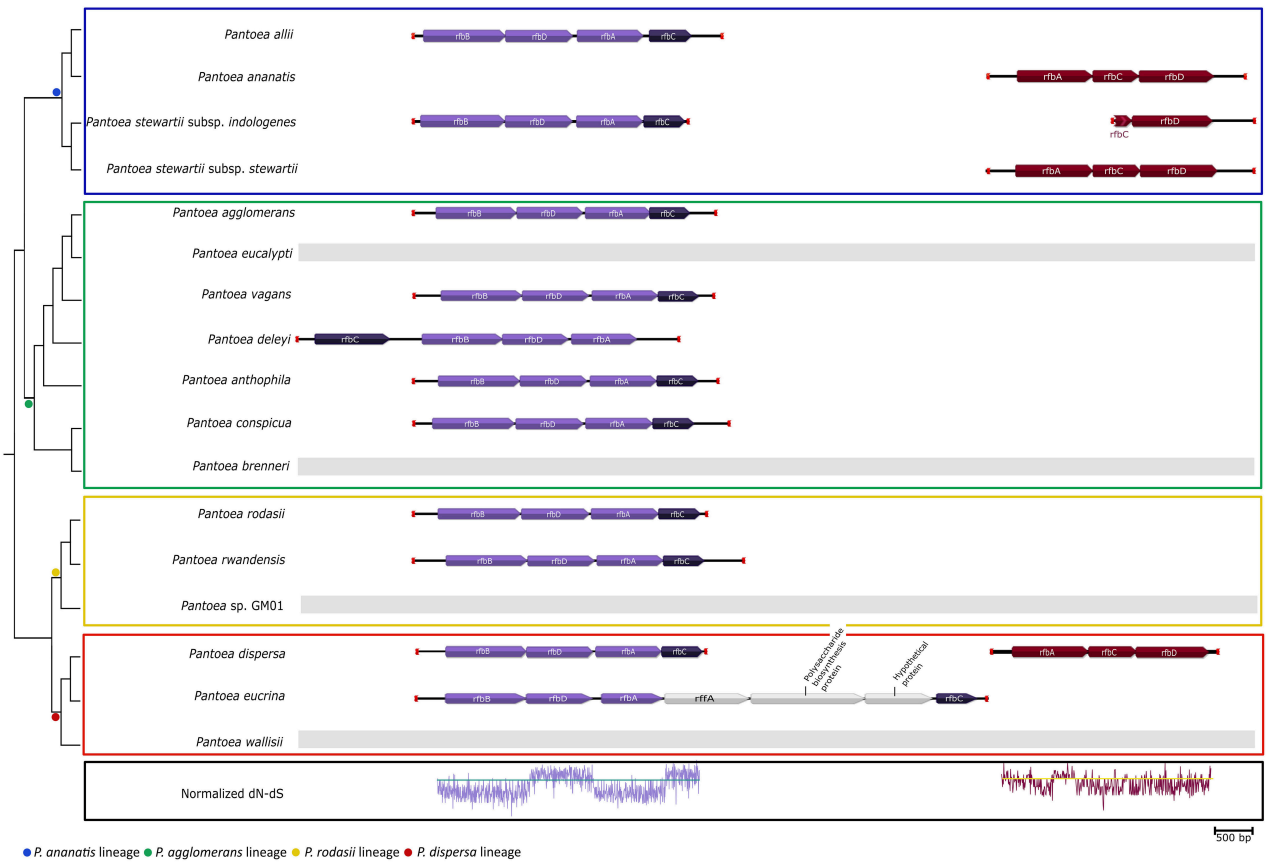


FIGURE 4 | The gene cluster involved in "Polyketide sugar unit biosynthesis" across the lineages of *Pantoea*. The dendrogram was inferred from the amino acid based topology of the core genome of Palmer et al. (2017). The different lineages are indicated with colored blocks. The first locus, containing *rfbA, rfbB, rfbC,* and *rfbD* (purple), was conserved in most members of *Pantoea*. The normalized dN-dS values for each codon position were plotted as an indication of the selective pressures upon the codons. Both *rfbA* and *rbfB* could be observed to experience mainly purifying selection (proportion of non-synonymous substitutions < proportion of synonymous substitutions), while *rfbC* and *rfbD* evolved mainly under diversifying selection (proportion of non-synonymous substitutions > proportion of synonymous substitutions). The second locus was identified in *P. ananatis*, *P. stewartii* subsp. *stewartii* and *P. dispersa*, with a partial locus present in *P. stewartii* subsp. *indologenes* (maroon). This locus lacked an *rfbB* gene and evolved mainly under purifying selection.

Comparison of the processes involved in "Cysteine and methionine metabolism" showed differences in the synthesis of spermidine (*speD* and *speE*) and the methionine salvage pathway (*mtnA*, *mtnB*, *mtnC*, *mtnD,* and *mtnK*) (**Table 3**, **Supplementary Files S3**, **S4**). The two genes required for the biosynthesis of spermidine allow for the conversion of S-adenosyl-L-methionine and putrescine to 5′-methylthioadenosine and spermidine (*speE*). These two genes were present in all members of the genus except *Pantoea* sp. GM01 (*P. rodasii* lineage), *P. eucrina* and *P. wallisii* (both from the *P. dispersa* lineage; **Supplementary File S4**). Genes involved in the methionine salvage pathway allow conversion of 5-methylthio-D-ribose to 3-(methylthio)-propanoate through 5 different intermediate reactions (**Supplementary File S3**). These methionine salvage pathway genes were absent in the *P. dispersa* lineage (*P. dispersa*, *P. eucrina,* and *P. wallisii*) (**Supplementary File S4**).

We also found several multi-gene systems for two-component systems (2 systems), ABC Transporters (14 systems) and PTSs (2 systems) that were differentially present within these lineages (**Supplementary File S5**). Local BLAST analyses allowed identification of taxa where these genes were indeed present, despite not being conserved within the specific gene sets (**Figure 6**, **Supplementary File S5**). The two-component systems identified were that for citrate as well as nitrate/nitrite uptake. The ABC transporters identified were the systems for nitrate/nitrite/cyanate, HMP/FAMP, spermidine/putrescine, putrescine, maltose/maltodextrin, D-xylose, myoinositol-1-phophate, phosphonate, glutamine, arginine, urea, glutathione and iron(II)/manganese. The PTSs detected were those for cellobiose and L-ascorbate.

Annotation of Lineage-Specific Genes without KEGG Associations

A total of 264 genes were identified as being differentially present in the *Pantoea* lineages, to which no KO term assignment could be made. This set of uncharacterized genes consisted of 62 genes in the *P. ananatis* lineage, 75 genes in the *P. agglomerans*

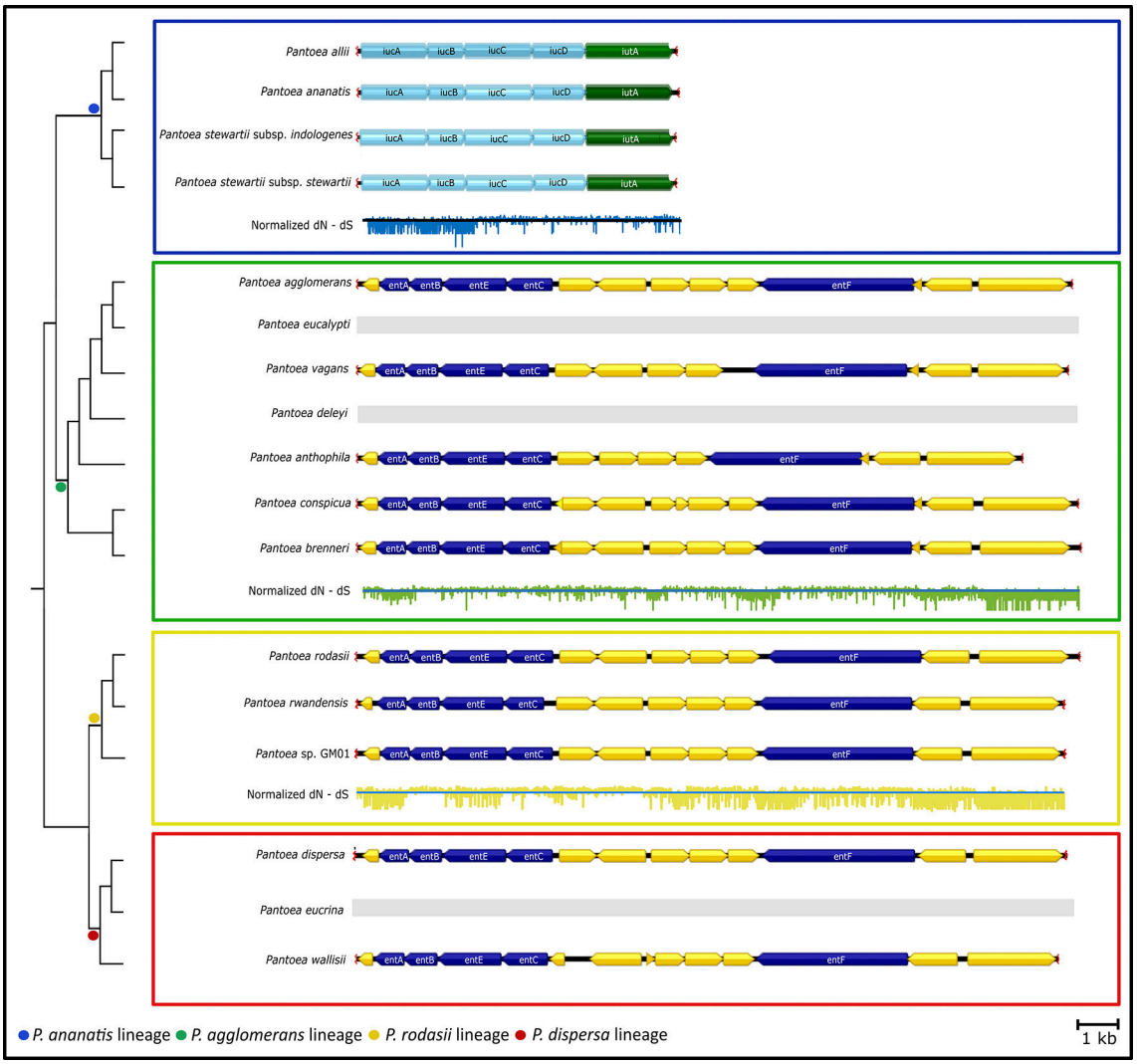


FIGURE 5 | The gene clusters involved in "Lysine degradation" and "Biosynthesis of siderophore group non-ribosomal peptides". The dendrogram was inferred from the species tree of Palmer et al. (2017). Lineages are indicated with colored blocks. Both these clusters encode for the biosynthesis of siderophores, namely aerobactin and enterobactin, respectively. The locus required for the production of aerobactin was conserved in members of the *P. ananatis* lineage, while the locus required for enterobactin biosynthesis was present in most other members of *Pantoea*. The enterobactin biosynthesis locus was completely absent from the genomes of the members of the *P. ananatis* lineage, while the aerobactin locus was lacking in all other members of *Pantoea*. As an indication of selective pressures on the loci, the normalized dN-dS value at each codon position was plotted across the clusters.

lineage, 98 genes in the *P. rodasii* lineage and 29 genes in the *P. dispersa* lineage. Analysis of these genes with Blast2GO allowed annotation of 182 genes. A further six genes could be assigned GO terms, but could not be fully annotated upon merging of annotations due to a lack in InterProScan hits. A total of 76 genes had no functional associations. These unannotated genes could however be used for blastp analyses to identify potential sources of horizontally acquired genes.

The 62 genes of the *P. ananatis* lineage were subjected to Blast2GO analyses, leading to the annotation of 38 genes. In terms of biological processes (GO Level 3), the highest number of genes were involved in "cellular metabolic processes", "primary metabolic processes" and "organic substance metabolic processes", followed by "regulation of cellular processes" and "nitrogen compound metabolic processes" (**Figure 7**). This lineage thus contained 24 genes present in all members of the lineage, without KEGG or GO functional annotations. Based on blastp hits, 15 of the 24 genes had their closest homologs within other members of the *Enterobacteriaceae*, while two genes had homologs in members of the *Rhizobiaceae* (α-Proteobacteria). The closest homolog for three genes was respectively from the *Aurantimonadaceae* (*Martelella mediterranea*; α-Proteobacteria), *Corchorus olitorius* (bush okra; Malvaceae, Eudicots), and *Erwinia* phage ENT90. The remaining four genes had no BLAST hits (blastp) on the non-redundant database (**Supplementary File S6**).

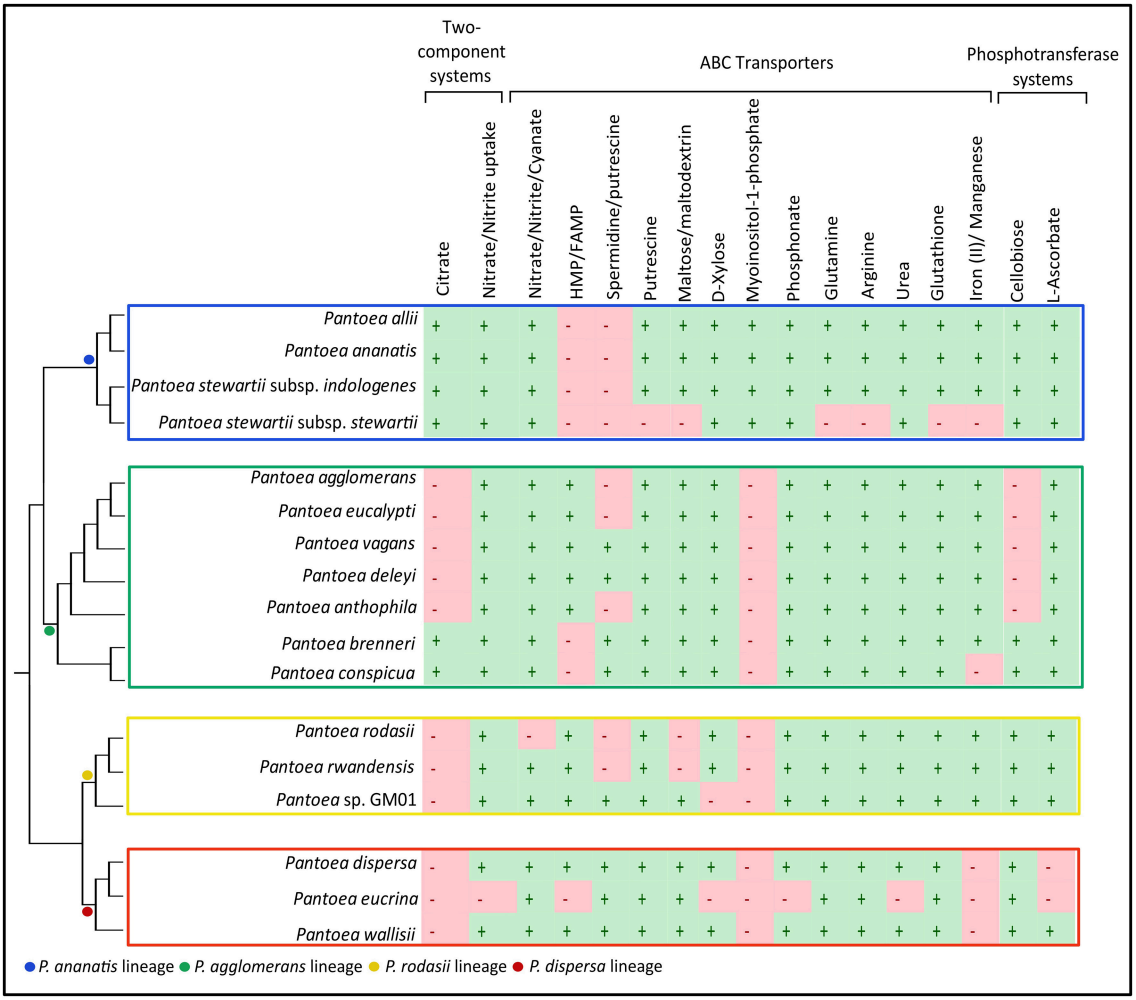


FIGURE 6 | Differences between the lineages in processes involved in "Environmental Information Processing". The presence (+) or absence (–) of complete (all genes required for functional system) two component systems, ABC transporters and PTSs in the genomes of the species in the main lineages within *Pantoea*. The dendrogram of the relationships within and between lineages were inferred from Palmer et al. (2017). The separate lineages are indicated with colored blocks.

Of the 75 genes conserved within the *P. agglomerans* lineage not annotated with KEGG, 54 genes could be annotated with Blast2GO. Most of these genes were involved, in descending order, in "cellular metabolic processes", "organic substance metabolic processes", "establishment of localization", "primary metabolic processes" and "biosynthetic processes" (BP GO Level 3; **Figure 7**). The remaining 21 genes with no associated GO terms could not be functionally classified. Homologs of these genes were however, identified in other members of the *Enterobacteriaceae*, often pathogens, with a single gene having its closest homolog in the metagenome of a soil sample from an unknown source (**Supplementary File S6**).

Of the 98 genes conserved within the *P. rodasii* lineage without KEGG annotations, 76 could be annotated with Blast2GO. The five highest biological processes in which these genes were involved were "organic substance metabolic processes", "cellular metabolic processes", "primary metabolic processes", "regulation of cellular processes" and "nitrogen compound metabolic processes" (GO Level 3; **Figure 7**). This resulted in 22 genes without any functional annotation with either KEGG or GO analyses. Homologs for all 22 genes were identified in other members of the *Enterobacteriaceae*, of which 21 genes were most closely related to genes from human pathogens (**Supplementary File S6**).

Of the 29 unique genes in the *P. dispersa* lineage without any KEGG annotations, 14 genes could be annotated with Blast2GO. These genes were primarily involved in "cellular metabolic processes", "organic substance metabolic processes", "primary metabolic processes", "nitrogen compound metabolic processes" and "biosynthetic processes" (**Figure 7**). Of the 15 unannotated genes, homologs for all genes were identified in other members of the *Enterobacteriaceae*, particularly those associated with the stinkbug, *Plautia stali* (**Supplementary File S6**).

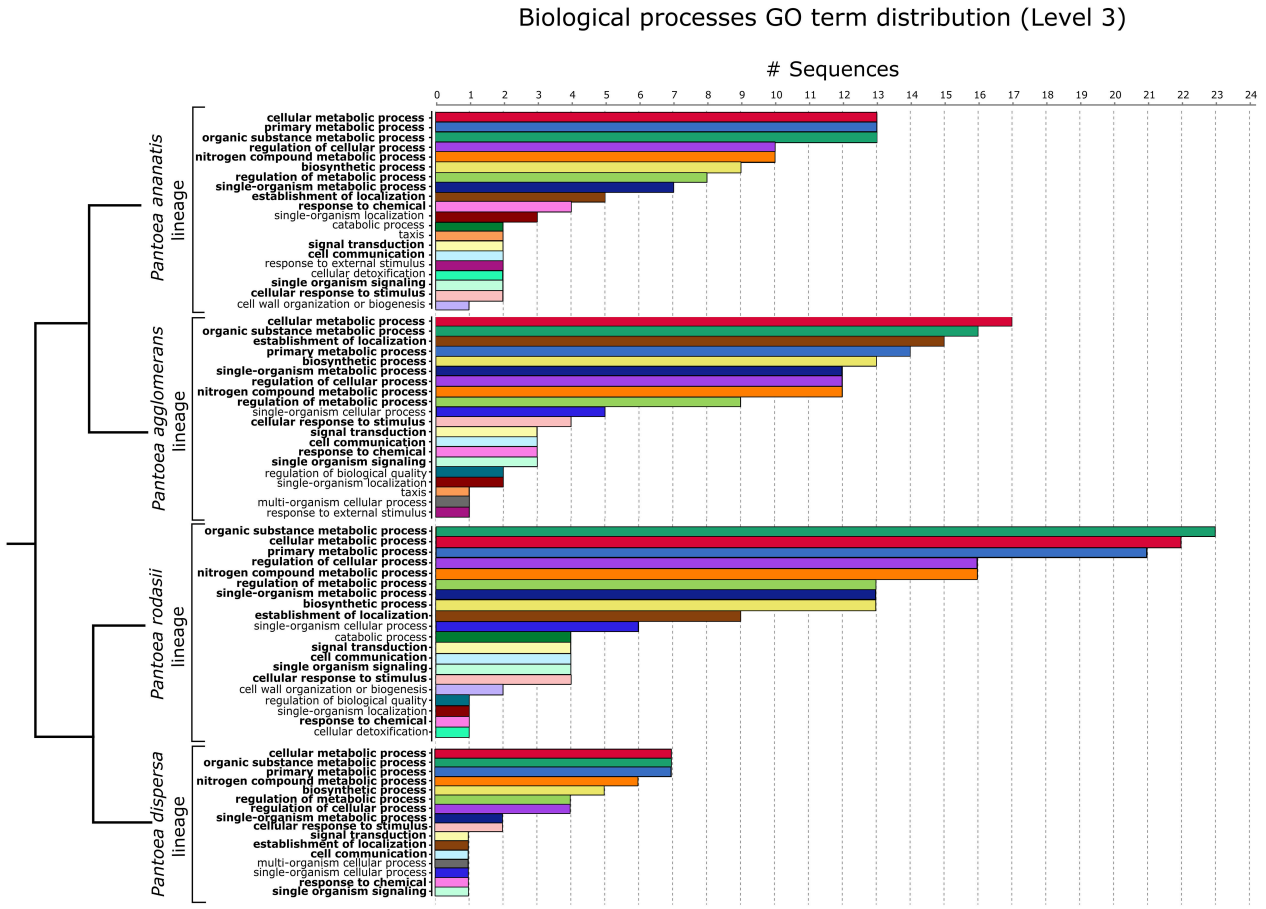


FIGURE 7 | The top 20 biological processes (Level 3) for the *P. ananatis*, *P. agglomerans* and *P. rodasii* lineages and the 16 biological processes for the *P. dispersa* lineage, of the 182 genes annotated with Blast2GO. The dendrogram indicating the relationships between the lineages was inferred from the species tree of Palmer et al. (2017). The GO terms for the shared processes are indicated in the same color across the lineages. Processes highlighted in bold are shared by all four lineages.

DISCUSSION

Our findings suggest that *in silico* mining of bacterial genome sequences is a feasible approach for inferring large sets of biological characters for particular taxa. This approach is invaluable for unveiling large repertoires of potential bacterial phenotypes and can thus contribute hugely toward identifying biologically relevant diagnostic characteristics from whole genome sequences. Furthermore, by superimposing such characters onto the phylogeny of a particular bacterial group it appears to be possible to identify those traits that might have contributed toward the initial emergence of a taxon and/or its subsequent stable persistence in nature. Here we identified extensive sets of biological characters specific to *Pantoea* and its main phylogenetic lineages. Our study thus outlines the initial steps toward linking biological functions (based on the variable genomic components) to taxonomy (based on the stable, conserved genomic components).

Genome-Based Comparisons of Specific Processes between Genera

The methodology employed in this study allowed for the identification of biological characters that potentially define and differentiate bacterial genera from one another. Despite the necessity of these taxonomic ranks, our understanding of what constitutes and distinguishes genera is mostly limited. Previous attempts to obtain natural and logical groupings have always been based on a limited view of the organisms' metabolic potential, often with a focus on what was considered to be clinically relevant data rather than from a biological outlook (Konstantinidis and Tiedje, 2005a,b; Sutcliffe, 2015). Although, the current classification system aims to identify and describe naturally occurring groups by employing an evolution-based approach, it still does not provide any biologically meaningful information for the organisms (Cohan, 2002; Konstantinidis and Tiedje, 2005a; Tindall et al., 2010). However, our study of *Pantoea* and *Tatumella* clearly highlights how diverse sets of biological characters for bacterial genera may be inferred

from genome data. Apart from so-called genus-defining traits that can potentially be used to differentiate these taxa, these characters also provide information on the general biology of the taxa investigated (see **Table 3**). Our findings indicate that such genome-based analyses provide a more informed view of the biology of the organisms, and the information emerging from comparing metabolic differences can be linked to the shared ancestry of groups of organisms.

Pantoea appears to be metabolically more versatile than its sister genus *Tatumella*. Different from *Tatumella*, it encodes a range of additional pathways potentially enabling it to use diverse compounds as nutrient sources [e.g., fatty acids (Schulz, 1991; Fujita et al., 2007; Liu et al., 2010) and various carbohydrate derivatives (Mehler and Tabor, 1953; Anderson and Magasanik, 1971; Berkowitz, 1971; Yoshimoto et al., 1976; Baecker et al., 1986; Fouet et al., 1986; Deeg et al., 1987; Amemura et al., 1988; Sprenger, 1993; Van Beers et al., 1995; Boer et al., 1996; Schwede et al., 1999; Yoshida et al., 2004; Sim et al., 2008; Chandra et al., 2011; Yang et al., 2011)]. *Pantoea* also encodes additional nutrient cycling or salvage systems [e.g., purine, pyrimidine and co-factor cycling (Kimiyoshi et al., 1993; Colloc'h et al., 1997; Giorgelli et al., 1997; Nygaard et al., 2000; Moffatt and Ashihara, 2002; Yang et al., 2003; Ollagnier-de Choudens et al., 2005; Katoh et al., 2006; Cendron et al., 2007; Gossmann et al., 2012; Armenta-Medina et al., 2014)]. These systems have been shown to allow for the recycling of compounds that are no longer utilized in the cell, and might enhance *Pantoea*'s ability to perform basic, yet essential cellular functions under nutrient limiting conditions (Krismer et al., 2014; Shimizu, 2014).

Pantoea and *Tatumella* differ markedly in terms of KEGG's "Environmental Information Processing" functional category, which includes all signaling and membrane transport pathways (Kanehisa et al., 2002). Various ABC transporters (Nohno et al., 1986; Scripture et al., 1987; Gowrishankar, 1989; Stirling et al., 1989; Kehres and Hogg, 1992; Kempf and Bremer, 1995; Van der Ploeg et al., 1996; Walshaw et al., 1997; Kappes et al., 1999; Ko and Smith, 1999; Hosie and Poole, 2001; Schneider, 2001; Vanneste et al., 2001; Suzuki et al., 2005; Javaux et al., 2007; Novikova et al., 2007; Metlitskaya et al., 2009) and phosphotransferase systems (PTSs) (Hall and Xu, 1992; Dahl et al., 2004; Uehara et al., 2006; Jaeger and Mayer, 2008; Plumbridge, 2009) were differentially present in the two genera. Among the various ABC transporters identified only in *Pantoea*, one has been associated with susceptibility to microcin C in the absence of a microcin C-specific efflux pump (Metlitskaya et al., 2009), which is part of a group of antibiotic produced by certain *Enterobacteriaceae* (Vanneste et al., 2001; Metlitskaya et al., 2009). The absence of this ABC transporter in *Tatumella* and the concomitant antibiotic resistance may increase ecological competitiveness of species exposed to these compounds (Hacker and Carniel, 2001). Ecological advantages are likely also obtained from some of the predicted PTSs, which have previously been linked to enhanced recycling of cell wall components under nutrient-poor conditions (e.g., PTSs involving N-acetylmuramic acid and N-acetylglucosamine; Jaeger and Mayer, 2008), and the uptake of plant-derived carbon compounds (e.g., PTS involving arbutin and salicin; Zangoui et al., 2015).

Taken together, these findings suggest that evolution has equipped *Pantoea* with extensive repertoires of metabolic processes that make them generally more versatile in their ability to adapt to changing environments. Compared to *Tatumella*, they can utilize a wider range of carbon sources and use available resources more efficiently by recycling metabolic byproducts. This potentially also provides them with a competitive advantage in nutrient-poor environments such as mammalian blood. The various genus-defining traits we identified for *Pantoea* may thus contribute to our understanding of the complex, and often opportunistic, relationships these species have with their plant and animal hosts (De Baere et al., 2004; Cruz et al., 2007; De Maayer et al., 2012a).

Genome-Based Comparisons of Specific Processes between Lineages of *Pantoea*

Comparisons of the metabolic processes inferred from whole genome sequences allowed for the identification of various sets of traits specific to one or more lineages of *Pantoea*. Based on previous work in diverse bacteria (including *Pantoea*), we attempted to correlate these processes to the lifestyles of the taxa investigated. Although a number of the identified processes were likely related to pathogenicity (see **Table 4**), most probably play roles in niche adaptation and utilization in a non-pathogenic capacity (see **Table 5**).

The processes likely associated with pathogenesis, particularly in the *Enterobacteriaceae*, were those involved in O-antigen (Stevenson et al., 1994; Whitfield, 1995; Wang and Reeves, 1998; Kohchi et al., 2006; Greenfield and Whitfield, 2012), siderophore (Montgomerie et al., 1984; Williams and Carbonetti, 1986; Opal et al., 1990; Fecteau et al., 2001; Fiedler et al., 2001; Torres et al., 2001; Hubertus et al., 2003; Raymond et al., 2003; Garcia et al., 2011; Gao et al., 2012) and polyamine (Khan et al., 1992; Ha et al., 1998; Gugliucci and Menini, 2003; Shah and Swiatlo, 2008; Pegg, 2016) biosynthesis. Differences in the locus involved in O-antigen biosynthesis (of which some *Pantoea* species have two) were previously associated with a pathogen's ability to escape host responses (Whitfield, 1995; Kohchi et al., 2006; Greenfield and Whitfield, 2012). Our results further showed that all species in the *P. ananatis* lineage likely produce aerobactin, while many of those in the other lineages produce enterobactin. Although these siderophores essentially perform the same function, aerobactin is more efficient at scavenging iron during nutrient limitation, and may in some instances even assist in resistance against iron-dependent antimicrobials (Williams and Carbonetti, 1986; Torres et al., 2001; Garcia et al., 2011). Also, all of the examined species in the *P. agglomerans* and *P. ananatis* lineages are predicted to be capable of producing the polyamine spermidine (this process was detected in only some of the species in the other two lineages). Apart from their essential cellular functions (Ha et al., 1998; Shah and Swiatlo, 2008; Pegg, 2016), polyamines have been implicated in biofilm formation, escape from host phagolysosomes and toxin production and activity (Shah and Swiatlo, 2008).

The group of processes likely associated with niche adaptation and utilization were those related to nutrient metabolism and "Environmental Information Processing"

TABLE 4 | Pathogenicity-associated processes with differences between the lineages.

Process	Gene cluster	Function, biological role and contribution to pathogenicity	Distribution across lineages[a]	References
O-antigen biosynthesis	*rfb* locus 1	• Encodes for the synthesis of polysaccharide unit of O-antigen • Cause variation on outer membrane of cells • Assists in host immune response evasion	*P. ananatis* lineage (2/4) *P. agglomerans* lineage (5/7) *P. rodasii* lineage (2/3) *P. dispersa* lineage (2/3)	Stevenson et al., 1994; Whitfield, 1995; Wang and Reeves, 1998; Kohchi et al., 2006; Greenfield and Whitfield, 2012
	rfb locus 2	• Encodes for the synthesis of the O-antigen's polysaccharide • Cause variation on outer membrane of cells • Assists in host immune response evasion	*P. ananatis* lineage (3/4) *P. dispersa* lineage (1/4)	Stevenson et al., 1994; Whitfield, 1995; Wang and Reeves, 1998; Kohchi et al., 2006; Greenfield and Whitfield, 2012
Siderophore production	*iuc* locus	• Encodes for the synthesis of aerobactin • Efficient at scanvenging iron during nutrient limitation • Assists in resistance against iron-dependent antimicrobials	*P. ananatis* lineage (4/4)	Montgomerie et al., 1984; Williams and Carbonetti, 1986; Opal et al., 1990; Fecteau et al., 2001; Torres et al., 2001; Garcia et al., 2011; Gao et al., 2012
	ent locus	• Encodes for the synthesis of enterobactin • High affinity for iron, but susceptible to inactivation by lipocalin-2 produced by mammalian hosts • Iron acquisitioning in hosts	*P. agglomerans* lineage (5/7) *P. rodasii* lineage (3/3) *P. dispersa* lineage (2/3)	Fiedler et al., 2001; Torres et al., 2001; Hubertus et al., 2003; Raymond et al., 2003; Garcia et al., 2011
Polyamine biosynthesis	*spe* locus	• Synthesis of spermidine from S-adenosyl-L-methionine • Involved in protein and nucleic acid biosynthesis, expression regulation, membrane functioning and ROS scavenging • Contributes toward pathogenesis through involvement in biofilm formation, escape from phagolysosomes and toxin production	*P. ananatis* lineage (4/4) *P. agglomerans* lineage (7/7) *P. rodasii* lineage (2/3) *P. dispersa* lineage (1/3)	Khan et al., 1992; Ha et al., 1998; Gugliucci and Menini, 2003; Shah and Swiatlo, 2008; Pegg, 2016

[a] *Brackets indicate the number of taxa for which the specific locus is present out of all taxa in a lineage. For complete distribution patterns see* ***Supplementary File S4****.*

(**Supplementary Table S1**). For example, other than those in the *P. dispersa* lineage, all *Pantoea* species were predicted to be capable of converting the environmental mutagen chloroacetaldehyde to glycolate, thus providing the dual means of disposing of the mutagen and accessing glycolate as carbon source (Young Kim et al., 2007; Maciejewska et al., 2010). Similarly, in all species, bar those of the *P. dispersa* lineage, the methionine salvage pathway likely allow increased efficacy under sulfur cycling in nutrient-poor conditions (Sekowska et al., 2000, 2004; Albers, 2009). Species in the *P. ananatis* lineage encode the benzoate degradation pathways, which likely enable their utilization of protocatechuate as carbon source (Song, 2009; Brady et al., 2011; Gueule et al., 2015). Additionally, differences were also observed in rhamnose (Badía et al., 1985; Moralejo et al., 1993; Saxena et al., 2010) and galacturonate (Walton, 1994; Hématy et al., 2009; Richard and Hilditch, 2009) utilization. The four lineages further differed in terms of their ability to transport various nutrients (e.g., the myoinositol-1-phosphate ABC transporter occurred only in the *P. ananatis* lineage). The same was also true for the predicted two-component signaling systems and other PTSs (e.g., except for two *P. agglomerans* lineage species, only the *P. ananatis* lineage encoded a two-component signaling system for citrate utilization) (**Supplementary Table S1**).

Overall, we could correlate the versatility in lifestyle and host to increased metabolic potential (in terms of compounds utilized) as well as pathogen-associated traits between the different lineages within *Pantoea*. Despite the association of various *Pantoea* species with only clinical infections (*P. brenneri, P. conspicua,* and *P. eucrina*), it appears that species associated with opportunistic clinical infections (De Baere et al., 2004; Cruz et al., 2007), possessed genes often associated with animal pathogenicity within the *Enterobacteriaceae*. In general, it appears that the lineages with the most diverse niche associated processes also corresponded to those species isolated from the most diverse environments. For example, members of the *P. agglomerans* and *P. ananatis* lineages are routinely isolated from various plant species, as epiphytes, endophytes, or pathogens, as well as from insects, animals and humans (Coutinho and Venter, 2009; Walterson and Stavrinides, 2015), while members of the *P. dispersa* and *P. rodasii* lineages are usually only associated with a single host organism, with the exception of *P. dispersa*.

TABLE 5 | Niche-associated processes (non-pathogenic) with differences between the lineages.

Process	Gene cluster	Function, biological role and contribution to niche adaptation	Distribution across lineages[a]	References
L-rhamnose utlization	*rha* locus	• Convert L-rhamnose to S-lactaldehyde • Catabolism of L-rhamnose as carbon source • Potentially contributes toward niche expansion	• *P. ananatis* lineage (2/4) • *P. agglomerans* lineage (7/7) • *P. rodasii* lineage (3/3) • *P. dispersa* lineage (2/3)	Badía et al., 1985; Moralejo et al., 1993; Saxena et al., 2010
D-galacturonate utilization	*uxa* locus	• Involved in the utilization of glucuronate and galacturonate as carbon sources • Integral parts of plant cell walls • Can contribute toward colonization of plants	• *P. ananatis* lineage (4/4) • *P. agglomerans* lineage (7/7) • *P. rodasii* lineage (3/3) • *P. dispersa* lineage (2/3 partial)	Walton, 1994; Hématy et al., 2009; Richard and Hilditch, 2009
Benzoate degradation	*pca* locus	• Involved in the utilization of protochatechuate • Catabolism of protochatechuate as carbon source • Potentially contributes toward niche expansion	• *P. ananatis* lineage (4/4) • (note, *P. stewartii* ssp. *stewartii* • contains deletion)	Song, 2009; Brady et al., 2011; Gueule et al., 2015
Chloroalkane and chloroalkene degradation	-	• Conversion of chloroacetaldehyde to glycolate • Chloroacetaldehyde binds to DNA and causes conformational changes leading to mutations • Removes mutagen from environment and utilize it as carbon source via glycolate	• *P. ananatis* lineage (4/4 • *P. agglomerans* lineage (7/7) • *P. rodasii* lineage (3/3)	Young Kim et al., 2007; Maciejewska et al., 2010
Methionine salvage pathway	*mtn* locus	• Salvage methionine through the conversion of 5-methylthio-D-ribose to 3-(methylthio)-propanoate • Allows sulfur cycling • Can survive in low-sulfur environments	• *P. ananatis* lineage (4/4 • *P. agglomerans* lineage (7/7) • *P. rodasii* lineage (3/3)	Sekowska et al., 2000, 2004; Albers, 2009

[a] *Brackets indicate the number of taxa for which the specific locus is present out of all taxa in a lineage. For complete distribution patterns see* ***Supplementary File S4****.*

These characteristics may contribute to the opportunistic nature of lineages containing species like *P. ananatis* and *P. agglomerans* that are proven plant pathogens, but isolated from diverse environments including the clinical setting.

Evolution of Multi-gene Pathways in *Pantoea* and Its Lineages

Various evolutionary mechanisms likely shaped the presence and distribution of the multi-gene pathways inferred for *Pantoea*. Bacteria propagate asexually and progeny are anticipated to contain the same genetic material as the parent (Daubin et al., 2002). Any changes in an individual's genetic makeup can become fixed in populations if they provide a competitive advantage, or at least have no deleterious effects (Cohan, 2001, 2005; Caro-Quintero and Konstantinidis, 2012). The most common forces facilitating genetic change are random mutations (point mutations as well as insertions and deletions) and horizontal gene transfer (HGT) (Gogarten et al., 2002; Cohen et al., 2011). Accordingly, more closely related species are likely to encode similar pathways, while those subject to HGT would have a more spurious distribution (Gogarten et al., 2002; Cohen et al., 2011).

Evolution of most of the multi-gene pathways in *Pantoea* and its lineages involved complex processes, involving vertical descent with lineage- and species-specific gene losses/gains via duplication or HGT. For example, lineage- and species-specific gene losses would be characterized by distribution patterns where particular gene clusters are present in all taxa neighboring a species lacking it, because of gene losses at a specific ancestral node (Koskiniemi et al., 2012). In our study, such processes were likely involved in the loci required for the conversion of chloroacetaldehyde to glycolate and in the methionine salvage pathway. In contrast to this, the sudden appearance of genes or loci in a lineage that are lacking in all neighboring taxa suggest they were acquired via HGT (Zaneveld et al., 2008). An example of this is the locus required for protochatechuate utilization. The evolution of the siderophore loci likely involved gene losses together with the horizontal acquisition of genes. This is evident from the absence of the enterobactin locus (despite its presence in all closely related lineages) from the *P. ananatis* lineage, and the presence of the locus encoding aerobactin in only this lineage, thus acquired through an HGT event.

Our results suggested that the gene clusters encoding the various *Pantoea* pathways examined in this study, mainly experience purifying selection. Numerous hypotheses attempt to explain why genes cluster and how clusters are maintained (Carbone et al., 2007; Geddy and Brown, 2007; Fang et al., 2008; Sorrels et al., 2009). However, various studies have showed that purifying selection may contribute to the stability and functionality of gene clusters once they have formed (Carbone

et al., 2007; Geddy and Brown, 2007; Fang et al., 2008; Sorrels et al., 2009). Purifying selection seems to facilitate the maintenance of ancient or ancestral gene clusters by limiting the possibility of non-synonymous mutations becoming fixed, which in turn allows out-competition of individuals undergoing deleterious or lethal mutations inhibiting the functioning of the gene cluster (Fang et al., 2008; Sorrels et al., 2009). In our study, purifying selection appeared to play a role in the maintenance and functioning of the loci for siderophore biosynthesis (aerobactin and enterobactin loci) and protochatechuate utilization, which are evolving mainly under the influence of purifying selection. However, we also detected diversifying selection for some of the genes/gene regions examined [e.g., *rfbC* and *rfbD* (*rfb* locus 1), both associated with the conversion of dTDP-6-deoxy-D-xylo-4-hexulose to dTDP-L-rhamnose (Graninger et al., 1999)]. In these cases, selection is causing non-synonymous changes in the sequences of genes or parts of genes, thus driving the appearance of new alleles.

In Silico Predictions vs. Experimentally Confirmed Multi-gene Processes

Although a large number of differential characters were identified from the genome sequences of these organisms, little has been done in terms of experimental verification. Also, a number of characters could not be correlated to current experimental knowledge as false negative and positive results for phenotypic tests are common (Sutcliffe et al., 2012). For instance, despite the presence of genes required for the utilization of arbutin and salicin as carbon source in *Tatumella*, previous phenotypic tests have previously tested negative (Brady et al., 2010b; Tracz et al., 2015). This was also observed for the intra-generic comparisons. Examples within *Pantoea* included the uptake and utilization of cellobiose (Brady et al., 2009) and D-galacturonate, previously identified as a genus-defining attribute (Gavini et al., 1989; Brady et al., 2010a), as well as citrate (Brady et al., 2010a, 2012). In these cases, phenotypic tests previously confirmed the ability to perform these functions, but these genes were lacking in various *Pantoea* species (absence confirmed in available genome sequences of additional isolates of these species). This lack in correlation may either be as a result of gene expression complexities during phenotypic tests, to sequencing, assembly or annotation errors or the presence of as of yet uncharacterized alternate pathways.

There were, however, also a number of *in silico* functional predictions that correlated well with the results of previous phenotypic tests. For example, the lack of genes required for the utilization of histidine in *Tatumella*, correlated entirely with negative results of all previous phenotypic tests (Brady et al., 2010b; Tracz et al., 2015). Moreover, the utilization of protochatechuate has previously tested positive in *P. allii* (Brady et al., 2011), and our findings showed that the locus encoding the necessary gene products is indeed present within this lineage. Other examples of characters supporting previously performed phenotypic tests are the transport for the utilization of D-xylose, maltose, myo-inositol-1-P, and sucrose. In all of the above mentioned examples gene clusters were only observed in taxa previously testing positive for the associated phenotypic characters (Brady et al., 2010a,b, 2011).

Perspectives and Relevance

The increase in data obtained from bacterial genome sequences has superseded the rate at which gene discovery, characterization, and verification can occur. This means that a number of genes could not be assigned definitive functions as no homologs were detected for these genes in the KEGG database. The majority of these genes, at both the generic and lineage-specific level, could be annotated with gene ontologies with Blast2GO, although the functions of some genes remained unknown. Of the uncharacterized genes, the most abundant GO terms were generally quite similar (data for genera not shown). This indicates that the unknown genes could be performing similar functions within the different lineages or genera, despite not being orthologous amongst the different gene sets. The genes to which no functional annotation could be assigned seemed to have originated mostly within the *Enterobacteriaceae*, either acquired through lateral acquisition of genes from other taxa within this family, or, in the case where genes are conserved in most of the lineages, acquired by an ancestor and subsequently lost in some lineages when the genes were no longer required. A number of unique genes were also identified with no known homologs, but expression of these genes should be confirmed before they are considered for further investigation. Although no functional information is available for the unannotated genes, the distribution across various taxa provides insight into potential HGT events.

The availability of whole genome sequence data has transformed the approaches available for understanding bacterial evolution. This has, however, not yet replaced traditional methods such as DNA-DNA hybridization, monophyly in phylogenetic analyses and physiological and metabolic tests for the delineation of bacterial taxa. Although physiologic capabilities provide an insight into what potential metabolic differentiation may have occurred during speciation, inconsistencies may still occur due to differential expression of genes in different isolates or regulation of expression in specific environments. In recent years, we have been moving toward approaches aiming to identify natural groups at higher taxonomic levels by implementing monophyly as a prerequisite for taxon descriptions, but no light can be shed on the biology of the organisms through these approaches. Instead we need to investigate the more variable genomic compartments reflecting the biology of the organisms to obtain a more natural and robust taxonomic system. By employing this approach, one would be able to supplement or supplant the available diagnostic characters used in bacterial taxonomy. Additionally, obtaining a holistic biological perspective from the genome will provide power to predict the lifestyle and ecology of the organisms and is essentially much more informative than only having discriminative power between taxa. We thus believe that this approach of identifying genome-based characteristics in metabolic networks for the taxonomic levels higher than the species, provide an approach of identifying biologically relevant differences along the course of speciation.

AUTHOR CONTRIBUTIONS

All authors contributed toward the original concept or data analyses, together with the drafting, revision and approval of the final manuscript. MP, ES, MC, and SV: was involved in the conceptualization and design of the work; MP and JB: was involved in data acquisition and analysis and all authors were involved in the interpretation of data.

FUNDING

The authors would like to acknowledge the National Research Foundation for student funding through the Centre of Excellence in Tree Health Biotechnology (CTHB) and the University of Pretoria.

ACKNOWLEDGMENTS

We would like to acknowledge the Centre of Excellence in Tree Health Biotechnology (CTHB) and the Tree Protection Co-operative Programme (TPCP) at the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, for access to computing infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00113/full#supplementary-material

Supplementary Table S1 | Environmental information processing traits with differences among the lineages.

Supplementary File S1 | Output from GhostKOALA for the *Pantoea* and *Tatumella* shared gene sets. This file contains the differences observed for the overview maps and the specific pathways for *Pantoea* and *Tatumella*. The overlay figures of the overview and pathway maps are also indicated.

Supplementary File S2 | Output from GhostKOALA for the lineages within *Pantoea*. This file contains the differences observed for the overview maps as well as the overlay figures of the overview maps for the different lineages.

Supplementary File S3 | The summary of differences in pathways requiring 2 or more genes, as well as a summary of the BLAST confirmations of these genes.

Supplementary File S4 | The results from the selection analyses and the figures for the gene clusters not indicated in text.

Supplementary File S5 | The differences for pathways involved in "Environmental Information Processing". A summary of the BLAST confirmation is also included as well as the maps for each lineage for the ABC transporters, two-component systems and the PTSs.

Supplementary File S6 | Summary of the Blast2GO analyses as well as the BLAST hits for genes not annotated with Blast2GO. A pie chart indicating the distribution of BLAST hits is also indicated.

REFERENCES

Albers, E. (2009). Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5′-methylthioadenosine. *IUBMB Life* 61, 1132–1142. doi: 10.1002/iub.278

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

Amemura, A., Chakraborty, R., Fujita, M., Noumi, T., and Futai, M. (1988). Cloning and nucleotide sequence of the isoamylase gene from *Pseudomonas amyloderamosa* SB-15. *J. Biol. Chem.* 263, 9271–9275.

Anderson, W. A., and Magasanik, B. (1971). The pathway of myo-inositol degradation in *Aerobacter aerogenes*: conversion of 2-deoxy-5-keto-d-gluconic acid to glycolytic intermediates. *J. Biol. Chem.* 246, 5662–5675.

Armenta-Medina, D., Segovia, L., and Perez-Rueda, E. (2014). Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *BMC Genomics* 15:800. doi: 10.1186/1471-2164-15-800

Badía, J., Ros, J., and Aguilar, J. (1985). Fermentation mechanism of fucose and rhamnose in *Salmonella typhimurium* and *Klebsiella pneumoniae*. *J. Bacteriol.* 161, 435–437.

Baecker, P. A., Greenberg, E., and Preiss, J. (1986). Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* 1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-alpha-D-(1, 4-alpha-D-glucano)-transferase as deduced from the nucleotide sequence of the glg B gene. *J. Biol. Chem.* 261, 8738–8743.

Bennett, J. S., Jolley, K. A., Earle, S. G., Corton, C., Bentley, S. D., Parkhill, J., et al. (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology.* 158, 1570–1580. doi: 10.1099/mic.0.056077-0

Berkowitz, D. (1971). D-Mannitol utilization in *Salmonella typhimurium*. *J. Bacteriol.* 105, 232–240.

Beukes, C. W., Palmer, M., Manyaka, P., Chan, W. Y., Avontuur, J. R., van Zyl, E., et al. (2017). Genome Data provides high support for generic boundaries in burkholderia sensu lato. *Front. Microbiol.* 8:1154. doi: 10.3389/fmicb.2017.01154

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44, W22–W28. doi: 10.1093/nar/gkw255

Boer, H., ten Hoeve-Duurkens, R. H., and Robillard, G. T. (1996). Relation between the oligomerization state and the transport and phosphorylation function of the *Escherichia coli* mannitol transport protein: interaction between mannitol-specific enzyme II monomers studied by complementation of inactive site-directed mutants. *Biochemistry* 35, 12901–12908. doi: 10.1021/bi9611016

Brady, C. L., Cleenwerck, I., van der Westhuizen, L., Venter, S. N., Coutinho, T. A., and De Vos, P. (2012). *Pantoea rodasii* sp. nov., *Pantoea rwandensis* sp. nov. and *Pantoea wallisii* sp. nov., isolated from *Eucalyptus*. *Int. J. Syst. Evol. Microbiol.* 62, 1457–1464. doi: 10.1099/ijs.0.032615-0

Brady, C. L., Cleenwerck, I., Venter, S. N., Engelbeen, K., De Vos, P., and Coutinho, T. A. (2010a). Emended description of the genus Pantoea, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. *Int. J. Syst. Evol. Microbiol.* 60, 2430–2440. doi: 10.1099/ijs.0.017301-0

Brady, C. L., Goszczynska, T., Venter, S. N., Cleenwerck, I., De Vos, P., Gitaitis, R. D., et al. (2011). *Pantoea allii* sp. nov., isolated from onion plants and seed. *Int. J. Syst. Evol. Microbiol.* 61, 932–937. doi: 10.1099/ijs.0.022921-0

Brady, C. L., Venter, S. N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J., et al. (2009). *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. *Int. J. Syst. Evol. Microbiol.* 59, 2339–2345. doi: 10.1099/ijs.0.009241-0

Brady, C. L., Venter, S. N., Cleenwerck, I., Vandemeulebroecke, K., De Vos, P., and Coutinho, T. A. (2010b). Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. and description of *Tatumella morbirosei* sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 484–494. doi: 10.1099/ijs.0.012070-0

Brady, C., Cleenwerck, I., Venter, S., Coutinho, T., and De Vos, P. (2013). Taxonomic evaluation of the genus Enterobacter based on multilocus sequence analysis (MLSA): proposal to reclassify E. *nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into

Pluralibacter gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii, E. radicincitans, E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as *Kosakonia cowanii* comb. nov., *Kosakonia radicincitans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis, E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. *Syst. Appl. Microbiol.* 36, 309–319. doi: 10.1016/j.syapm.2013.03.005

Carbone, I., Ramirez-Prado, J. H., Jakobek, J. L., and Horn, B. W. (2007). Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. *BMC Evol. Biol.* 7:111. doi: 10.1186/1471-2148-7-111

Caro-Quintero, A., and Konstantinidis, K. T. (2012). Bacterial species may exist, metagenomics reveal. *Environ. Microbiol.* 14, 347–355. doi: 10.1111/j.1462-2920.2011.02668.x

Cendron, L., Berni, R., Folli, C., Ramazzina, I., Percudani, R., and Zanotti, G. (2007). The structure of 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline decarboxylase provides insights into the mechanism of uric acid degradation. *J. Biol. Chem.* 282, 18182–18189. doi: 10.1074/jbc.M701297200

Chan, J. Z. M., Halachev, M. R., Loman, N. J., Constantinidou, C., and Pallen, M. J. (2012). Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. *BMC Microbiol.* 12:302. doi: 10.1186/1471-2180-12-302

Chandra, G., Chater, K. F., and Bornemann, S. (2011). Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. *Microbiology* 157, 1565–1572. doi: 10.1099/mic.0.044263-0

Cohan, F. M. (2001). Bacterial species and speciation. *Syst. Biol.* 50, 513–524. doi: 10.1080/10635150118398

Cohan, F. M. (2002). What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487. doi: 10.1146/annurev.micro.56.012302.160634

Cohan, F. M. (2005). "Periodic selection and ecological diversity in bacteria," in *Selective Sweep*, ed D. Nurminsky (Georgetown, TX: Springer), 78–93.

Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* 28, 1481–1489. doi: 10.1093/molbev/msq333

Colloc'h, N., Hajji, M. E., Bachet, B., L'Hermite, G., Schiltz, M., Prange, T., et al. (1997). Crystal structure of the protein drug urate oxidase-inhibitor complex at 2.05 A resolution. *Nat. Struct. Mol. Biol.* 4, 947–952. doi: 10.1038/nsb1197-947

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Coutinho, T. A., and Venter, S. N. (2009). Pantoea ananatis: an unconventional plant pathogen. *Mol. Plant Pathol.* 10, 325–335. doi: 10.1111/j.1364-3703.2009.00542.x

Cruz, A. T., Cazacu, A. C., and Allen, C. H. (2007). *Pantoea agglomerans*, a plant pathogen causing human disease. *J. Clin. Microbiol.* 45, 1989–1992. doi: 10.1128/JCM.00632-07

Dahl, U., Jaeger, T., Nguyen, B. T., Sattler, J. M., and Mayer, C. (2004). Identification of a phosphotransferase system of *Escherichia coli* required for growth on N-Acetylmuramic acid. *J. Bacteriol.* 186, 2385–2392. doi: 10.1128/JB.186.8.2385-2392.2004

Daubin, V., Gouy, M., and Perrière, G. (2002). A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 12, 1080–1090. doi: 10.1101/gr.187002

De Baere, T., Verhelst, R., Labit, C., Verschraegen, G., Wauters, G., Claeys, G., et al. (2004). Bacteremic infection with *Pantoea ananatis*. *J. Clin. Microbiol.* 42, 4393–4395. doi: 10.1128/JCM.42.9.4393-4395.2004

De Maayer, P., Chan, W. Y., Rezzonico, F., Bühlmann, A., Venter, S. N., Blom, J., et al. (2012a). Complete genome sequence of clinical isolate *Pantoea ananatis* LMG5342. *J. Bacteriol.* 194, 1615–1616. doi: 10.1128/JB.06715-11.

De Maayer, P., Chan, W.-Y., Blom, J., Venter, S. N., Duffy, B., Smits, T. H. M., et al. (2012b). The large universal *Pantoea plasmid* LPP-1 plays a major role in biological and ecological diversification. *BMC Genomics* 13:625. doi: 10.1186/1471-2164-13-625

De Maayer, P., Chan, W.-Y., Rubagotti, E., Venter, S. N., Toth, I. K., Birch, P. R. J., et al. (2014). Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. *BMC Genomics* 15:404. doi: 10.1186/1471-2164-15-404

Deeg, R., Roder, A., Siedel, J., Gauhl, H., Ziegenhorn, J., and Boehringer Mannheim Gmbh (1987). *Process and Reagent for the Determination of N-Carbamoylsarcosine with the Use of a New Enzyme*. U.S. Patent 4,645,739.

Fang, G., Rocha, E. P., and Danchin, A. (2008). Persistence drives gene clustering in bacterial genomes. *BMC Genomics* 9:4. doi: 10.1186/1471-2164-9-4

Fecteau, G., Fairbrother, J. M., Higgins, R., Van Metre, D. C., Par,é, J., Smith, B. P., et al. (2001). Virulence factors in *Escherichia coli* isolated from the blood of bacteremic neonatal calves. *Vet. Microbiol.* 78, 241–249. doi: 10.1016/S0378-1135(00)00299-6

Fiedler, H.-P., Krastel, P., Müller, J., Gebhardt, K., and Zeeck, A. (2001). Enterobactin: the characteristic catecholate siderophore of Enterobacteriaceae is produced by *Streptomyces* species. *FEMS Microbiol. Lett.* 196, 147–151. doi: 10.1111/j.1574-6968.2001.tb10556.x

Fouet, A., Klier, A., and Rapoport, G. (1986). Nucleotide sequence of the sucrase gene of *Bacillus subtilis*. *Gene* 45, 221–225. doi: 10.1016/0378-1119(86)90258-1

Fujita, Y., Matsuoka, H., and Hirooka, K. (2007). Regulation of fatty acid metabolism in bacteria. *Mol. Microbiol.* 66, 829–839. doi: 10.1111/j.1365-2958.2007.05947.x

Gao, Q., Wang, X., Xu, H., Xu, Y., Ling, J., Zhang, D., et al. (2012). Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol.* 12:143. doi: 10.1186/1471-2180-12-143

Garcia, E. C., Brumbaugh, A. R., and Mobley, H. L. (2011). Redundancy and specificity of *Escherichia coli* iron acquisition systems during urinary tract infection. *Infect. Immun.* 79, 1225–1235. doi: 10.1128/IAI.01222-10

Garrity, G. M., Bell, J. A., and Lilburn, T. (2005). "The revised road map to the *manual*," in *Bergey's Manual® of Systematic Bacteriology*, eds D. J. Brenner, N. R. Krieg, J. R. Staley, and G. M. Garrity (Boston, MA: Springer), 159–187.

Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K., et al. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. *nov.* as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int. J. Syst. Bacteriol.* 39, 337–345. doi: 10.1099/00207713-39-3-337

Geddy, R., and Brown, G. G. (2007). Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* 8:30. doi: 10.1186/1471-2164-8-130

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Re-evaluating prokaryotic species. *Nature Rev.* 3, 733–739. doi: 10.1038/nrmicro1236

Giorgelli, F., Bottai, C., Mascia, L., Scolozzi, C., Camici, M., and Ipata, P. L. (1997). Recycling of α-d-ribose 1-phosphate for nucleoside interconversion. *Biochim. Biophys. Acta* 1335, 16–22. doi: 10.1016/S0304-4165(96)00117-1

Glaeser, S. P., and Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst. Appl. Microbiol.* 38, 237–245. doi: 10.1016/j.syapm.2015.03.007

Gogarten, J. P., Doolittle, W. F., and Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238. doi: 10.1093/oxfordjournals.molbev.a004046

Gossmann, T. I., Ziegler, M., Puntervoll, P., de Figueiredo, L. F., Schuster, S., and Heiland, I. (2012). NAD+ biosynthesis and salvage–a phylogenetic perspective. *FEBS J.* 279, 3355–3363. doi: 10.1111/j.1742-4658.2012.08559.x

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Gowrishankar, J. (1989). Nucleotide sequence of the osmoregulatory proU operon of *Escherichia coli*. *J. Bacteriol.* 171, 1923–1931. doi: 10.1128/jb.171.4.1923-1931.1989

Graninger, M., Nidetzky, B., Heinrichs, D. E., Whitfield, C., and Messner, P. (1999). Characterization of dTDP-4-dehydrorhamnose 3,5-Epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-l-rhamnose biosynthesis in *Salmonella enterica* S*erovar Typhimurium* LT2. *J. Biol. Chem.* 274, 25069–25077. doi: 10.1074/jbc.274.35.25069

Greenfield, L. K., and Whitfield, C. (2012). Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr. Res.* 356, 12–24. doi: 10.1016/j.carres.2012.02.027

Gueule, D., Fourny, G., Ageron, E., Le Flèche-Matéos, A., Vandenbogaert, M., Grimont, P. A. D., et al. (2015). *Pantoea coffeiphila* sp. nov., cause of the 'potato taste' of Arabica coffee from the African Great Lakes region. *Int. J. Syst. Evol. Microbiol.* 65, 23–29. doi: 10.1099/ijs.0.063545-0

Gugliucci, A., and Menini, T. (2003). The polyamines spermine and spermidine protect proteins from structural and functional damage by AGE precursors: a new role for old molecules? *Life Sci.* 72, 2603–2616. doi: 10.1016/S0024-3205(03)00166-8

Ha, H. C., Sirisoma, N. S., Kuppusamy, P., Zweier, J. L., Woster, P. M., and Casero, R. A. (1998). The natural polyamine spermine functions directly as a free radical scavenger. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11140–11145. doi: 10.1073/pnas.95.19.11140

Hacker, J., and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2, 376–381. doi: 10.1093/embo-reports/kve097

Hall, B. G., and Xu, L. (1992). Nucleotide sequence, function, activation, and evolution of the cryptic asc operon of *Escherichia coli* K12. *Mol. Biol. Evol.* 9, 688–706.

Hématy, K., Cherk, C., and Somerville, S. (2009). Host–pathogen warfare at the plant cell wall. *Curr. Opin. Plant Biol.* 12, 406–413. doi: 10.1016/j.pbi.2009.06.007

Hillis, D. M., and Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66, 411–453. doi: 10.1086/417338

Hong, K.-W., Gan, H. M., Low, S.-M., Lee, P. K. Y., Chong, Y.-M., Yin, W.-F., et al. (2012). Draft genome sequence of *Pantoea* sp. strain A4, a Rafflesia-associated bacterium that produces N-acylhomoserine lactones as quorum-sensing molecules. *J. Bacteriol.* 194:6610. doi: 10.1128/JB.01619-12

Hosie, A. H. F., and Poole, P. S. (2001). Bacterial ABC transporters of amino acids. *Res. Microbiol.* 152, 259–270. doi: 10.1016/S0923-2508(01)01197-4

Hubertus, H., Schoeser, M., Lesuisse, E., Ernst, J. F., Parson, W., Beate, A., et al. (2003). Characterization of the *Aspergillus nidulans* transporters for the siderophores enterobactin and triacetylfusarinine C. *Biochem. J.* 371, 505–513. doi: 10.1042/bj20021685

Jaeger, T., and Mayer, C. (2008). N-acetylmuramic acid 6-phosphate lyases (MurNAc etherases): role in cell wall metabolism, distribution, structure, and mechanism. *Cell. Mol. Life Sci.* 65, 928–939. doi: 10.1007/s00018-007-7399-x

Javaux, C., Joris, B., and De Witte, P. (2007). Functional Characteristics of TauA binding protein from TauABC *Escherichia coli* system. *Protein J.* 26, 231–238. doi: 10.1007/s10930-006-9064-x

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kamber, T., Smits, T. H. M., Rezzonico, F., and Duffy, B. (2012). Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. *Trees* 26, 227–238. doi: 10.1007/s00468-011-0619-x

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. doi: 10.1093/nar/30.1.42

Kanehisa, M., Sato, Y., and Morishima, K. (2016a). 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

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016b). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kappes, R. M., Kempf, B., Kneip, S., Boch, J., Gade, J., Meier-Wagner, J., et al. (1999). Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Mol. Microbiol.* 32, 203–216. doi: 10.1046/j.1365-2958.1999.01354.x

Katoh, A., Uenohara, K., Akita, M., and Hashimoto, T. (2006). Early steps in the biosynthesis of NAD in arabidopsis start with aspartate and occur in the plastid. *Plant Physiol.* 141, 851–857. doi: 10.1104/pp.106.081091

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

Kehres, D., and Hogg, R. (1992). *Escherichia coli* K12 arabinose-binding protein mutants with altered transport properties. *Protein Sci.* 1, 1652–1660. doi: 10.1002/pro.5560011213

Kempf, B., and Bremer, E. (1995). OpuA, an osmotically regulated binding protein-dependent transport system for the osmoprotectant glycine betaine in *Bacillus subtilis*. *J. Biol. Chem.* 270, 16701–16713. doi: 10.1074/jbc.270.28.16701

Khan, N. A., Quemener, V., and Moulinoux, J.-P. (1992). Phorbol esters augment polyamine transport by influencing Na^+-K^+ pump in murine leukemia cells. *Exp. Cell Res.* 199, 378–382. doi: 10.1016/0014-4827(92)90448-H

Kimiyoshi, I., Yoshihiro, A., Kumi, N., Shinsei, M., Tatsuo, H., Osamu, S., et al. (1993). Cloning of the cDNA encoding human xanthine dehydrogenase (oxidase): structural analysis of the protein and chromosomal location of the gene. *Gene* 133, 279–284. doi: 10.1016/0378-1119(93)90652-J

Ko, R., and Smith, L. T. (1999). Identification of an ATP-Driven, osmoregulated glycine betaine transport system in *Listeria monocytogenes*. *Appl. Environ. Microbiol.* 65, 4040–4048.

Kohchi, C., Inagawa, H., Nishizawa, T., Yamaguchi, T., Nagai, S., and Soma, G.-I. (2006). Applications of lipopolysaccharide derived from *Pantoea agglomerans* (IP-PA1) for health care based on macrophage network theory. *J. Biosci. Bioeng.* 102, 485–496. doi: 10.1263/jbb.102.485

Konstantinidis, K. T., and Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Konstantinidis, K. T., and Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264. doi: 10.1128/JB.187.18.6258-6264.2005

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504–509. doi: 10.1016/j.mib.2007.08.006

Koskiniemi, S., Sun, S., Berg, O. G., and Andersson, D. I. (2012). Selection-driven gene loss in bacteria. *PLoS Genet.* 8:e1002787. doi: 10.1371/journal.pgen.1002787

Krismer, B., Liebeke, M., Janek, D., Nega, M., Rautenberg, M., Hornig, G., et al. (2014). Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. *PLoS Pathog.* 10:e1003862. doi: 10.1371/journal.ppat.1003862

Linghu, B., Snitkin, E. S., Holloway, D. T., Gustafson, A. M., Xia, Y., and DeLisi, C. (2008). High-precision high-coverage functional inference from integrated data sources. *BMC Bioinformatics* 9:119. doi: 10.1186/1471-2105-9-119

Liu, T., Vora, H., and Khosla, C. (2010). Quantitative analysis and engineering of fatty acid biosynthesis in E. *coli*. *Metab. Eng.* 12, 378–386. doi: 10.1016/j.ymben.2010.02.003

Ma, Y., Yin, Y., Rong, C., Chen, S., Liu, Y., Wang, S., et al. (2016). *Pantoea pleuroti* sp. nov., Isolated from the fruiting bodies of *Pleurotus eryngii*. *Curr. Microbiol.* 72, 207–212. doi: 10.1007/s00284-015-0940-5

Maciejewska, A. M., Ruszel, K. P., Nieminuszczy, J., Lewicka, J., Sokołowska, B., Grzesiuk, E., et al. (2010). Chloroacetaldehyde-induced mutagenesis in *Escherichia coli*: the role of AlkB protein in repair of 3, N 4-ethenocytosine and 3, N 4-α-hydroxyethanocytosine. *Mutat. Res. Fundamental Mol. Mech. Mutagenesis* 684, 24–34. doi: 10.1016/j.mrfmmm.2009.11.005

McInerney, J. O., Pisani, D., Bapteste, E., and O'Connell, M. J. (2011). The public goods hypothesis for the evolution of life on earth. *Biol. Direct* 6:41. doi: 10.1186/1745-6150-6-41

Mehler, A. H., and Tabor, H. (1953). Deamination of histidine to form urocanic acid In liver. *J. Biol. Chem.* 201, 775–784.

Mergaert, J., Verdonck, L., and Kersters, K. (1993). Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea emend.* as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. indologenes subsp. nov. *Int. J. Syst. Bacteriol.* 43, 162–173. doi: 10.1099/00207713-43-1-162

Metlitskaya, A., Kazakov, T., Vondenhoff, G. H., Novikova, M., Shashkov, A., Zatsepin, T., et al. (2009). Maturation of the translation inhibitor Microcin C. *J. Bacteriol.* 191, 2380–2387. doi: 10.1128/JB.00999-08

Moffatt, B. A., and Ashihara, H. (2002). Purine and pyrimidine nucleotide synthesis and metabolism. *Arabidopsis Book* 1:e0018. doi: 10.1199/tab.0018

Montgomerie, J., Bindereif, A., Neilands, J., Kalmanson, G., and Guze, L. (1984). Association of hydroxamate siderophore (aerobactin) with Escherichia coli isolated from patients with bacteremia. *Infect. Immun.* 46, 835–838.

Moralejo, P., Egan, S., Hidalgo, E., and Aguilar, J. (1993). Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. *J. Bacteriol.* 175, 5585–5594. doi: 10.1128/jb.175.17.5585-5594.1993

Nohno, T., Saito, T., and Hong, J.-S. (1986). Cloning and complete nucleotide sequence of the *Escherichia coli* glutamine permease operon (glnHPQ). *Mol. Gen. Genet.* 205, 260–269. doi: 10.1007/BF00430437

Novikova, M., Metlitskaya, A., Datsenko, K., Kazakov, T., Kazakov, A., Wanner, B., et al. (2007). The *Escherichia coli* Yej transporter is required for the uptake of translation inhibitor Microcin C. *J. Bacteriol.* 189, 8361–8365. doi: 10.1128/JB.01028-07

Nygaard, P., Bested, S. M., Andersen, K. A. K., and Saxild, H. H. (2000). *Bacillus subtilis* guanine deaminase is encoded by the yknA gene and is induced during growth with purines as the nitrogen source. *Microbiology* 146, 3061–3069. doi: 10.1099/00221287-146-12-3061

Ollagnier-de Choudens, S., Loiseau, L., Sanakis, Y., Barras, F., and Fontecave, M. (2005). Quinolinate synthetase, an iron–sulfur enzyme in NAD biosynthesis. *FEBS Lett.* 579, 3737–3743. doi: 10.1016/j.febslet.2005.05.065

Opal, S. M., Cross, A. S., Gemski, P., and Lyhte, L. W. (1990). Aerobactin and α-hemolysin as virulence determinants in *Escherichia coli* isolated from human blood, urine, and stool. *J. Infect. Dis.* 161, 794–796. doi: 10.1093/infdis/161.4.794

Ormeno-Orrillo, E., Servín-Garcidue-as, L. E., Rogel, M. A., González, V., Peralta, H., Mora, J., et al. (2015). Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. *Syst. Appl. Microbiol.* 38, 287–291. doi: 10.1016/j.syapm.2014.12.002

Palmer, M., de Maayer, P., Poulsen, M., Steenkamp, E. T., van Zyl, E., Coutinho, T. A., et al. (2016). Draft genome sequences of *Pantoea agglomerans* and *Pantoea vagans* isolates associated with termites. *Stand. Genomic Sci.* 11, 1–11. doi: 10.1186/s40793-016-0144-z

Palmer, M., Steenkamp, E. T., Coetzee, M. P. A., Chan, W.-Y., van Zyl, E., De Maayer, P., et al. (2017). Phylogenomic resolution of the bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella*. *Antonie van Leeuwenhoek* 110, 1287–1309. doi: 10.1007/s10482-017-0852-4

Pegg, A. E. (2016). Functions of polyamines in mammals. *J. Biol. Chem.* 291, 14904–14912. doi: 10.1074/jbc.R116.731661

Plumbridge, J. (2009). An alternative route for recycling of N-Acetylglucosamine from peptidoglycan involves the N-Acetylglucosamine phosphotransferase system in *Escherichia coli*. *J. Bacteriol.* 191, 5641–5647. doi: 10.1128/JB.00448-09

Pond, S. L. K., and Muse, S. V. (2005). "HyPhy: hypothesis testing using phylogenies," in *Statistical Methods in Molecular Evolution*, ed R. Nielsen (New York, NY: Springer New York), 125–181.

Rahman, N. A., Parks, D. H., Vanwonterghem, I., Morrison, M., Tyson, G. W., and Hugenholtz, P. (2015). A phylogenomic analysis of the bacterial phylum Fibrobacteres. *Front. Microbiol.* 6:1469. doi: 10.3389/fmicb.2015.01469

Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003). Enterobactin: an archetype for microbial iron transport. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3584–3588. doi: 10.1073/pnas.0630018100

Richard, P., and Hilditch, S. (2009). D-galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Appl. Microbiol. Biotechnol.* 82, 597–604. doi: 10.1007/s00253-009-1870-6

Richards, V. P., Palmer, S. R., Pavinski Bitar, P. D., Qin, X., Weinstock, G. M., Highlander, S. K., et al. (2014). Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol. Evol.* 6, 741–753. doi: 10.1093/gbe/evu048

Rong, C., Ma, Y., Wang, S., Liu, Y., Chen, S., Huang, B., et al. (2016). *Pantoea hericii* sp. nov., isolated from the fruiting bodies of *Hericium erinaceus*. *Curr. Microbiol.* 72, 738–743. doi: 10.1007/s00284-016-1011-2

Saxena, R., Anand, P., Saran, S., Isar, J., and Agarwal, L. (2010). Microbial production and applications of 1, 2-propanediol. *Indian J. Microbiol.* 50, 2–11. doi: 10.1007/s12088-010-0017-x

Schneider, E. (2001). ABC transporters catalyzing carbohydrate uptake. *Res. Microbiol.* 152, 303–310. doi: 10.1016/S0923-2508(01)01201-3

Schubert, R. H. (1968). The taxonomy and nomenclature of the genus Aeromonas Kluyver and van Niel 1936. *Int. J. Syst. Evol. Microbiol.* 18, 1–7.

Schulz, H. (1991). Beta oxidation of fatty acids. *Biochimica Biophys. Acta* 1081, 109–120. doi: 10.1016/0005-2760(91)90015-A

Schwartz, A. R., Potnis, N., Timilsina, S., Wilson, M., Patané, J., Martins, J., et al. (2015). Phylogenomics of Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. *Front. Microbiol.* 6:535. doi: 10.3389/fmicb.2015.00535

Schwede, T. F., Rétey, J., and Schulz, G. E. (1999). Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* 38, 5355–5361. doi: 10.1021/bi982929q

Scripture, J. B., Voelker, C., Miller, S., O'Donnell, R. T., Polgar, L., Rade, J., et al. (1987). High-affinity l-arabinose transport operon. *J. Mol. Biol.* 197, 37–46. doi: 10.1016/0022-2836(87)90607-3

Sekowska, A., Dénervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A., et al. (2004). Bacterial variations on the methionine salvage pathway. *BMC Microbiol.* 4:9. doi: 10.1186/1471-2180-4-9

Sekowska, A., Kung, H.-F., and Danchin, A. (2000). Sulfur metabolism in *Escherichia coli* and related bacteria: facts and fiction. *J. Mol. Microbiol. Biotechnol.* 2, 145–177.

Shah, P., and Swiatlo, E. (2008). A multifaceted role for polyamines in bacterial pathogens. *Mol. Microbiol.* 68, 4–16. doi: 10.1111/j.1365-2958.2008.06126.x

Shimizu, K. (2014). Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. *Metabolites* 4:1. doi: 10.3390/metabo4010001

Sim, L., Quezada-Calvillo, R., Sterchi, E. E., Nichols, B. L., and Rose, D. R. (2008). Human intestinal maltase–glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. *J. Mol. Biol.* 375, 782–792. doi: 10.1016/j.jmb.2007.10.069

Song, Y.-J. (2009). Characterization of aromatic hydrocarbon degrading bacteria isolated from pine litter. *Kor. J. Microbiol. Biotechnol.* 37, 333–339.

Sorrels, C. M., Proteau, P. J., and Gerwick, W. H. (2009). Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl. Environ. Microbiol.* 75, 4861–4869. doi: 10.1128/AEM.02508-08

Sprenger, G. A. (1993). Two open reading frames adjacent to the *Escherichia coli* K-12 transketolase (tkt) gene show high similarity to the mannitol phosphotransferase system enzymes from *Escherichia coli* and various Gram-positive bacteria. *Biochim. Biophys. Acta* 1158, 103–106. doi: 10.1016/0304-4165(93)90103-F

Stanier, R. Y., and Van Niel, C. (1941). The main outlines of bacterial classification. *J. Bacteriol.* 42:437.

Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., et al. (1994). Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its rfb gene cluster. *J. Bacteriol.* 176, 4144–4156. doi: 10.1128/jb.176.13.4144-4156.1994

Stirling, D., Hulton, C., Waddell, L., Park, S., Stewart, G., Booth, I., et al. (1989). Molecular characterization of the proU loci of *Salmonella typhimurium* and *Escherichia coli* encoding osmoregulated glycine betaine transport systems. *Mol. Microbiol.* 3, 1025–1038. doi: 10.1111/j.1365-2958.1989.tb00253.x

Sutcliffe, I. C. (2015). Challenging the anthropocentric emphasis on phenotypic testing in prokaryotic species descriptions: rip it up and start again. *Front. Genet.* 6:218. doi: 10.3389/fgene.2015.00218

Sutcliffe, I. C., Trujillo, M. E., and Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie Van Leeuwenhoek* 101, 13–20. doi: 10.1007/s10482-011-9664-0

Suzuki, M. (1991). *U.S. Patent No. 5,047,329*. Washington, DC: U.S. Patent and Trademark Office.

Suzuki, H., Koyanagi, T., Izuka, S., Onishi, A., and Kumagai, H. (2005). The yliA, -B, -C, and -D genes of *Escherichia coli* K-12 encode a novel glutathione importer with an ATP-binding cassette. *J. Bacteriol.* 187, 5861–5867. doi: 10.1128/JB.187.17.5861-5867.2005

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

Tindall, B. J., Rosselló-Mora, R., Busse, H.-J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0

Torres, A. G., Redford, P., Welch, R. A., and Payne, S. M. (2001). TonB-dependent systems of uropathogenic *Escherichia coli*: aerobactin and heme transport and

TonB are required for virulence in the mouse. *Infect. Immun.* 69, 6179–6185. doi: 10.1128/IAI.69.10.6179-6185.2001

Tracz, D. M., Gilmour, M. W., Mabon, P., Beniac, D. R., Hoang, L., Kibsey, P., et al. (2015). *Tatumella saanichensis* sp. nov., isolated from a cystic fibrosis patient. *Int. J. Syst. Evolut. Microbiol.* 65, 1959–1966. doi: 10.1099/ijs.0.000207

Uehara, T., Suefuji, K., Jaeger, T., Mayer, C., and Park, J. T. (2006). MurQ etherase is required by *Escherichia coli* in order to metabolize anhydro-N-acetylmuramic acid obtained either from the environment or from its own cell wall. *J. Bacteriol.* 188, 1660–1662. doi: 10.1128/JB.188.4.1660-1662.2006

Van Beers, E. H., Büller, H. A., Grand, R. J., Einerhand, A. W. C., and Dekker, J. (1995). Intestinal brush border glycohydrolases: structure, function, and development. *Crit. Rev. Biochem. Mol. Biol.* 30, 197–262. doi: 10.3109/10409239509085143

Van der Ploeg, J., Weiss, M. A., Saller, E., Nashimoto, H., Saito, N., Kertesz, M. A., et al. (1996). Identification of sulfate starvation-regulated genes in *Escherichia coli*: a gene cluster involved in the utilization of taurine as a sulfur source. *J. Bacteriol.* 178, 5438–5446. doi: 10.1128/jb.178.18.5438-5446.1996

Vanneste, J., Cornish, D., Yu, J., and Voyle, M. (2001). "The peptide antibiotic produced by *Pantoea agglomerans* Eh252 is a microcin," in *IX International Workshop on Fire Blight* (Napier), 285–290.

Walshaw, D. L., Lowthorpe, S., East, A., and Poole, P. S. (1997). Distribution of a sub-class of bacterial ABC polar amino acid transporter and identification of an N-terminal region involved in solute specificity. *FEBS Lett.* 414, 397–401. doi: 10.1016/S0014-5793(97)01023-5

Walterson, A. M., and Stavrinides, J. (2015). Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiol. Rev.* 39, 968–984. doi: 10.1093/femsre/fuv027

Walton, J. D. (1994). Deconstructing the cell wall. *Plant Physiol.* 104:1113. doi: 10.1104/pp.104.4.1113

Wan, X., Hou, S., Phan, N., Malone Moss, J. S., Donachie, S. P., and Alam, M. (2015). Draft genome sequence of pantoea anthophila strain 11-2 from hypersaline lake laysan, Hawaii. *Genome Announc.* 3:e00321-15. doi: 10.1128/genomeA.00321-15

Wang, L., and Reeves, P. R. (1998). Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect. Immun.* 66, 3545–3551.

Whitfield, C. (1995). Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol.* 3, 178–185. doi: 10.1016/S0966-842X(00)88917-9

Williams, P. H., and Carbonetti, N. H. (1986). Iron, siderophores, and the pursuit of virulence: independence of the aerobactin and enterochelin iron uptake systems in *Escherichia coli*. *Infect. Immun.* 51, 942–947.

Woese, C. (1998). The universal ancestor. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6854–6859. doi: 10.1073/pnas.95.12.6854

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* 51:221.

Woese, C. R. (1994). There must be a prokaryote somewhere: microbiology's search for itself. *Microbiol. Rev.* 58, 1–9.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87, 4576–4579. doi: 10.1073/pnas.87.12.4576

Yang, Q., Liu, Y., Huang, F., and He, Z.-G. (2011). Physical and functional interaction between d-ribokinase and topoisomerase I has opposite effects on their respective activity in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Arch. Biochem. Biophys.* 512, 135–142. doi: 10.1016/j.abb.2011.05.018

Yang, Z., Savchenko, A., Yakunin, A., Zhang, R., Edwards, A., Arrowsmith, C., et al. (2003). Aspartate dehydrogenase, a novel enzyme identified from structural and functional studies of TM1643. *J. Biol. Chem.* 278, 8804–8808. doi: 10.1074/jbc.M211892200

Yoshida, K.-I., Yamaguchi, M., Ikeda, H., Omae, K., Tsurusaki, K.-I., and Fujita, Y. (2004). The fifth gene of the iol operon of *Bacillus subtilis*, iolE, encodes 2-keto-myo-inositol dehydratase. *Microbiology* 150, 571–580. doi: 10.1099/mic.0.26768-0

Yoshimoto, T., Oka, I., and Tsuru, D. (1976). Purification, crystallization, and some properties of creatine amidinohydrolase from *Pseudomonus putida*. *J. Biochem.* 79, 1381–1383. doi: 10.1093/oxfordjournals.jbchem.a131193

Young Kim, M., Zhou, X., Delaney, J. C., Taghizadeh, K., Dedon, P. C., Essigmann, J. M., et al. (2007). AlkB influences the chloroacetaldehyde-induced mutation spectra and toxicity in the pSP189 supF shuttle vector. *Chem. Res. Toxicol.* 20, 1075–1083. doi: 10.1021/tx700167v

Zaneveld, J. R., Nemergut, D. R., and Knight, R. (2008). Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* 154, 1–15. doi: 10.1099/mic.0.2007/011833-0

Zangoui, P., Vashishtha, K., and Mahadevan, S. (2015). Evolution of aromatic β-glucoside utilization by successive mutational steps in *Escherichia coli*. *J. Bacteriol.* 197, 710–716. doi: 10.1128/JB.02185-14

Zhang, Z.-G., Ye, Z.-Q., Yu, L., and Shi, P. (2011). Phylogenomic reconstruction of lactic acid bacteria: an update. *BMC Evol. Biol.* 11:1. doi: 10.1186/1471-2148-11-1

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 Palmer, Steenkamp, Coetzee, Blom and Venter. 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: 01 May 2018
doi: 10.3389/fmicb.2018.00771

GET_PHYLOMARKERS, a Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus *Stenotrophomonas*

Pablo Vinuesa[1]*, *Luz E. Ochoa-Sánchez*[1] and *Bruno Contreras-Moreira*[2,3†]

[1] Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, [2] Estación Experimental de Aula Dei – Consejo Superior de Investigaciones Científicas, Zaragoza, Spain, [3] Fundación Agencia Aragonesa para la Investigacion y el Desarrollo (ARAID), Zaragoza, Spain

OPEN ACCESS

Edited by:
Jesus L. Romalde,
Universidade de Santiago de Compostela, Spain

Reviewed by:
Francisco Jose Roig,
Universitat de València, Spain
Anne-Kristin Kaster,
Karlsruher Institut für Technologie, Germany

***Correspondence:**
Pablo Vinuesa
vinuesa@ccg.unam.mx
orcid.org/0000-0001-6119-2956

†Bruno Contreras-Moreira
orcid.org/0000-0002-5462-907X

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

Received: 16 January 2018
Accepted: 05 April 2018
Published: 01 May 2018

Citation:
Vinuesa P, Ochoa-Sánchez LE and Contreras-Moreira B (2018) GET_PHYLOMARKERS, a Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus Stenotrophomonas. Front. Microbiol. 9:771. doi: 10.3389/fmicb.2018.00771

The massive accumulation of genome-sequences in public databases promoted the proliferation of genome-level phylogenetic analyses in many areas of biological research. However, due to diverse evolutionary and genetic processes, many loci have undesirable properties for phylogenetic reconstruction. These, if undetected, can result in erroneous or biased estimates, particularly when estimating species trees from concatenated datasets. To deal with these problems, we developed GET_PHYLOMARKERS, a pipeline designed to identify high-quality markers to estimate robust genome phylogenies from the orthologous clusters, or the pan-genome matrix (PGM), computed by GET_HOMOLOGUES. In the first context, a set of sequential filters are applied to exclude recombinant alignments and those producing anomalous or poorly resolved trees. Multiple sequence alignments and maximum likelihood (ML) phylogenies are computed in parallel on multi-core computers. A ML species tree is estimated from the concatenated set of top-ranking alignments at the DNA or protein levels, using either FastTree or IQ-TREE (IQT). The latter is used by default due to its superior performance revealed in an extensive benchmark analysis. In addition, parsimony and ML phylogenies can be estimated from the PGM. We demonstrate the practical utility of the software by analyzing 170 *Stenotrophomonas* genome sequences available in RefSeq and 10 new complete genomes of Mexican environmental *S. maltophilia* complex (Smc) isolates reported herein. A combination of core-genome and PGM analyses was used to revise the molecular systematics of the genus. An unsupervised learning approach that uses a goodness of clustering statistic identified 20 groups within the Smc at a core-genome average nucleotide identity (cgANIb) of 95.9% that are perfectly consistent with strongly supported clades on the core- and pan-genome trees. In addition, we identified 16 misclassified RefSeq genome sequences, 14 of them labeled as *S. maltophilia*, demonstrating

the broad utility of the software for phylogenomics and geno-taxonomic studies. The code, a detailed manual and tutorials are freely available for Linux/UNIX servers under the GNU GPLv3 license at https://github.com/vinuesa/get_phylomarkers. A docker image bundling GET_PHYLOMARKERS with GET_HOMOLOGUES is available at https://hub.docker.com/r/csicunam/get_homologues/, which can be easily run on any platform.

Keywords: phylogenetics, genome-phylogeny, maximum-likelihood, species-tree, species delimitation, *Stenotrophomonas maltophilia* complex, Mexico

INTRODUCTION

Accurate phylogenies represent key models of descent in modern biological research. They are applied to the study of a broad spectrum of evolutionary topics, ranging from the analysis of populations up to the ecology of communities (Dornburg et al., 2017). The way microbiologists describe and delimit species is undergoing a major revision in the light of genomics (Vandamme and Peeters, 2014; Rosselló-Móra and Amann, 2015), as reflected in the emerging field of microbial genomic taxonomy (Konstantinidis and Tiedje, 2007; Thompson et al., 2009, 2013). Current geno-taxonomic practice is largely based on the estimation of (core-)genome phylogenies (Daubin et al., 2002; Lerat et al., 2003; Tettelin et al., 2005; Ciccarelli et al., 2006; Wu and Eisen, 2008) and the computation of diverse overall genome relatedness indices (OGRIs) (Chun and Rainey, 2014), such as the popular genomic average nucleotide identity (gANI) values (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009). These indices are rapidly and effectively replacing the traditional DNA-DNA hybridization values used for species delimitation in the pre-genomic era (Stackebrandt and Goebel, 1994; Vandamme et al., 1996; Stackebrandt et al., 2002).

The ever-increasing volume of genome sequences accumulating in public sequence repositories provides a huge volume of data for phylogenetic analysis. This significantly improves our capacity to understand the evolution of species and any associated traits (Dornburg et al., 2017). However, due to diverse evolutionary forces and processes, many loci in genomes have undesirable properties for phylogenetic reconstruction. If undetected, these can lead to erroneous or biased estimates (Shen et al., 2017; Parks et al., 2018), although, ironically, with strong branch support (Kumar et al., 2012). Their impact is particularly strong in concatenated datasets (Kubatko and Degnan, 2007; Degnan and Rosenberg, 2009), which are standard in microbial phylogenomics (Wu and Eisen, 2008). Hence, robust phylogenomic inference requires the selection of well-suited markers for the task (Vinuesa, 2010).

For this study we developed GET_PHYLOMARKERS, an open-source and easy-to-use software package designed with the aim of inferring robust genome-level phylogenies and providing tools for microbial genome taxonomy. We describe the implementation details of the pipeline and how it integrates with GET_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013; Vinuesa and Contreras-Moreira, 2015). The latter is a popular and versatile genome-analysis software package designed to identify robust clusters of homologous sequences. It has been widely used in microbial pan-genomics and comparative genomics (Lira et al., 2017; Nourdin-Galindo et al., 2017; Savory et al., 2017; Sandner-Miranda et al., 2018), including recent bacterial geno-taxonomic (Gauthier et al., 2017; Gomila et al., 2017), and plant pan-genomic studies (Contreras-Moreira et al., 2017; Gordon et al., 2017). Regularly updated auxiliary scripts bundled in the GET_HOMOLOGUES package compute diverse OGRIs, at the protein, CDS and transcript levels, provide graphical and statistical tools for a range of pan-genome analyses, including inference of pan-genome phylogenies under the parsimony criterion. GET_PHYLOMARKERS was designed to work both at the core-genome and pan-genome levels, using either the homologous gene clusters or the pan-genome matrix (PGM) computed by GET_HOMOLOGUES. In the first context, it identifies single-copy orthologous gene families with optimal attributes (listed further down) and concatenates them to estimate a genomic species tree. In the second scenario, it uses the PGM to estimate phylogenies under the maximum likelihood (ML) or parsimony optimality criteria. In addition, we implemented unsupervised learning methods that automatically identify species-like genome clusters based on the statistical analysis of the PGM and core-genome average nucleotide identity matrices (cgANIb).

To demonstrate these capabilities and benchmark performance, we applied the pipeline to critically evaluate the molecular systematics and taxonomy of the genus *Stenotrophomonas.* Species delimitation is problematic and far from resolved in this genus (Ochoa-Sánchez and Vinuesa, 2017), despite recent efforts using genomic approaches with a limited number of genome sequences (Patil et al., 2016; Yu et al., 2016; Lira et al., 2017).

The genus *Stenotrophomonas* (Gammaproteobacteria, *Xhanthomonadales*, *Xanthomonadaceae*) (Palleroni and Bradbury, 1993; Palleroni, 2005) groups ubiquitous, aerobic, non-fermenting bacteria that thrive in diverse aquatic and edaphic habitats, including human-impacted ecosystems (Ryan et al., 2009). As of March 2018, 14 validly described species were listed in Jean Euzeby's list of prokaryotic names with standing in nomenclature (http://www.bacterio.net/stenotrophomonas.html). By far, its best-known species is *S. maltophilia.* It is considered a globally emerging, multidrug-resistant (MDR) and opportunistic pathogen (Brooke, 2012; Chang et al., 2015). *S. maltophilia*-like organisms display high genetic, ecological and phenotypic diversity (Valdezate et al., 2004; Vasileuskaya-Schulz et al., 2011), forming the so-called *S. maltophilia* complex (Smc) (Svensson-Stadler et al., 2012; Berg

and Martinez, 2015). Heterogeneous resistance and virulence phenotypes have been reported for environmental isolates of diverse ecological origin classified as *S. maltophilia* (Adamek et al., 2011; Deredjian et al., 2016). We have recently shown that this phenotypic heterogeneity largely results from problems in species delimitations within the Smc (Ochoa-Sánchez and Vinuesa, 2017). We analyzed the genetic diversity of a collection of 108 *Stenotrophomonas* isolates recovered from several water bodies in Morelos, Central Mexico, based on sequence data generated for the 7 loci used in the Multilocus Sequence Typing (MLST) scheme available for *S. maltophilia* at https://pubmlst.org. We assembled a large set of reference sequences retrieved from the MLST database (Kaiser et al., 2009; Vasileuskaya-Schulz et al., 2011) and from selected genome sequences (Crossman et al., 2008; Lira et al., 2012; Davenport et al., 2014; Vinuesa and Ochoa-Sánchez, 2015; Patil et al., 2016), encompassing 11 out of the 12 validly described species at the time. State-of-the-art phylogenetic and population genetics methods, including the multispecies coalescent model coupled with Bayes factor analysis and Bayesian clustering of the multilocus genotypes consistently resolved five conservatively-defined genospecies within the Smc clade, which were named *S. maltophilia* and Smc1-Smc4. The approach also delimited Smc5 as a sister clade of *S. rhizophila*. Importantly, we showed that (*i*) only members of the Smc clade that we designed as *S. maltophilia* were truly MDR and (*ii*) that *S. maltophilia* was the only species that consistently expressed metallo-beta-lactamases (Ochoa-Sánchez and Vinuesa, 2017). Strains of the genospecies Smc1 and Smc2 were only recovered from the Mexican rivers and displayed significantly lower resistance levels than sympatric *S. maltophilia* isolates, revealing well-defined species-specific phenotypes.

Given this context, the present study was designed with two major goals. The first one was to develop GET_PHYLOMARKERS, a pipeline for the automatic and robust estimation of genome phylogenies using state-of-the art methods. The emphasis of the pipeline is on selecting top-ranking markers for the task, based on the following quantitative/statistical criteria: (*i*) they should not present signs of recombination, (*ii*) the resulting gene trees should not be anomalous or deviating from the distribution of tree topologies and branch lengths expected under the multispecies coalescent model, and (*iii*) they should have a strong phylogenetic signal. The top-scoring markers are concatenated to estimate the species phylogeny under the ML optimality criterion using either FastTree (Price et al., 2010) or IQ-TREE (IQT) (Nguyen et al., 2015). The second aim was to apply GET_PHYLOMARKERS to challenge and refine the species delimitations reported in our previous MLSA study (Ochoa-Sánchez and Vinuesa, 2017) using a genomic approach, focusing on resolving the geno-taxonomic structure of the Smc and *S. maltophilia sensu lato* clades. For this purpose we sequenced five strains from the new genospecies Smc1 and Smc2 and analyzed them together with all reference genome sequences available for the genus *Stenotrophomonas* as of August 2017 using the methods implemented in GET_PHYLOMARKERS. The results were used to critically revise the molecular systematics of the genus in light of genomics, identify misclassified genome sequences, suggest correct classifications for them and discover multiple novel genospecies within *S. maltophilia*.

MATERIALS AND METHODS

Genome Sequencing, Assembly, and Annotation

Ten *Stenotrophomonas* strains from our collection were selected (**Table 1**) for genome sequencing using a MiSeq instrument (2×300 bp) at the Genomics Core Sequencing Service provided by Arizona State University (DNASU). They were all isolated from rivers in the state of Morelos, Central Mexico, and classified as genospecies 1 (Smc1) or 2 (Smc2), as detailed in a previous publication (Ochoa-Sánchez and Vinuesa, 2017). Adaptors at the 5′-ends and low quality residues at the 3′ ends of reads were trimmed-off using ngsShoRT v2.1 (Chen et al., 2014) and passed to Spades v3.10.1 (Bankevich et al., 2012) for assembly (with options –careful -k 33,55,77,99,127,151). The resulting assembly scaffolds were filtered to remove those with low coverage (<7X) and short length (< 500 nt). All complete genome sequences available in RefSeq for *Stenotrophomonas* spp. were used as references for automated ordering of assembly scaffolds using MeDuSa v1.6 (Bosi et al., 2015). A final assembly polishing step was performed by remapping the quality-filtered sequence reads on the ordered scaffolds using BWA (Li and Durbin, 2009) and passing the resulting sorted binary alignments to SAMtools (Li et al., 2009) for indexing. The indexed alignments were used by Pilon 1.21 (Walker et al., 2014) for gap closure and filling, correction of indels and single nucleotide polymorphisms (SNPs), as previously described (Vinuesa and Ochoa-Sánchez, 2015). The polished assemblies were annotated with NCBI's Prokaryotic Genome Annotation Pipeline (PGAP v4.2) (Angiuoli et al., 2008). BioProject and BioSample accession numbers are provided in Table S1.

Reference Genomes

On August 1st, 2017, a total of 169 annotated *Stenotrophomonas* genome sequences were available in RefSeq, 134 of which were labeled as *S. maltophilia*. The corresponding GenBank files were retrieved, as well as the corresponding table with assembly metadata. Seven complete *Xanthomonas* spp. genomes were also downloaded to use them as outgroup sequences. In January 2018, the genome sequence of *S. bentonitica* strain VV6 was added to RefSeq and included in the revised version of this work to increase the taxon sampling.

Computing Consensus Core- and Pan-Genomes With GET_HOMOLOGUES

We used GET_HOMOLOGUES (v05022018) (Contreras-Moreira and Vinuesa, 2013) to compute clusters of homologous gene families from the input genome sequences, as previously detailed (Vinuesa and Contreras-Moreira, 2015). Briefly, the source GenBank-formatted files were passed to get_homologues.pl and instructed to compute homologous gene clusters by running either our heuristic (fast) implementation of the bidirectional best-hit (BDBH) algorithm ("-b") to explore the complete dataset, or the full BDBH, Clusters of Orthologous

TABLE 1 | Overview of key annotation features for the 10 new genome assemblies reported in this study for environmental isolates recovered from Mexican rivers and classified as genospecies 1 (Smc1) and 2 (Smc2) in the study of Ochoa-Sánchez and Vinuesa (2017).

Genome	Size_nt	CDSs (coding)	rRNAs	tRNAs	Pseudo-genes	RefSeq Acc. num.
Stenotrophomonas genospecies 1 (Smc1; Sgn1) ESTM1D MKCIP4 1	4,475,880	3,904	6	59	67	CP026004
Stenotrophomonas genospecies 1 (Smc1; Sgn1) SAU14A NAIMI4 5	4,570,883	4,020	6	69	66	CP026003
Stenotrophomonas genospecies 1 (Smc1; Sgn1) ZAC14A NAIMI4 1	4,698,328	4,150	7	45	66	CP026002
Stenotrophomonas genospecies 1 (Smc1; Sgn1) ZAC14D1 NAIMI4 1	4,702,461	4,131	6	42	66	CP026001
Stenotrophomonas genospecies 1 (Smc1; Sgn1) ZAC14D1 NAIMI4 6	4,700,343	4,128	6	45	63	CP026000
Stenotrophomonas genospecies 2 (Smc2; Sgn2) SAU14A NAIMI4 8	4,479,100	3,893	5	54	69	CP025999
Stenotrophomonas genospecies 2 (Smc2; Sgn2) YAU14A MKIMI4 1	4,487,007	3,918	7	43	67	CP025998
Stenotrophomonas genospecies 2 (Smc2; Sgn2) YAU14D1 LEIMI4 1	4,319,112	3,819	6	51	66	CP025997
Stenotrophomonas genospecies 2 (Smc2; Sgn2) ZAC14D2 NAIMI4 6	4,431,104	3,882	6	52	66	CP025996
Stenotrophomonas genospecies 2 (Smc2; Sgn2) ZAC14D2 NAIMI4 7	4,468,731	3,918	6	66	62	CP025995

Details of their isolation sites and antimicrobial resistance phenotypes can be found therein. All genomes consist of a single gapped chromosome. Table S1 provides additional information of the assemblies. Their phylogenetic placement within the Stenotrophomonas maltophilia complex is shown in ***Figure 5*** *(clades Sgn1/Smc1 and Sgn2/Smc2).*

Groups—triangles (COGtriangles), and OrthoMCL (Markov Clustering of orthologs, OMCL) algorithms for the different sets of selected genomes, as detailed in the relevant sections and explained in the GET_HOMOLOGUES's online manual (eead-csic-compbio.github.io/get_homologues/manual/manual.html). PFAM-domain scanning was enabled for the latter runs (-D flag). BLASTP hits were filtered by imposing a minimum of 90% alignment coverage (-C 90). The directories holding the results from the different runs were then passed to the auxiliary script compare_clusters.pl to compute either the consensus core genome (-t number_of_genomes) or pan-genome clusters (-t 0). The commands to achieve this can be found in the online tutorial https://vinuesa.github.io/get_phylomarkers/#get_homologues-get_phylomarkers-tutorials provided with the distribution.

Overview of the Computational Steps Performed by the GET_PHYLOMARKERS Pipeline

Figure 1 presents a flow-chart that summarizes the computational steps performed by the pipeline, which are briefly described below. For an in-depth description of each step and associated parameters, as well as for a full version of the pipeline's flow-chart, the reader is referred to the online manual (https://vinuesa.github.io/get_phylomarkers/). The pipeline is primarily intended to run DNA-based phylogenies ("-R 1 -t DNA") on a collection of genomes from different species of the same genus or family. However, it can also select optimal markers for population genetics ("-R 2 -t DNA"), when the source genomes belong to the same species (not shown here). For more divergent genomes, the pipeline should be run using protein sequences ("-R 1 -t PROT"). The analyses are started from the directory holding single-copy core-genome clusters generated either by "get_homologues.pl -e -t number_of_genomes" or by "compare_clusters.pl -t number_of_genomes." Note that both the protein (faa) and nucleotide (fna) FASTA files for the clusters are required, as detailed in the online tutorial (https://vinuesa.github.io/get_phylomarkers/#get_homologues-get_phylomarkers-tutorials). The former are first aligned with clustal-omega (Sievers et al., 2012) and then used by pal2nal (Suyama et al., 2006) to generate codon alignments. These are subsequently scanned with the Phi-test (Bruen et al., 2005) to identify and discard those with significant evidence for recombinant sequences. Maximum-likelihood phylogenies are inferred for each of the non-recombinant alignments using by default IQT v.1.6.3 (Nguyen et al., 2015), which will perform model selection with ModelFinder (Kalyaanamoorthy et al., 2017) and the "-fast" flag enabled for rapid computation, as detailed in the online manual. Alternatively, FastTree v2.1.10 (Price et al., 2010) can be executed using the "-A F" option, which will estimate phylogenies under the GTR+Gamma model. FastTree was compiled with double-precision enabled for maximum accuracy (see the manual for details). The resulting gene trees are screened to detect "outliers" with help of the R package kdetrees (v.0.1.5) (Weyenberg et al., 2014, 2017). It implements a non-parametric test based on the distribution of tree topologies and branch lengths expected under the multispecies coalescent, identifying those phylogenies with unusual topologies or branch lengths. The stringency of the test can be controlled with the *-k* parameter (inter-quartile range multiplier for outlier detection, by default set to the standard 1.5). In a third step, the phylogenetic signal of each gene-tree is computed based on mean branch support values (Vinuesa et al., 2008), keeping only those above a user-defined mean Shimodaira-Hasegawa-like (SH-alrt) bipartition support (Anisimova and Gascuel, 2006) threshold ("-m 0.75" by default). To make all the previous steps as fast as possible, they are run in parallel on multi-core machines using GNU parallel (Tange, 2011). The set of alignments passing all filters are concatenated and subjected to maximum-likelihood (ML) tree searching, using by default IQT with model fitting, to estimate the genomic species-tree.

The complete GET_PHYLOMARKERS pipeline is launched with the master script run_get_phylomarkers_pipeline.sh, which calls a subset of auxiliary Bash, Perl and R programs to perform specific tasks. This architecture allows the user to run the

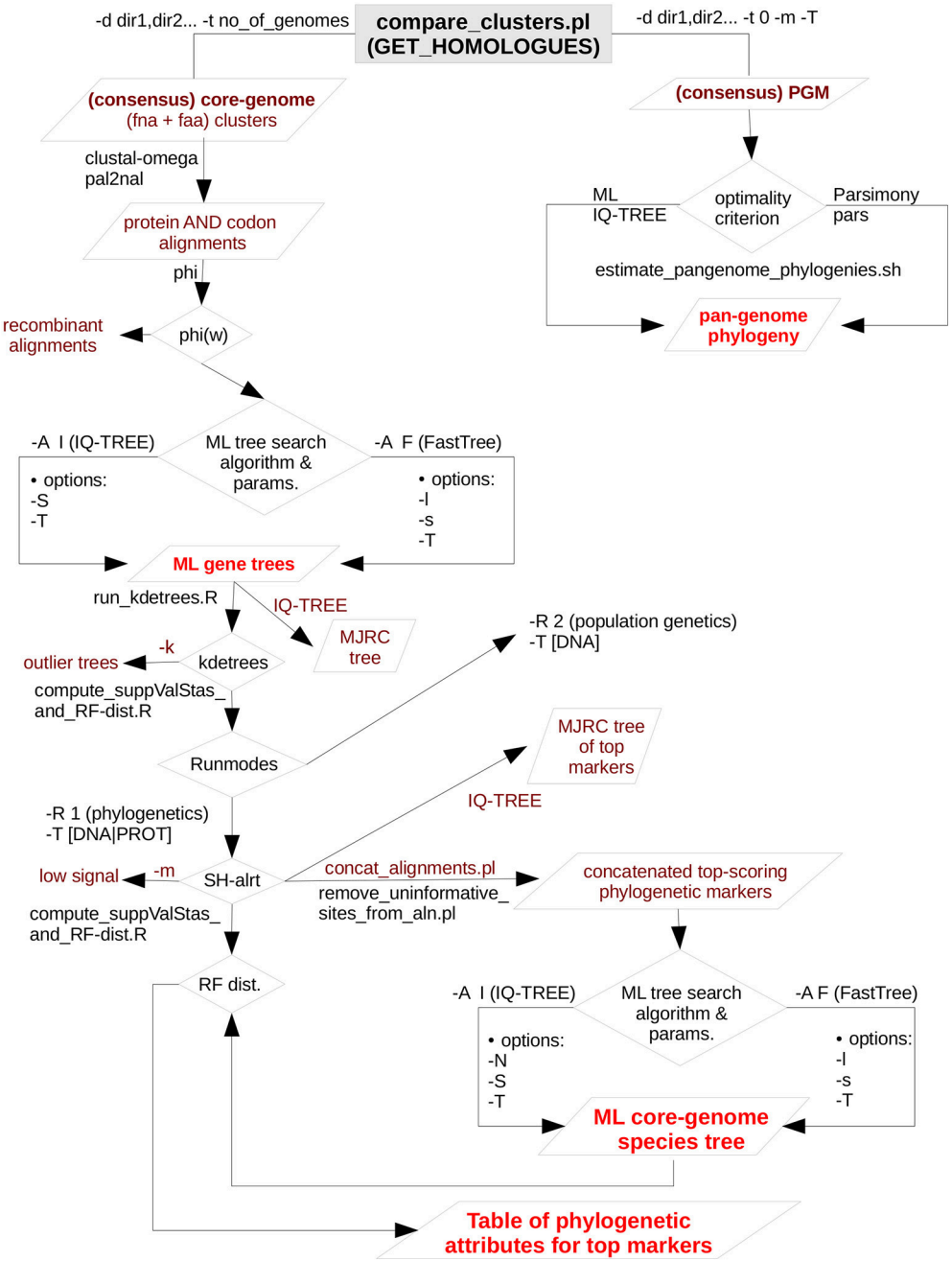


FIGURE 1 | Simplified flow-chart of the GET_PHYLOMARKERS pipeline showing only those parts used and described in this work. The left branch, starting at the top of the diagram, is fully under control of the master script run_get_phylomarkes_pipeline.sh. The names of the worker scripts called by the master program are indicated on the relevant points along the flow. Steps involving repetitive computational processes, like generating multiple sequence alignments or inferring the corresponding gene trees, are run in parallel with the aid of GNU parallel, which is called from run_parallel_cmmds.pl. The right-hand branch, at the top of the diagram, summarizes the analyses that can be performed on the pan-genome matrix (PGM). In this work we only present the estimation of maximum-likelihood and parsimony pan-genome phylogenies. However, unsupervised learning approaches are provided by the hcluster_pangenome_matrix.sh script (not shown) for statistical analysis of the PGM. In addition, the plot_matrix_heatmap.sh script was used to analyze average nucleotide identity matrices generated by get_homologues.pl. It implements the unsupervised learning method described in this work to define the optimal number of clusters in such matrices. The plot_matrix_heatmap.sh script is distributed with the GET_HOMOLOGUES suite.

individual steps separately, which adds convenient flexibility for advanced users (examples provided in the Supplementary Materials). The pipeline is highly customizable, and the reader is referred to the latest version of the online manual for the details of each option. However, the default values should produce satisfactory results for most purposes, as these were carefully

selected based on the benchmark analysis presented in this work. All the source code is freely available under the GNU GENERAL PUBLIC LICENSE V3 from https://github.com/vinuesa/get_phylomarkers. Detailed installation instructions are provided (https://github.com/vinuesa/get_phylomarkers/blob/master/INSTALL.md), along with a hands-on tutorial (https://vinuesa.github.io/get_phylomarkers/). The software has been extensively tested on diverse Linux distributions (CentOS, Ubuntu and RedHat). In addition, a docker image bundling GET_HOMOLOGUES and GET_PHYLOMARKERS is available at https://hub.docker.com/r/csicunam/get_homologues/. We recommend running the docker image to avoid potential trouble with the installation and configuration of diverse dependencies (second party binaries, as well as Perl and R packages), making it easy to install on any architecture, including Windows, and to reproduce analyses with exactly the same software.

Estimating Maximum Likelihood and Parsimony Pan-Genome Trees From the Pan-Genome Matrix (PGM)

The GET_PHYLOMARKERS package contains auxiliary scripts to perform diverse clustering and phylogenetic analyses based on the pangenome_matrix_t0.* files returned by the compare_clusters.pl script (options "-t 0 –m") from the GET_HOMOLOGUES suite. In this work, consensus PGMs (Vinuesa and Contreras-Moreira, 2015) were computed as explained in the online tutorial (https://vinuesa.github.io/get_phylomarkers/#get_homologues-get_phylomarkers-tutorials). These represent the intersection of the clusters generated by the COGtriangles and OMCL algorithms. Adding the -T flag to the previous command instructs compare_clusters.pl to compute a Wagner (multistate) parsimony tree from the PGM, launching a tree search with 50 taxon jumbles using pars from the PHYLIP (Felsenstein, 2004b) package (v.3.69). A more thorough and customized ML or parsimony analysis of the PGM can be performed with the aid of the auxiliary script estimate_pangenome_phylogenies.sh, bundled with GET_PHYLOMARKERS. By default this script performs a ML tree-search using IQT v1.6.3 (Nguyen et al., 2015). It will first call ModelFinder (Kalyaanamoorthy et al., 2017) using the JC2 and GTR2 base models for binary data, the latter accounting for unequal state frequencies. The best fitting base model + ascertain bias correction + among-site rate variation parameters are selected using the Akaike Information Criterion (AIC). IQT (Nguyen et al., 2015) is then called to perform a ML tree search under the selected model with branch support estimation. These are estimated using approximate Bayesian posterior probabilities (aBypp), a popular single branch test (Guindon et al., 2010), as well as the recently developed ultrafast-bootstrap2 (UFBoot2) test (Hoang et al., 2017). In addition, the user may choose to run a parsimony analysis with bootstrapping on the PGM, as detailed in the online manual and illustrated in the tutorial. Note however, that the parsimony search with bootstrapping is much slower than the default ML search.

Unsupervised Learning Methods for the Analysis of Pairwise Average Nucleotide (ANI) and Aminoacid (AAI) Identity Matrices

The GET_HOMOLOGUES distribution contains the plot_matrix_heatmap.sh script which generates ordered heatmaps with attached row and column dendrograms from squared tab-separated numeric matrices. These can be presence/absence PGM matrices or similarity/identity matrices, as those produced with the get_homologues -A option. Optionally, the input cgANIb matrix can be converted to a distance matrix to compute a neighbor joining tree, which makes the visualization of relationships in large ANI matrices easier. Recently added functionality includes reducing excessive redundancy in the tab-delimited ANI matrix file (-c max_identity_cut-off_value) and sub-setting the matrix with regular expressions, to focus the analysis on particular genomes extracted from the full cgANIb matrix. From version 1.0 onwards, the mean silhouette-width (Rousseeuw, 1987) goodness of clustering statistics is included to determine the optimal number of clusters automatically. The script currently depends on the R packages ape (Popescu et al., 2012), dendextend (https://cran.r-project.org/package=dendextend), factoextra (https://cran.r-project.org/package=factoextra) and gplots (https://CRAN.R-project.org/package=gplots).

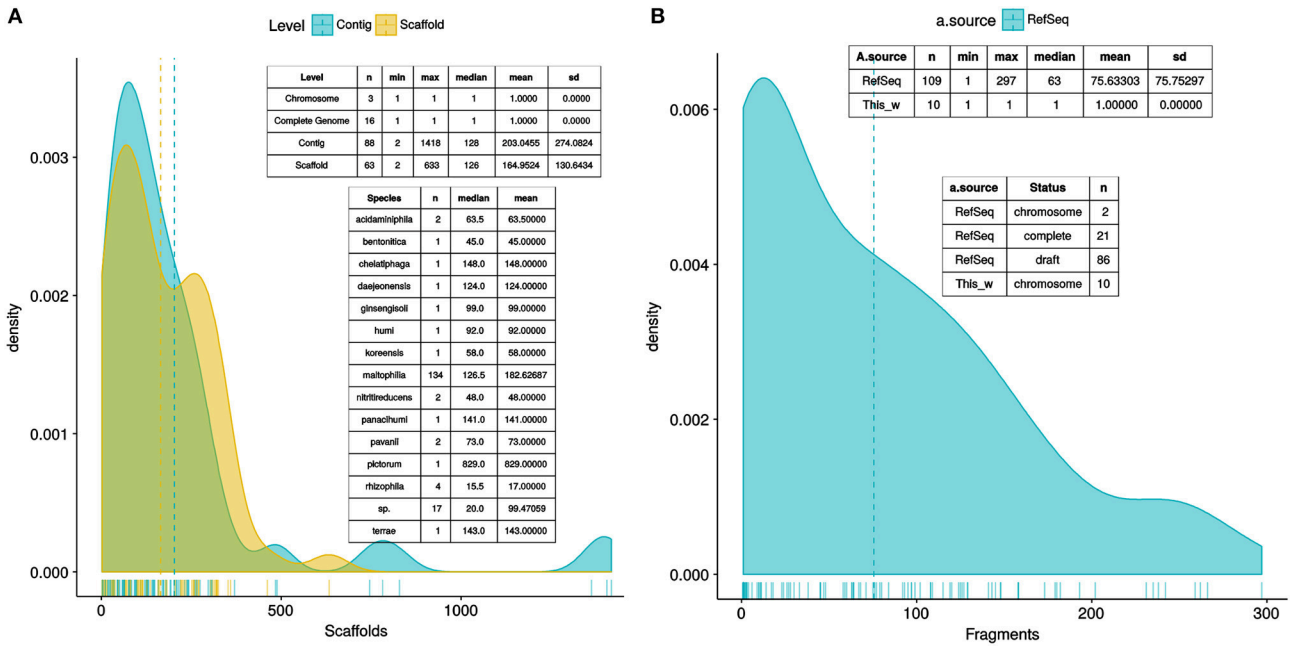
RESULTS

Ten New Complete Genome Assemblies for the Mexican Environmental *Stenotrophomonas maltophilia* Complex Isolates Previously Classified as Genospecies 1 (Smc1) and 2 (Smc2)

In this study we report the sequencing and assembly of five isolates each from the genospecies 1 (Smc1) and 2 (Smc2) recovered from rivers in Central Mexico, previously reported in our extensive MLSA study of the genus *Stenotrophomonas* (Ochoa-Sánchez and Vinuesa, 2017). All assemblies resulted in a single chromosome with gaps. No plasmids were detected. A summary of the annotated features for each genome are presented in **Table 1**. Assembly details are provided in Table S1.

Rapid Phylogenetic Exploration of *Stenotrophomonas* Genome Sequences Available at NCBI's RefSeq Repository Running GET_PHYLOMARKERS in Fast Runmode

A total of 170 *Stenotrophomonas* and 7 *Xanthomonas* reference genomes were retrieved from RefSeq (see methods). **Figure 2A** depicts parallel density plots showing the distribution of the number of fragments for the *Stenotrophomonas* assemblies at the Complete ($n = 16$), Chromosome ($n = 3$), Scaffold ($n = 63$), and Contig ($n = 88$) finishing levels. The distributions have conspicuous long tails, with an overall mean and median number of fragments of ∼238 and ∼163, respectively. The table insets in **Figure 2A** provide additional descriptive statistics of the



Level	n	min	max	median	mean	sd
Chromosome	3	1	1	1	1.0000	0.0000
Complete Genome	16	1	1	1	1.0000	0.0000
Contig	88	2	1418	128	203.0455	274.0824
Scaffold	63	2	633	126	164.9524	130.6434

Species	n	median	mean
acidaminiphila	2	63.5	63.50000
bentonitica	1	45.0	45.00000
chelatiphaga	1	148.0	148.00000
daejeonensis	1	124.0	124.00000
ginsengisoli	1	99.0	99.00000
humi	1	92.0	92.00000
koreensis	1	58.0	58.00000
maltophilia	134	126.5	182.62687
nitritireducens	2	48.0	48.00000
panacihumi	1	141.0	141.00000
pavanii	2	73.0	73.00000
pictorum	1	829.0	829.00000
rhizophila	4	15.5	17.00000
sp.	17	20.0	99.47059
terrae	1	143.0	143.00000

A.source	n	min	max	median	mean	sd
RefSeq	109	1	297	63	75.63303	75.75297
This_w	10	1	1	1	1.00000	0.00000

a.source	Status	n
RefSeq	chromosome	2
RefSeq	complete	21
RefSeq	draft	86
This_w	chromosome	10

FIGURE 2 | Density plots showing the distribution of the number of fragments of the *Stenotrophomonas* genomes available in RefSeq as of August 2017, plus the genome of *S. bentonitica* VV6, released in January 2018. **(A)** Distribution of the number of fragments in the assemblies of 170 annotated *Stenotrophomonas* genomes, as a function of assembly status (contigs vs. scaffolds), plus 7 *Xanthomonas* genomes used as outgroup to root the tree. Inset tables provide additional summary statistics of the RefSeq assemblies. **(B)** Distribution of the number of fragments in the assemblies of the 119 genomes selected for the analyses presented in this study, which include 102 reference *Stenotrophomonas* genomes, 10 new genomes generated for this study, and 7 complete *Xanthomonas* spp. genomes.

distributions. A first GET_HOMOLOGUES run was launched using this dataset ($n = 177$) with two objectives: (*i*) to test its performance with a relatively large set of genomes and (*ii*) to get an overview of their evolutionary relationships to select a non-redundant set of those with the best assemblies. For this analysis, GET_HOMOLOGUES was run in its "fast-BDBH" mode (-b), on 60 cores (-n 60; AMD Opteron™ Processor 6380, 2500.155 MHz), and imposing a stringent 90% coverage cut-off for BLASTP alignments (-C 90), excluding inparalogues (-e). This analysis took 1 h:32 m:13 s to complete and identified 132 core genes. These were fed into the GET_PHYLOMARKERS pipeline, which was executed using a default FastTree search with the following command line: "run_get_phylomarkers_pipeline.sh -R 1 -t DNA -A F," which took 8 m:1 s to complete on the same number of cores. Only 79 alignments passed the Phi recombination test. Thirteen of them failed to pass the downstream kdetree test. The phylogenetic signal test excluded nine additional loci with average SH-alrt values < 0.70. Only 57 alignments passed all filters and were concatenated into a supermatrix of 38,415 aligned residues, which were collapsed to 19,129 non-gapped and variable sites. A standard FastTree maximum-likelihood tree-search was launched, and the resulting phylogeny ($\ln L = -475237.540$) is shown in Figure S1. Based on this tree and the level of assembly completeness for each genome (**Figure 2A**), we decided to discard those with >300 contigs (**Figure 2B**). This resulted in the loss of 19 genomes labeled as *S. maltophilia*. However, we retained *S. pictorum* JCM 9942, a highly fragmented genome with 829 contigs (Patil et al., 2016) to maximize taxon sampling. Several *S. maltophilia* subclades contained identical sequences (Figure S1) and were trimmed, retaining only the assembly with the lowest numbers of scaffolds or contigs.

Selection of a Stringently Defined Set of Orthologous Genes Using GET_HOMOLOGUES

After the quality and redundancy filtering described in the previous section, 109 reference genomes (102 *Stenotrophomonas* + 7 *Xanthomonas*) were retained for more detailed investigation. Table S2 provides an overview of them. To this set we added the 10 new genomes reported in this study (**Table 1**). **Figure 2B** depicts a density plot and two inset tables summarizing the distribution of number of contigs/scaffolds in the selected reference genomes and the new genomes for the Mexican environmental Smc isolates previously classified as genospecies 1 (Smc1) and 2 (Smc2) (Ochoa-Sánchez and Vinuesa, 2017). A high stringency consensus core-genome containing 239 gene families was computed as the intersection of the clusters generated by the BDBH, COG-triangles and OMCL algorithms (**Figure 3A**).

GET_PHYLOMARKERS in Action: Benchmarking the Performance of FastTree and IQ-Tree to Select Top-Scoring Markers for Phylogenomics

The set of 239 consensus core-genome clusters (**Figure 3A**) was used to launch multiple instances of the GET_PHYLOMARKERS pipeline to evaluate the phylogenetic

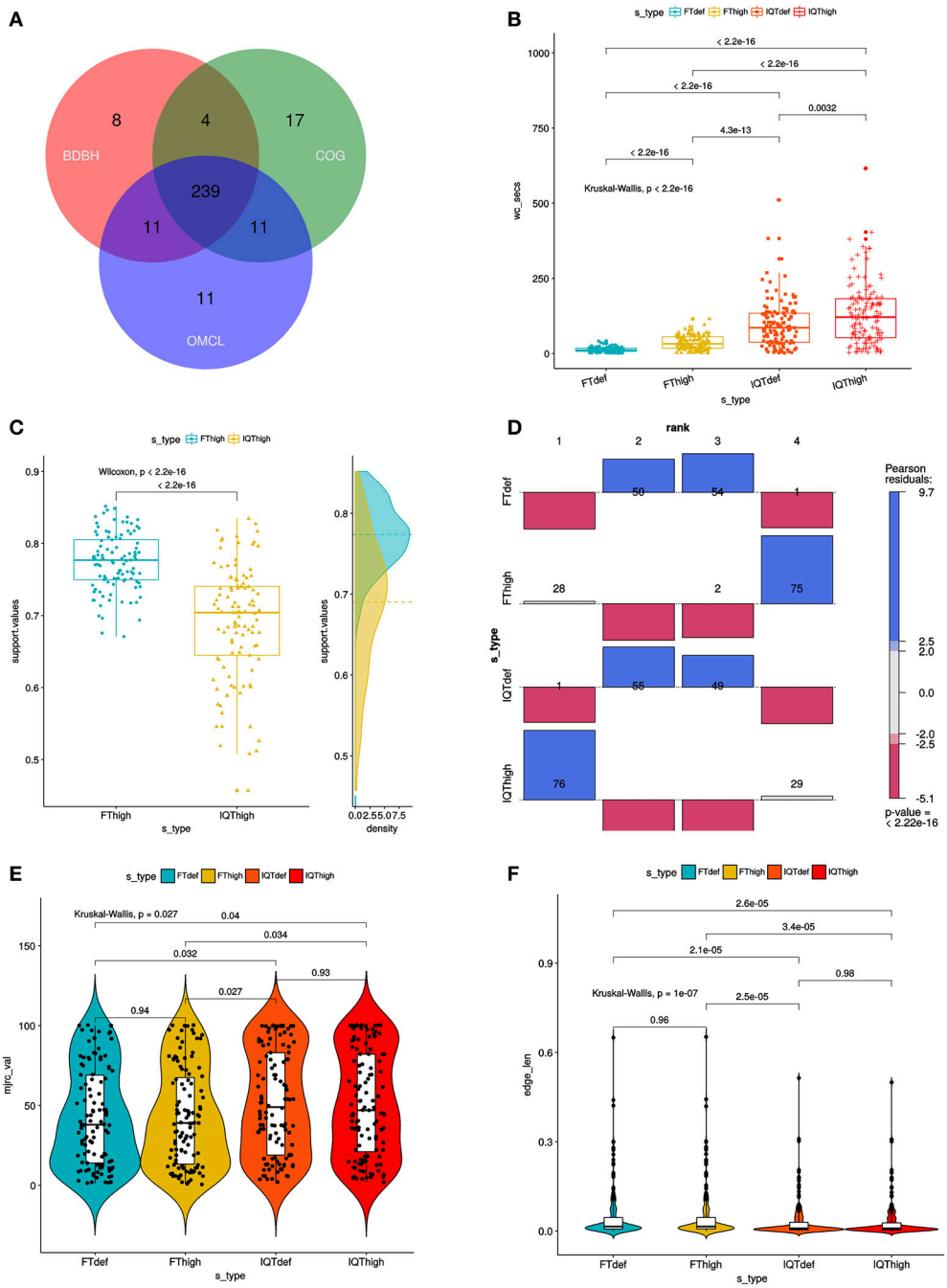


FIGURE 3 | Combined filtering actions performed by GET_HOMOLOGUES and GET_PHYLOMARKERS to select top-ranking phylogenetic markers to be concatenated for phylogenomic analyses, and benchmark results of the performance of the FastTree (FT) and IQ-TREE (IQT) maximum-likelihood (ML) phylogeny inference programs. **(A)** Venn-diagram indicating the number consensus and algorithm-specific core-genome orthologous clusters. **(B)** Parallel box-plots summarizing the computation time required by FT and IQT when run under "default" (FTdef, IQTdef) and thorough (FThigh, IQThigh) search modes (s_type) on the 239 consensus clusters, as detailed in the main text. Statistical significance of differences between treatments were computed with the Kruskal-Wallis (robust, non-parametric, ANOVA-like) test. **(C)** Distribution of SH-alrt branch support values of gene-trees found by the FThigh and IQThigh searches. Statistical significance of differences between the paired samples was computed with the Wilcoxon signed-rank test. This is a non-parametric alternative to paired *t*-test used to compare paired data when they are not normally distributed. **(D)** Association plot (computed with the vcd package) summarizing the results of multi-way Chi-Square analyses of the ln*L* score ranks (1–4, meaning best to worst) of the ML gene-trees computed from the set of 105 codon alignments passing the kdetrees filter in the IQThigh run (**Table 2**) for each search-type. The height and color-shading of the bars indicate the magnitude and significance level of the Pearson residuals. **(E)** Statistical analysis (Kruskal-Wallis test) of the distribution of consensus values from majority-rule consensus trees computed from the gene trees passing all the filters, as a function of search-type. **(F)** Statistical analysis (Kruskal-Wallis test) of the distribution of the edge-lengths of species-trees computed from the concatenated top-scoring markers, as a function of search-type.

TABLE 2 | Comparative benchmark analysis of the filtering performance of the GET_PHYLOMARKERS pipeline when run using the FastTree (FT) and IQ-TREE (IQT) maximum-likelihood algorithms, under default and high search-intensity levels.

Test	FTdef	FThigh	IQTdef	IQThigh
Alignments passing the Phi recombination test	127/239 (53.14%)	125/239 (52.30%)	125/239 (52.30%)	127/239 (53.14%)
Outlier phylogenies (kdetrees test; $k = 1.0$) out of the indicated number of non-recombinant alignments	22/127 (17.32%) passing: 105	18 (14.17%) passing: 107	19 (14.96%) passing: 106	22 (17.32%) passing: 105
Alignments passing the phylogenetic signal (mean SH-alrt bipartition support; $m \geq 0.7$) test	98/105 (93.33%)	99/107 (92.52%)	52/106 (49.05%)	55/105 (52.38%)
Concatenated top-scoring markers, ln*L* score, substitution model and number of independent searches	98 markers var. sites = 36,082 ln*L* = −917444.522 GTR+G Searches = 1	99 markers var. sites = 35,509 ln*L* = −899898.614 GTR+G Searches = 1	52 markers var. sites = 25,383 ln*L* = −666437.563 GTR+F+ASC+R6 Searches = 1	55 markers var. sites = 26,988 ln*L* = −707933.476 GTR+F+ASC+R6 Searches = 5
Total wall-clock time of runs on 50 cores	0 h:13 m:39 s	0 h:38 m:30 s	1 h:22 m:18 s	2 h:40 m:13 s

The analyses were started with the stringently defined set of 239 consensus core-genome clusters computed by GET_HOMOLOGUES for a dataset of 119 genomes (112 Stenotrophomonas spp. and 7 Xanthomonas spp.).

performance of FastTree (FT; v2.1.10) and IQ-TREE (IQT; v1.6.3), two popular fast maximum-likelihood (ML) tree searching algorithms. Our benchmark was designed to compare: (*i*) the execution times of the FT vs. IQT runs under default (FTdef, IQTdef) and thorough (FThigh, IQThigh) search modes (see methods and online manual for their parameterization details); (*ii*) the phylogenetic resolution (average support values) of gene trees estimated by FT and IQT under both search modes; (*iii*) the rank of ln*L* scores of the gene trees found in those searches for each locus; (*iv*) the distribution of consensus values of each node in majority rule consensus trees computed from the gene trees found by each search type; (*v*) the distribution of edge-lengths in the species-trees computed by each search type. The results of these analyses are summarized in **Table 2** and in **Figure 3**. The first steps of the pipeline (**Figure 1**) comprise the generation of codon alignments and their analysis to identify potential recombination events. Only 127 alignments (53.14%) passed the Phi-test (**Table 2**). Phylogenetic analyses start downstream of the recombination test (**Figure 1**). The computation times required by the two algorithms and search intensity levels were significantly different (Kruskal-Wallis, $p < 2.2e\text{-}16$), FastTree being always the fastest, and displaying the lowest dispersion of compute times across trees (**Figure 3B**). This is not surprising, as IQT searches involved selecting the best substitution model among a range of base models (see methods and online manual) and fitting additional parameters (+G+ASC+I+F+R) to account for heterogeneous base frequencies and rate-variation across sites. In contrast, FT searches just estimated the parameter values for the general time-reversible (GTR) model, and among-site rate variation was modeled fitting a gamma distribution with 20 rate categories (+G), as summarized in **Table 2**. Similar numbers of "outlier" trees (range 18:22) were detected by the kdetrees-test in the four search types (**Table 2**). However, the distributions of SH-alrt support values are strikingly different for both search algorithms (Wilcoxon, $p < 2.2e\text{-}16$), revealing that gene-trees found by IQT have a much lower average support than those found by FT (**Figure 3C**). Consequently, the former searches were significantly more efficient to identify gene trees with low average branch support values (**Table 2** and **Figure 3C**). This result is in line with the well-established fact that poorly fitting and under-parameterized models produce less reliable tree branch lengths and overestimate branch support (Posada and Buckley, 2004), implying that the FT phylogenies may suffer from clade over-credibility. These results demonstrate that: (*i*) FT-based searches are significantly faster than those performed with IQT, and (*ii*) that IQT has a significantly higher discrimination power for phylogenetic signal than FT. Due to the fact that the number of top-scoring alignments selected by the two algorithms for concatenation is notably different (**Table 2**), the ln*L* scores of the resulting species-trees are not comparable (**Table 2**). Therefore, in order to further evaluate the quality of the gene-trees found by the four search strategies, we performed an additional benchmark under highly standardized conditions, based on the 105 optimal alignments that passed the kdetrees-test in the IQThigh search (**Table 2**). Gene trees were estimated for each of these alignments using the four search strategies (FTdef, IQTdef, FThigh, and IQThigh) and their ln*L* scores ranked for each gene tree. An association analysis (deviation from independence in a multi-way chi-squared test) was performed on the ln*L* ranks (1–4, coding for highest to lowest ln*L* scores, respectively) attained by each search type for each gene tree. As shown in **Figure 3D**, the IQThigh search was the winner, attaining the first rank (highest ln*L* score) in 76/105 of the searches (72.38%), way ahead of the number of FThigh (26%), and IQTdef (0.009%) searches that ranked in the first position (highest ln*L* score for a particular alignment). A similar analysis performed on the full set of input alignments ($n = 239$) indicated that when operating on an unfiltered set, the difference in performance was even more striking, with IQT-based searches occupying > 97% of the first rank positions (data not shown). These results highlight two points: (*i*) the importance of proper model selection and thorough tree searching in phylogenetic inference and (*ii*) that IQT generally finds better trees than FT. Finally, we evaluated additional phylogenetic attributes of the species-trees computed by each search type, either as the

majority rule consensus (mjrc) tree of top-scoring gene-trees, or as the tree estimated from the supermatrices of concatenated alignments. **Figure 3E** shows the distribution of mjrc values of the mjrc trees computed by each search type, which can be interpreted as a proxy for the level phylogenetic congruence among the source trees. These values were significantly higher for the IQT than in the FT searches (Kruskal-Wallis, $p = 0.027$), with a higher number of 100% mjrc clusters found in the former than in the latter type of trees (**Figure 3E**). An analysis of the distribution of edge-lengths of the species-trees inferred from the concatenated alignments revealed that those found in IQT searches had significantly (Kruskal-Wallis, $p =$ 1e-07) shorter edges (branches) than those estimated by FT (**Figure 3F**). This highlights again the importance of adequate substitution models for proper edge-length estimation. Tree-lengths (sum of edge lengths) of the species-trees found in IQT-based searches are about 0.63 times shorter than those found by FT (Figure S2). As a final exercise, we computed the Robinson-Foulds (RF) distances of each gene tree found in a given search type to the species tree inferred from the corresponding supermatrix. The most striking result of this analysis was that no single gene-tree had the same topology as the species tree inferred from the concatenated top-scoring alignments (Figure S3).

Effect of Tree-Search Intensity on the Quality of the Species Trees Found by IQT-REE and FastTree

Given the astronomical number of different topologies that exist for 119 terminals, we decided to evaluate the effect of tree-search thoroughness on the quality of the trees found by FT and IQT, measured as their log-likelihood (ln*L*) score. To make the results comparable across search algorithms, we used the supermatrix of 55 top-scoring markers (25,896 variable, non-gapped sites) selected by the IQThigh run (**Table 2**). One thousand FT searches were launched from the same number of random topologies computed with the aid of a custom Perl script. In addition, a standard FT search was started from the default BioNJ tree. All these searches were run in "thorough" mode (-quiet -nt -gtr -bionj -slow -slownni -gamma -mlacc 3 -spr 16 -sprlength 10) on 50 cores. The resulting ln*L* profile for this search is presented in **Figure 4A**, which reached a maximal score of −717195.373. This is 121.281 ln*L* units better than the score of the best tree found in the search started from the BioNJ seed tree (ln*L* –717316.654, lower discontinuous blue line). In addition, 50 independent tree searches were run with IQT under the best fitting model previously found (**Table 2**), using the shell loop command (# 5) provided in the Supplementary Material. The corresponding ln*L* profile of this search is shown in **Figure 4B**, which found a maximum-scoring tree with a score of –707932.468. This is only 8.105 ln*L* units better than the worst tree found in that same search (**Figure 4B**). Importantly, the best tree found in the IQT-search is 9262.905 ln*L* units better that of the best tree found in the FT search, despite the much higher number of seed trees used for the latter. This result clearly demonstrates the superiority of the IQT algorithm for ML tree searching. Based on this evidence, and that presented in the previous section (**Table 2**; **Figure 3**), IQT was chosen as the default tree-search algorithm used by GET_PHYLOMARKERS. The Robinson-Foulds distance between both trees was 46.

A Robust Genomic Species Phylogeny for the Genus *Stenotrophomonas*: Taxonomic Implications and Identification of Multiple Misclassified Genomes

Figure 5 displays the best ML phylogeny found in the IQT search (**Figure 4B**) described in the previous section. This is a highly resolved phylogeny. All bipartitions have an approximate

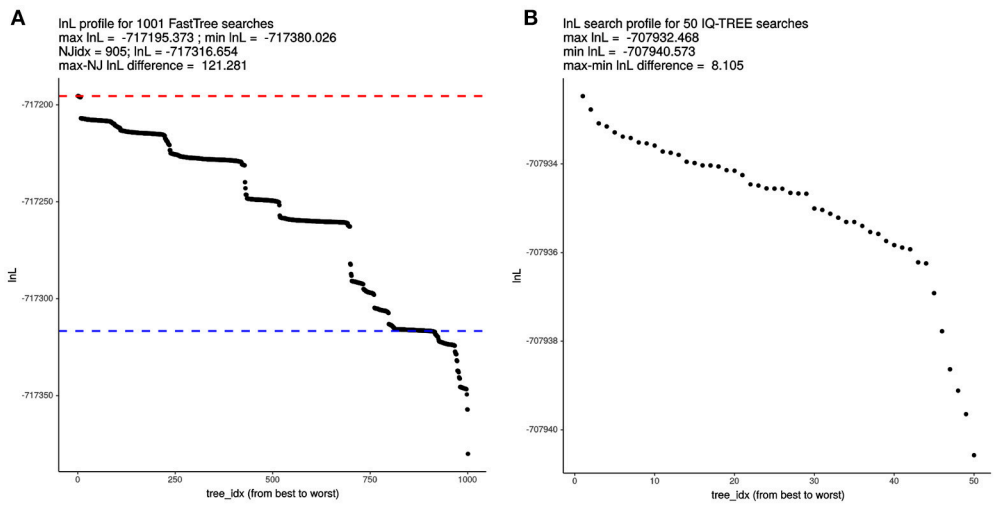


FIGURE 4 | Comparative analysis of log-likelihood tree search profiles. **(A)** Sorted ln*L* profile of FastTree (FT) tree searches launched from 1,000 random trees + 1 BioNJ phylogeny, using the "thorough" tree-search settings described in the main text and the 55 top-ranking markers (26,988 non-gapped, variable sites) selected by the IQThigh run for 119 genomes (**Table 2**). The dashed blue line indicates the score of the search initiated from the BioNJ tree. **(B)** Sorted ln*L* profile of 50 independently launched IQ-TREE (IQT) searches under the best-fitting model using the same matrix as for the FT search.

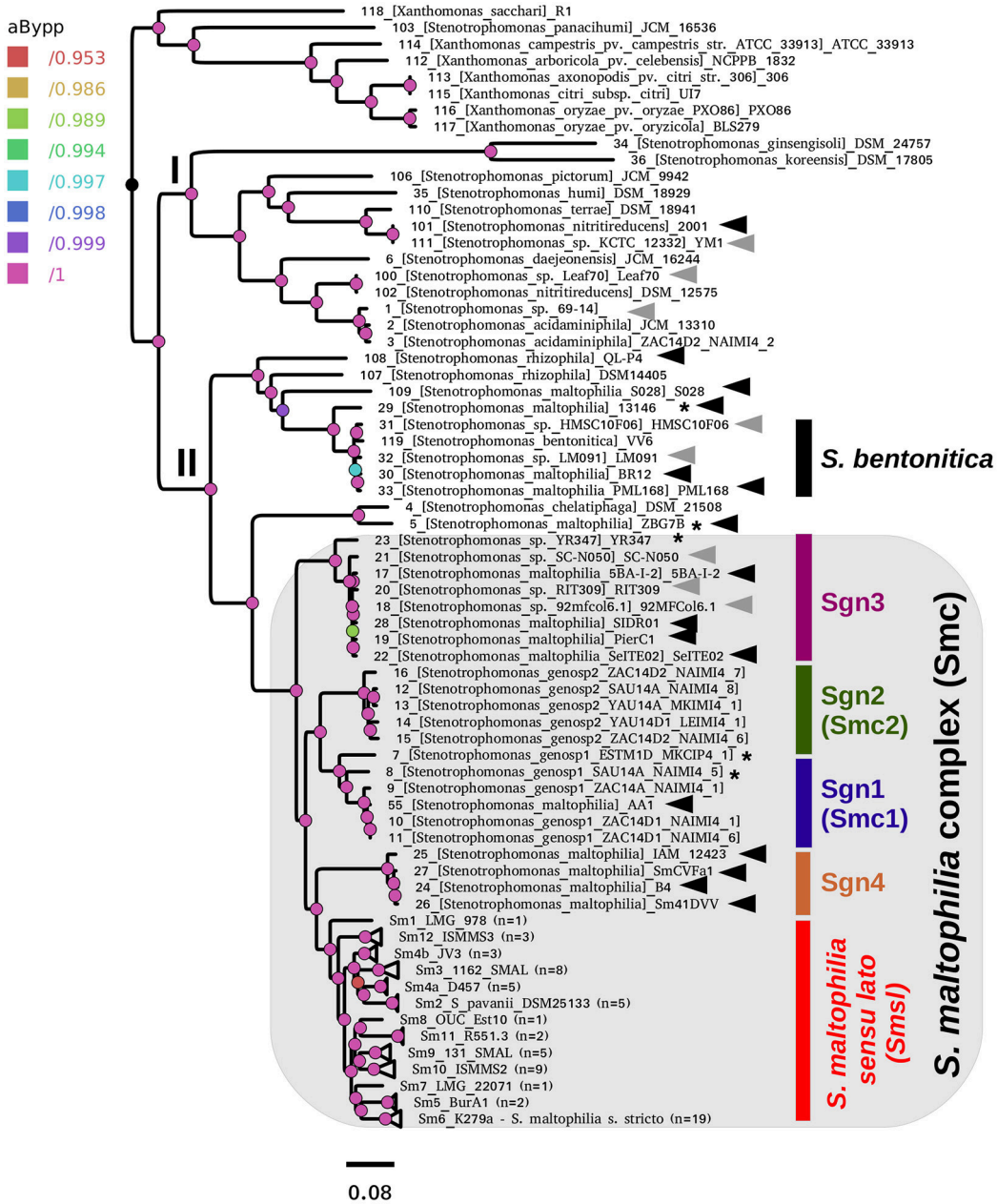


FIGURE 5 | Best maximum-likelihood core-genome phylogeny for the genus *Stenotrophomonas* found in the IQ-TREE search described in **Figure 4B**, based on the supermatrix obtained by concatenation of 55 top-ranking alignments (**Table 2**). The tree was rooted using the *Xanthomonas* spp. sequences as the outgroup. Arrows highlight genomes not grouping in the *S. maltophilia sensu lato* clade (Smsl), for which we suggest a reclassification, as summarized in **Table 3**. Black arrows indicate misclassified strains, while gray ones mark unclassified genomes. The shaded area highlights the strains considered as members of the *S. maltophilia* complex (Smc). The genospecies 1 and 2 (Sgn1 = Smc1; Sgn2 = Smc2) were previously recognized as separate species-like lineages by Ochoa-Sánchez and Vinuesa (2017). Strains grouped in the Smsl clade are collapsed into sub-clades that are perfectly consistent with the cluster analysis of core-genome average nucleotide identity (cgANIb) values presented in **Figure 7** at a cutoff-value of 95.9%. Integers in parentheses correspond to the number of genomes in each collapsed clade. Figure S4 displays the same tree in non-collapsed form. Strains from genospecies 1, 3, and 5 (Sgn1, Sgn3, Sgn5) marked with an asterisk may represent additional species, according to cgANIb values. Nodes are colored according to the lateral scale, which indicates the approximate Bayesian posterior probability values. The scale bar represents the number of expected substitutions per site under the best-fitting GTR+ASC+F+R6 model.

Bayesian posterior probability (aBypp) $p \geq 0.95$. It was rooted at the branch subtending the *Xanthomonas* spp. clade, used as an outgroup. A first taxonomic inconsistency revealed by this phylogeny is the placement of *S. panacihumi* within the latter clade, making the genus *Stenotrophomonas* paraphyletic. It is worth noting that *S. panacihumi* is a non-validly described,

and poorly characterized species (Yi et al., 2010). The genus *Stenotrophomonas*, as currently defined, and excluding *S. panacihumi*, consists of two major clades, labeled as I and II in **Figure 5**, as previously defined (Ochoa-Sánchez and Vinuesa, 2017).

Clade I groups environmental isolates, recovered from different ecosystems, mostly soils and plant surfaces, classified as *S. ginsengisoli* (Kim et al., 2010), *S. koreensis* (Yang et al., 2006), *S. daejeonensis* (Lee et al., 2011), *S. nitritireducens* (Finkmann et al., 2000), *S. acidaminiphila* (Assih et al., 2002), *S. humi*, and *S. terrae* (Heylen et al., 2007). The recently described *S. pictorum* (Ouattara et al., 2017) is also included in clade I. These are all rather poorly studied species, for which only one or a few strains have been considered in the corresponding species description or to study particular aspects of their biology. None of these species have been reported as opportunistic pathogens, but some contain promising strains for plant growth-promotion and bio-remediation. Particularly notorious are the disproportionally long terminal branches (heterotachy) of *S. ginsengisoli* and *S. koreensis* (**Figure 5**). The potential impact of these long branches on the estimated phylogeny needs to be evaluated in future work.

Clade II contains the species *S. rhizophila* (Wolf et al., 2002), *S. chelatiphaga* (Kaparullina et al., 2009), the recently described *S. bentonitica* (Sánchez-Castro et al., 2017), along with multiple species and genospecies lumped in the *S. maltophilia* complex (Smc; shaded area in **Figure 5**) (Svensson-Stadler et al., 2012; Berg and Martinez, 2015). The Smc includes the validly described *S. maltophilia* (Palleroni and Bradbury, 1993) and *S. pavanii* (Ramos et al., 2011) (collapsed subclades Sm6 and Sm2, respectively, located within the clade labeled as *S. maltophilia sensu lato* in **Figure 5**), along with at least four undescribed genospecies (Sgn1-Sgn4) recently identified in our MLSA study of the genus (Ochoa-Sánchez and Vinuesa, 2017). In light of this phylogeny, we discovered 16 misclassified RefSeq genome sequences (out of 119; 13.44%), 14 of them labeled as *S. maltophilia*. These genomes are highlighted with black arrows in **Figure 5**. The phylogeny also supports the classification, either as a validly published species, or as new genospecies, of 8 (∼6.72%) additional RefSeq genomes (gray arrows) lacking a species assignation in the RefSeq record, as summarized in **Table 3**. In addition, the phylogeny resolved 13 highly supported lineages (aBypp > 0.95) within the *S. maltophilia sensu lato* (Smsl) cluster, shown as collapsed clades. They have a cgANIb

TABLE 3 | RefSeq genome sequences reclassified in this study based on the genomic evidence presented herein (see **Figures 5**–**7**).

DEFINITION (RefSeq classification)	Species/Reclassification*	Status	Fragments	BioProject	BioSample	PMID
Stenotrophomonas sp. 69-14	*S. acidaminiphila*	Draft	27	PRJNA279279	SAMN05660631	NA
Stenotrophomonas maltophilia ZBG7B	*S. chelatiphaga*	Draft	145	PRJNA272355	SAMN03280975	26659682
Stenotrophomonas maltophilia AA1	Genospecies 2 (Sgn2/Smc2)	Complete	1	PRJNA224116	SAMN06130959	28275097
Stenotrophomonas maltophilia 5BA-I-2	Genospecies 3	Draft	4	PRJNA224116	SAMN02641498	24604648
Stenotrophomonas sp. 92mfcol6.1	Genospecies 3	Draft	11	PRJNA224116	SAMN04488690	NA
Stenotrophomonas maltophilia PierC1	Genospecies 3	Draft	59	PRJEB8824	SAMEA3309462	26276674
Stenotrophomonas sp. RIT309	Genospecies 3	Draft	45	PRJNA224116	SAMN02676627	24812212
Stenotrophomonas sp. SC-N050	Genospecies 3	Draft	24	PRJNA224116	SAMN05720615	NA
Stenotrophomonas maltophilia SelTE02	Genospecies 3	Draft	63	PRJNA224116	SAMEA3138997	24812214
Stenotrophomonas sp. YR347	Genospecies 3	Draft	11	PRJNA224116	SAMN05518671	NA
Stenotrophomonas maltophilia B4	Genospecies 4	Draft	180	PRJNA224116	SAMN03753636	NA
Stenotrophomonas maltophilia Sm41DVV	Genospecies 4	Draft	26	PRJNA323790	SAMN05188789	27540065
Stenotrophomonas maltophilia SmCVFa1	Genospecies 4	Draft	30	PRJNA323845	SAMN05190067	27540065
Stenotrophomonas maltophilia IAM 12423	Genospecies 4	Draft	9	PRJNA224116	SAMN04487782	NA
Stenotrophomonas maltophilia SIDR01	genospecies 4	Draft	2	PRJNA248909	SAMN06040735	NA
Stenotrophomonas maltophilia 13146	*S. bentonitica complex*	Draft	60	PRJNA224116	SAMN07237143	NA
Stenotrophomonas maltophilia BR12*S*	*S. bentonitica*	Draft	80	PRJNA224116	SAMN03456145	26472823
Stenotrophomonas sp. HMSC10F07	*S. bentonitica*	Draft	63	PRJNA269850	SAMN03287020	NA
Stenotrophomonas sp. LM091	*S. bentonitica*	Complete	1	PRJNA344031	SAMN05818440	27979933
Stenotrophomonas maltophilia PML168	*S. bentonitica*	Draft	97	PRJNA224116	SAMEA2272452	22887661
Stenotrophomonas sp. Leaf70	*S. nitritireducens*	Draft	11	PRJNA224116	SAMN04151613	26633631
Stenotrophomonas sp. KCTC 12332	*S. terrae complex*	Complete	1	PRJNA310387	SAMN04451766	28689013
Stenotrophomonas nitritireducens 2001	*S. terrae complex*	Complete	1	PRJNA224116	SAMN05428703	NA
Stenotrophomonas maltophilia S028	*Stenotrophomonas* sp.	Draft	297	PRJNA224116	SAMN02469568	23144428
Stenotrophomonas rhizophila QL-P4	*Stenotrophomonas* sp.	Complete	1	PRJNA326321	SAMN05276013	NA

*The numbered genospecies correspond to novel unnamed species identified by Ochoa-Sánchez and Vinuesa (2017) and in this study. Strains assigned to the S. bentonitica and the S. terrae complexes. Most likely represent novel species related to these species, respectively.

>96% (**Figure 5**). These lineages may represent 13 additional species in the Smsl clade, as detailed in following sections. Figure S4 shows the non-collapsed version of the species-tree displayed in **Figure 5**.

No genome sequences, nor MLSA data are available for the recently described *S. tumulicola* (Handa et al., 2016).

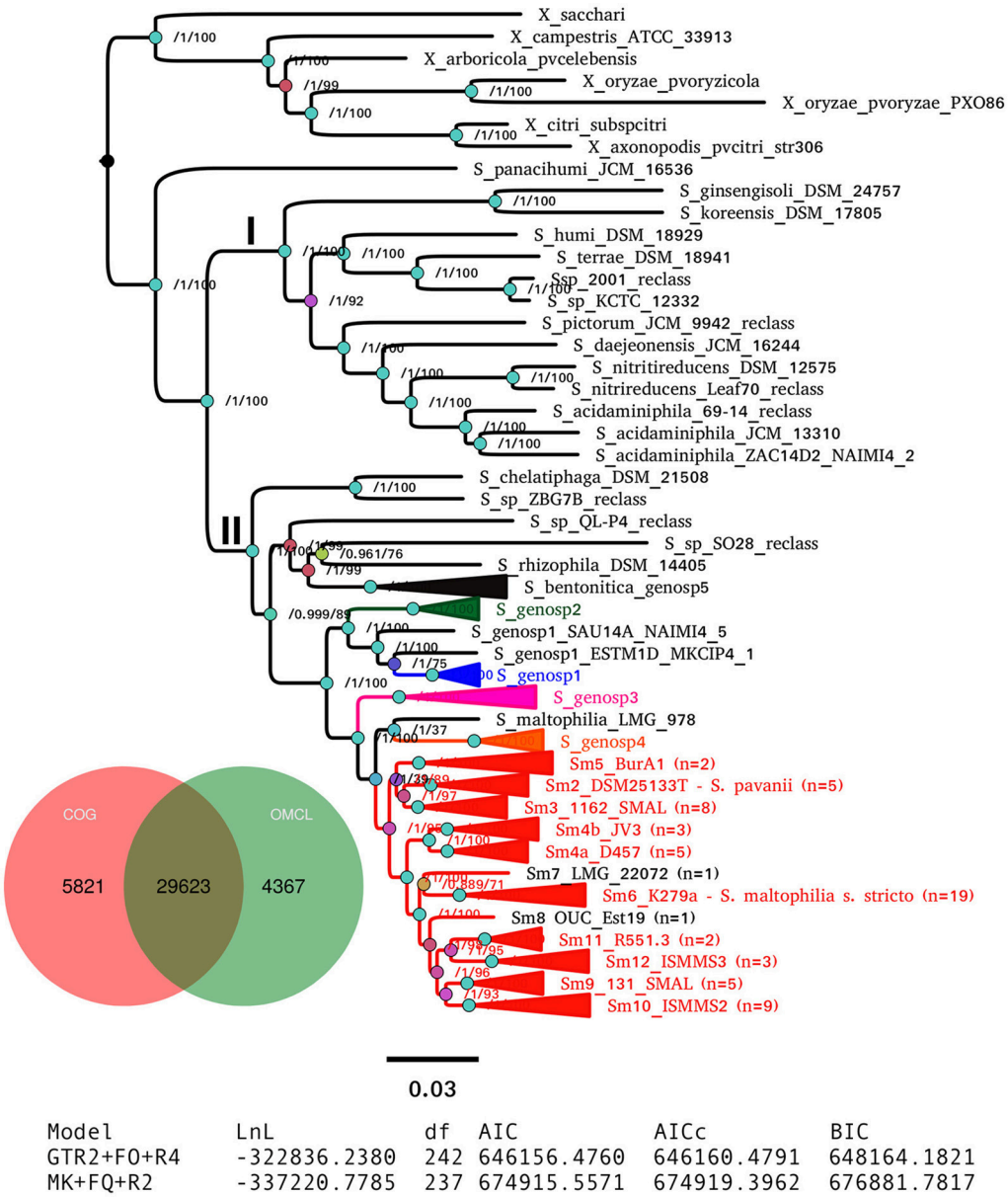
Pan-Genome Phylogenies for the Genus *Stenotrophomonas* Recover the Same Species Clades as the Core-Genome Phylogeny

A limitation of core-genome phylogenies is that they are estimated from the small fraction of single-copy genes shared by all organisms under study. Genes encoding adaptive traits relevant for niche-differentiation and subsequent speciation events typically display a lineage-specific distribution. Hence, phylogenetic analysis of pan-genomes, based on their differential gene-composition profiles, provide a complementary, more resolved and often illuminating perspective on the evolutionary relationships between species.

A consensus PGM containing 29,623 clusters was computed from the intersection of those generated by the COG-triangles and OMCL algorithms (**Figure 6**). This PGM was subjected to ML tree searching using the binary and morphological models implemented in IQT for phylogenetic analysis of discrete characters with the aid of the estimate_pangenome_phylogenies.sh script bundled with GET_PHYLOMARKERS (**Figure 1**). As shown in the tabular inset of **Figure 6**, the binary GTR2+FO+R4 model was by large the best-fitting one (with the smallest AIC and BIC values). Twenty five independent IQT searches were performed on the consensus PGM with the best-fitting model. The best tree found is presented in **Figure 6**, rooted with the *Xanthomonas* spp. outgroup sequences. It depicts the evolutionary relationships of the 119 genomes based on their gene content (presence-absence) profiles. The numbers on the nodes indicate the approximate Bayesian posterior probabilities (aBypp)/UFBoot2 support values (see methods). The same tree, but without collapsing clades, is presented in the Figure S5. This phylogeny resolves exactly the same species-like clades highlighted on the core-genome phylogeny presented in **Figure 5**, which are also grouped in the two major clades I and II. These are labeled with the same names and color-codes, for easy cross-comparison. However, there are some notorious differences in the phylogenetic relationships between species on both trees, like the placement of *S. panacihumi* outside of the *Xanthomonas* clade, and the sister relation of genospecies 3 (Sgnp3) to the *S. maltophilia sensu lato* clade. These same relationships were found in a multi-state (Wagner) parsimony phylogeny of the PGM shown in Figure S6. In summary, all core-genome and pan-genome analyses presented consistently support our previous claim that the five genospecies defined in our MLSA study represent distinct species and support the existence of multiple cryptic species within the Smsl clade, as defined in **Figure 5**.

Application of a Non-supervised Learning Approach to BLAST-Based Core-Genome Average Nucleotide Distance (cgANDb) Matrices to Identify Statistically-Consistent Species-Like Clusters

The final goal of any geno-taxonomic study is to identify species-like clusters. These should consist of monophyletic groups identified on genome trees that display average genome identity (gANI) values >94%, based on a widely accepted cutoff-value (Rosselló-Móra and Amann, 2015). In this section we searched for such species-clusters within the taxonomically problematic *Stenotrophomonas maltophilia* complex (Smc). Our core- and pan-genome phylogenies consistently identified potential species-clades within the Smc that grouped exactly the same strains (compare **Figures 5**, **6**). We additionally performed a cluster analysis of core-genome ANI values computed from the pairwise BLASTN alignments (cgANIb) used to define OMCL core-genome clusters for the 86 Smc genomes analyzed in this study. The resulting cgANIb matrix was then converted to a distance matrix (cgANDb = 100%–cgANIb) and clustered with the aid of the plot_matrix_heatmap.sh script from the GET_HOMOLOGUES suite. **Figure 7** shows the resulting cladogram, which resolves 16 clusters within the Smc at a conservative cgANDb cutoff value of 5% (cgANIb = 95%). At this distance level, the four genospecies labeled as Sgn1-Sgn4 on **Figure 5** are resolved as five clusters because the most divergent Sgn1 genome (ESTM1D_MKCIP4_1) is split as a separate lineage. This is the case also at cgANDb = 6 (**Figure 7**), reason why this strain most likely represents a sixth genospecies. All these genospecies are very distantly related to the large *S. maltophilia sensu lato* cluster, which gets split into 11 sub-clusters at the conservative cgANDb = 5% cutoff. Thirteen clusters are resolved at the 4% threshold, and a minimum of seven at the 6% level (cgANIb = 94%), as shown by the dashed lines (**Figure 7**). These results strongly suggest that the *S. maltophilia sensu lato* clade (**Figure 5**) actually comprises multiple species. The challenging question is how many? In an attempt to find a statistically-sound answer, we applied an unsupervised learning approach based on the evaluation of different goodness of clustering statistics to determine the optimal number of clusters (k) for the cgANDb matrix. The gap-statistic and a parametric, model-based cluster analysis, yielded k-values $\geq$ 35 (data not shown). These values seem too high for this dataset, as they correspond to a gANI value > 98%. However, the more conservative average silhouette width (ASW) method (Kaufman and Rousseeuw, 1990) identified an optimal k = 19 (inset in **Figure 7**) for the complete set of Smc genomes. This number of species-like clusters is much more reasonable for this data set, as it translates to a range of cgANDb between 4.5 and 4.7 (cgANIb range: 95.5–95.3%). Close inspection of the ASW profile reveals that the first peak is found at k = 13, which has an almost identical ASW as that of the maximal value and maps to a cgANDg = 5.7 (cgANIb of 94.3%). In summary, the range of reasonable numbers of clusters proposed by the ASW statistic (k = 13 to k = 19) corresponds to cgANDb values in the range of



Model	LnL	df	AIC	AICc	BIC
GTR2+FO+R4	-322836.2380	242	646156.4760	646160.4791	648164.1821
MK+FQ+R2	-337220.7785	237	674915.5571	674919.3962	676881.7817

FIGURE 6 | Maximum-likelihood pan-genome phylogeny estimated with IQ-TREE from the consensus pan-genome clusters displayed in the Venn diagram. Clades of lineages belonging to the *S. maltophilia* complex are collapsed and are labeled as in **Figure 5**. Numbers on the internal nodes represent the approximate Bayesian posterior probability/UFBoot2 bipartition support values (see methods). The tabular inset shows the results of fitting either the binary (GTR2) or morphological (MK) models implemented in IQ-TREE, indicating that the former has an overwhelmingly better fit. The scale bar represents the number of expected substitutions per site under the binary GTR2+F0+R4 substitution model.

5.7–4.5% (cgANIb range: 94.3–95.5%), which fits well with the new gold-standard for species delimitation (gANI > 94%), established in influential works (Konstantinidis and Tiedje, 2005; Richter and Rosselló-Móra, 2009). We noted however, that at a cgANDb = 4.1% (cgANIb = 95.9%) the strain composition of the clusters was 100% concordant with the monophyletic subclades shown in the core-genome (**Figure 5**) and pan-genome (**Figure 6**) phylogenies. Importantly, at this cutoff, the length of the branches subtending each cluster is maximal, both on the core-genome phylogeny (**Figure 5**) and on the cgANDb cladogram (**Figure 7**). Based on the combined and congruent evidence provided by these complementary approaches, we can safely conclude that: (*i*) the Smc genomes analyzed herein may actually comprise up to 19 or 20 different species-like lineages, and (*ii*) that only the strains grouped in the cluster labeled as Sm6 in **Figures 5**–**7** should be called *S. maltophilia*.

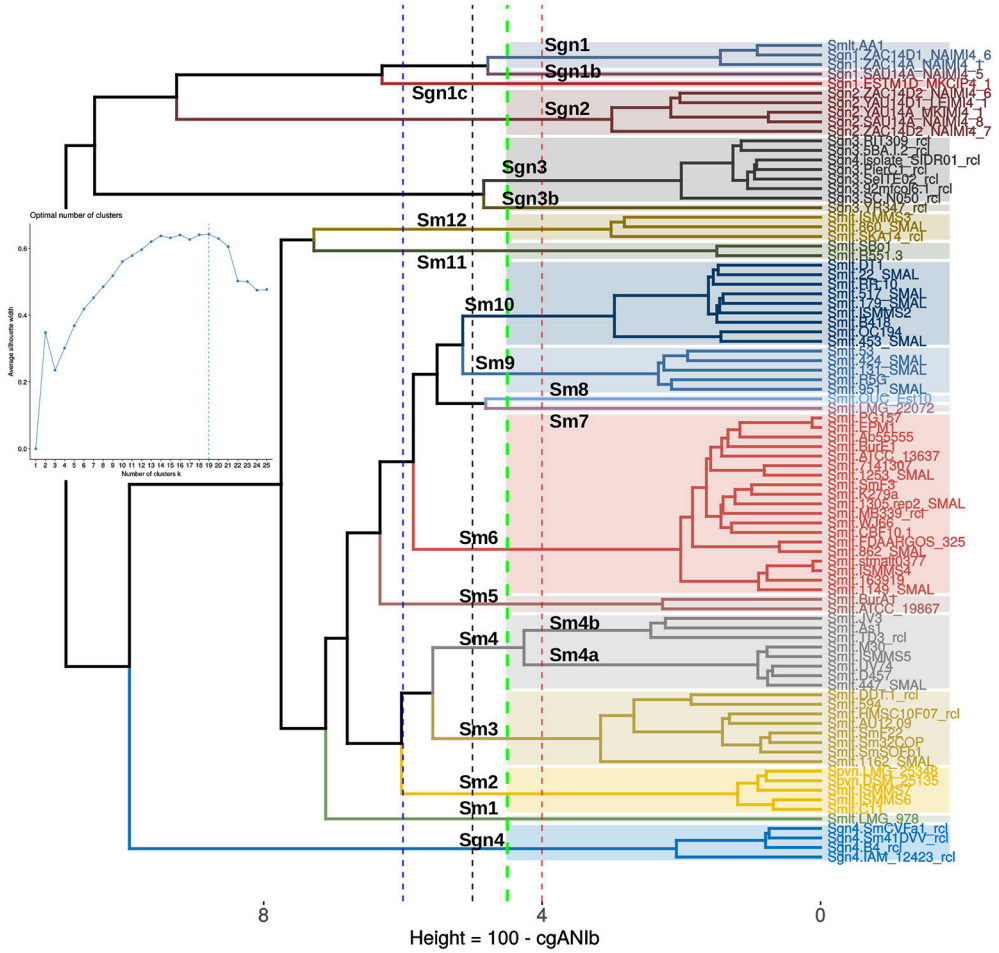


FIGURE 7 | Application of an unsupervised learning approach to the cgANIb distance matrix to identify statistically-consistent species-like clusters. The cgANIb matrix was converted to a distance matrix (cgANDb) and clustered using the Ward.D2 algorithm. The optimal number of clusters (*k*) was determined with the average silhouette-width statistic. The inset shows the statistic's profile, with $k = 19$ as the optimal number of clusters. This number corresponds to an cgANIb of 95.5% (gray dashed line). At a cgANDb of 4.1% (cgANIb = 95.9%) the groups delimited by the clustering approach are perfectly consistent with those delimited by the core- and pan-genome ML phylogenies displayed in **Figures 5**, **6**, respectively.

The latter is the most densely sampled species-like cluster ($n = 19$) and includes ATCC 13637^T, the type strain of the species.

On the Ecology and Other Biological Attributes of the Species-Like Clusters in the *Stenotrophomonas maltophilia* Complex

In this final section we present a brief summary of the ecological attributes reported for selected members of the species-like clusters resolved within the Smc (**Figures 5**, **7**). The four unnamed genospecies (Sgn1-Sgn4) group mainly environmental isolates. This is consistent with our previous evolutionary and ecological analyses of a comprehensive multilocus dataset of the genus (Ochoa-Sánchez and Vinuesa, 2017). In that study only Mexican environmental isolates were found to be members of the newly discovered genospecies Sgn1 and Sgn2 (named as Smc1 and Smc2, respectively). In this work we discovered that the recently sequenced maize root isolate AA1 (Niu et al., 2017), misclassified as *S. maltophilia*, clusters tightly with the Sgn1 strains (**Figure 5**). The *S. maltophilia sensu lato* clade is split into 12 or 13 groups based on cgANDb (**Figure 7**). Sm6 forms the largest cluster, grouping mostly clinical isolates related to the type strain *S. maltophilia* ATCC 13637^T, like the model strain K279a (Crossman et al., 2008), ISMMS4 (Pak et al., 2015), 862_SMAL, 1149_SMAL, and 1253_SMAL (Roach et al., 2015), as well as EPM1 (Sassera et al., 2013), recovered from the human parasite *Giardia duodenalis*. However, this group also comprises some environmental isolates like BurE1, recovered from a bulk soil sample (Youenou et al., 2015). In summary, cluster Sm6 holds the *bona fide S. maltophilia* strains (*sensu stricto*), which may be well-adapted to associate with different eukaryotic hosts and cause opportunistic infections in humans. Cluster Sm4a contains the

model strain D574 (Lira et al., 2012) along with four other clinical isolates (Conchillo-Solé et al., 2015) and therefore may represent a second clade enriched in strains with high potential to cause opportunistic pathogenic infections in humans. Noteworthy, this group is distantly related to Sm6 (**Figures 5**, **7**). Cluster Sm4b is closely related to Sm4a based on the pan-genome phylogeny and the cgANDd cladogram (**Figures 6**, **7**). It groups the Brazilian rhizosphere-colonizing isolate JV3, the Chinese highly metal-tolerant strain TD3 (Ge and Ge, 2016) and strain As1, isolated from the Asian malaria vector *Anopheles stephensi* (Hughes et al., 2016). The lineage Sm3 holds eight isolates of contrasting origin, including the Chinese soil isolate DDT-1, capable of using DDT as the sole source of carbon and energy (Pan et al., 2016), as well as clinical isolates like 1162_SMAL (Roach et al., 2015) and AU12-09, isolated from a vascular catheter (Zhang et al., 2013), and environmental isolates like SmF22, Sm32COP, and SmSOFb1, isolated from different manures in France (Bodilis et al., 2016). Cluster Sm2 groups the *S. pavanii* strains, including the type strain DSM_25135^T, isolated from the stems of sugar cane in Brazil (Ramos et al., 2011), together with the clinical isolates ISMMS6 and ISMMS7, that carry mutations conferring quinolone resistance and causing bacteremia (Pak et al., 2015), and strain C11, recovered from pediatric cystic fibrosis patients (Ormerod et al., 2015). Cluster Sm5 includes two strains recovered from soils, ATCC 19867 which was first classified as *Pseudomonas hibiscicola,* and later reclassified as *S. maltophilia* based on MLSA studies (Vasileuskaya-Schulz et al., 2011), and the African strain BurA1, isolated from bulk soil samples collected in sorghum fields in Burkina Faso (Youenou et al., 2015). Cluster Sm9 holds clinical isolates, like 131_SMAL, 424_SMAL, and 951_SMAL (Roach et al., 2015). Its sister group is Sm10. It holds 9 strains of contrasting geographic and ecological provenances, ranging from Chinese soil and plant-associated bacteria like the rice-root endophyte RR10 (Zhu et al., 2012), the grassland-soil tetracycline degrading isolate DT1 (Naas et al., 2008), and strain B418, isolated from a barley rhizosphere and displaying plant-growth promotion properties (Wu et al., 2015), to clinical isolates (22_SMAL, 179_SMAL, 453_SMAL, 517_SMAL) collected and studied in the context of a large genome sequencing project carried out at the University of Washington Medical Center (Roach et al., 2015). Cluster Sm11 tightly groups the well-characterized poplar endophyte R551-3, which is a model plant-growth-promoting bacterium (Ryan et al., 2009; Taghavi et al., 2009; Alavi et al., 2014) and SBo1, cultured from the gut of the olive fruit fly *Bactrocera oleae* (Blow et al., 2016). Cluster Sm 12 contains the environmental strain SKA14 (Adamek et al., 2014), along with the clinical isolates ISMMS3 (Pak et al., 2015) and 860_SMAL (Roach et al., 2015). Sm1, Sm7, and Sm8 each hold a single strain.

The following conclusions can be drawn from this analysis: (*i*) the species-like clusters within the *S. maltophilia sensu lato* (Smsl) clade (**Figure 5**) are enriched in opportunistic human pathogens, when compared with the Smc clusters Sgn1-Sgn4; (*ii*) most Smsl clusters also contain diverse non-clinical isolates isolated from a wide range of habitats, demonstrating the great ecological versatility found even within specific Smsl clusters like Sm3 or Sm10; (*iii*) taken together, these observations strongly suggest that the Smsl species-like clusters are all of environmental origin, with the potential for the opportunistic colonization of diverse human organs. This potential may be particularly high in certain lineages, like in *S. maltophilia sensu stricto* (Sm6) or Sm4a, both enriched in clinical isolates. However, a much denser sampling of genomes and associated phenotypes is required for all clusters to be able to identify statistically sound associations between them.

DISCUSSION

In this study we developed and benchmarked GET_PHYLOMARKERS, an open-source, comprehensive, and easy-to-use software package for phylogenomics and microbial genome taxonomy. Programs like amphora (Wu and Eisen, 2008) or phylosift (Darling et al., 2014) allow users to infer genome-phylogenies from huge genomic and metagenomic datasets by scanning new sequences against a reference database of conserved protein sequences to establish the phylogenetic relationships between the query sequences and database hits. The first program searches the input data for homologs to a set of 31 highly conserved proteins used as phylogenetic markers. Phylosift is more oriented toward the phylogenetic analysis of metagenome community composition and structure. Other approaches have been developed to study large populations of a single species. These are based on the identification of SNPs in sequence reads produced by high-throughput sequencers, using either reference-based or reference-free approaches, and subjecting them to phylogenetic analysis (Timme et al., 2013). The GET_PHYLOMARKERS software suite was designed with the aim of identifying orthologous clusters with optimal attributes for phylogenomic analysis and accurate species-tree inference. It also provides tools to infer phylogenies from pan-genomes, as well as non-supervised learning approaches for the analysis of overall genome relatedness indices (OGRIs) for geno-taxonomic studies of multiple genomes. These attributes make GET_PHYLOMARKERS unique in the field.

It is well-established that the following factors strongly affect the accuracy of genomic phylogenies: (*i*) correct orthology inference; (*ii*) multiple sequence alignment quality; (*iii*) presence of recombinant sequences; (*iv*) loci producing anomalous phylogenies, which may result for example from horizontal gene transfer, differential loss of paralogs between lineages, and (*v*) amount of the phylogenetic signal. GET_PHYLOMARKERS aims to minimize the negative impact of potentially problematic, or poorly performing orthologous clusters, by explicitly considering and evaluating these factors. Orthologous clusters were identified with GET_HOMOLGOUES (Contreras-Moreira and Vinuesa, 2013) because of its distinctive capacity to compute high stringency clusters of single-copy orthologs. In this study we used a combination of BLAST alignment filtering, imposing a high (90%) query coverage threshold, PFAM-domain composition scanning and calculation of a consensus core-genome from the orthologous gene families produced by three clustering algorithms (BDBH, COGtriangles and OMCL) to minimize errors in orthology inference.

Multiple sequence alignments were generated with CLUSTAL-OMEGA (Sievers et al., 2012), a state-of-the-art software under constant development, capable of rapidly aligning hundreds of protein sequences with high accuracy, as reported in recent benchmark studies (Le et al., 2017; Sievers and Higgins, 2018). GET_PHYLOMARKERS generates protein alignments and uses them to compute the corresponding DNA-alignments, ensuring that the codon structure is always properly maintained. Recombinant sequences have been known for a long time to strongly distort phylogenies because they merge independent evolutionary histories into a single lineage. Recombination erodes the phylogenetic signal and misleads classic treeing algorithms, which assume a single underlying history (Schierup and Hein, 2000; Posada and Crandall, 2002; Martin, 2009; Didelot and Maiden, 2010; Pease and Hahn, 2013; Turrientes et al., 2014). Hence, the first filtering step in the pipeline is the detection of putative recombinant sequences using the very fast, sensitive and robust phi(w) statistic (Bruen et al., 2005). The genus *Stenotrophomonas* has been previously reported to have high recombination rates (Yu et al., 2016; Ochoa-Sánchez and Vinuesa, 2017). It is therefore not surprising that the phi(w) statistic detected significant evidence for recombination in up to 47% of the orthologous clusters. The non-recombinant sequences are subsequently subjected to maximum-likelihood phylogenetic inference to identify anomalous gene trees using the non-parametric *kdetrees* statistic (Weyenberg et al., 2014, 2017). The method estimates distributions of phylogenetic trees over the "tree space" expected under the multispecies-coalescent, identifying outlier trees based on their topologies and branch lengths in the context of this distribution. Since this test is applied downstream of the recombination analysis, only a modest, although still significant proportion (14–17%) of outlier trees were detected (**Table 2**). The next step determines the phylogenetic signal content of each gene tree (Vinuesa et al., 2008). It has been previously established that highly informative trees are less prone to get stuck in local optima (Money and Whelan, 2012). They are also required to properly infer divergence at the deeper nodes of a phylogeny (Salichos and Rokas, 2013), and to get reliable estimates of tree congruence and branch support in large concatenated datasets typically used in phylogenomics (Shen et al., 2017). We found that IQT-based searches allowed a significantly more efficient filtering of poorly resolved trees than FastTree. This is likely due to the fact that the former fits more sophisticated models (with more parameters) to better account for among-site rate variation. Under-parameterized and poorly fitting substitution models partly explain the apparent overestimation of bipartition support values done by FastTree. This is also the cause of the poorer performance of FastTree, which finds gene trees that generally have lower ln*L* scores than those found by IQT. A recent comparison of the performance of four fast ML phylogenetic programs using large phylogenomic data sets identified IQT (Nguyen et al., 2015) as the most accurate algorithm. It consistently found the highest-scoring trees. FastTree (Price et al., 2010) was, by large, the fastest program evaluated, although at the price of being the less accurate one (Zhou et al., 2017). This is in line with our findings. We could show that the higher accuracy of IQT is particularly striking when using large concatenated datasets. As stated above, this is largely attributable to the richer choice of models implemented in the former. ModelFinder (Kalyaanamoorthy et al., 2017) selected GTR+ASC+F+R6 model for the concatenated supermatrix, which is much richer in parameters than the GTR+CAT+Gamma20 model fitted by FastTree. The +ASC is an ascertainment bias correction parameter, which should be applied to alignments without constant sites (Lewis, 2001), such as the supermatrices generated by GET_PHYLOMARKERS (see methods). The FreeRate model (+R) generalizes the +G model (fitting a discrete Gamma distribution to model among-site rate variation) by relaxing the assumption of Gamma-distributed rates (Yang, 1995). The FreeRate model typically fits data better than the +G model and is recommended for the analysis of large data sets (Soubrier et al., 2012).

The impact of substitution models in phylogenetics has been extensively studied (Posada and Buckley, 2004). However, the better models implemented in IQT are not the only reason for its superior performance. A key aspect strongly impacting the quality of phylogenomic inference with large datasets is tree-searching. This has been largely neglected in most molecular systematic and phylogenetic studies of prokaryotes (Vinuesa et al., 2008; Vinuesa, 2010; Ochoa-Sánchez and Vinuesa, 2017). Due to the factorial increase of the number of distinct bifurcating topologies possible with every new sequence added to an alignment (Felsenstein, 2004a), searching the tree-space for large datasets is an NP-hard (non-deterministic polynomial-time) problem that necessarily requires heuristic algorithms. This implies that once an optimum is found, there is no way of telling whether it is the global one. The strategy to gain quantitative evidence about the quality of a certain tree is to compare its score in the context of other trees found in searches initiated from a pool of different seed trees. Due to the high dimensionality of the likelihood space, and the strict "hill-climbing" nature of ML tree search algorithms (Felsenstein, 1981), they generally get stuck in local optima (Money and Whelan, 2012). The scores of the best trees found in each search can then be compared in the form of an "ln*L* score profile," as performed in our study. Available software implementations for fast ML tree searching use different branch-swapping strategies to try to escape from early encountered "local optima." IQT implements a more efficient tree-searching strategy than FastTree, based on a combination of hill-climbing and stochastic nearest-neighbor interchange (NNI) operations, always keeping a pool of seed trees, which help to escape local optima (Nguyen et al., 2015). This was evident when the ln*L* score profiles of both programs were compared. IQT found a much better scoring species tree despite the much higher number of independent searches performed with FastTree (50 vs. 1,001) using its most intensive branch-swapping regime. An important finding of our study is the demonstration that the ln*L* search profile of IQT is much shallower than that of FastTree. This suggests that the former finds trees much closer to the potential optimum than the latter. It has been shown that the highest-scoring (best) trees tend to have shorter branches, and overall tree-length, than those stuck in worse local optima (Money and Whelan, 2012). In agreement with this report, the best

species-tree found by IQT has a notoriously shorter total length and significantly shorter edges than those of the best species-tree found by FastTree.

Our extensive benchmark analysis conclusively demonstrated the superior performance of IQT. Based on this evidence, it was chosen as the default search algorithm for GET_PHYLOMARKERS. However, it should be noted that topological differences between the best trees found by both programs were minor, not affecting the composition of the major clades in the corresponding species trees. It is therefore safe to conclude that the reclassification of *Stenotrophomonas* genome sequences proposed in **Table 3** is robust. They are consistently supported by the species-trees estimated with both programs. This result underlines the utility of GET_PHYLOMARKERS to identify misclassified genomes in public sequence repositories, a problem found in many genera (Sangal et al., 2016; Gomila et al., 2017). GET_PHYLOMARKERS is unique in its ability to combine core-genome phylogenomics with ML and parsimony phylogeny estimation from the PGM. In line with other recent studies (Caputo et al., 2015; Tu and Lin, 2016), we demonstrate that pan-genome analyses are valuable in the context of microbial molecular systematics and taxonomy. All genomes found to be misclassified based on the phylogenomic analysis of core-genomes were corroborated by the ML and parsimony analyses of the PGM. Furthermore, the combined evidence gained from these independent approaches consistently revealed that the Smc contains up to 20 monophyletic and strongly supported species-like clusters. These are defined at the cgANIb 95.9% threshold, and include the previously identified genospecies Smc1-Smc4 (Ochoa-Sánchez and Vinuesa, 2017), and up to 13 genospecies within the *S. maltophilia sensu lato* clade. This threshold fits well with the currently favored gANI $>$ 94% cutoff for species delimitation (Konstantinidis and Tiedje, 2005; Richter and Rosselló-Móra, 2009). The consistency among all the different approaches strongly supports the proposed delimitations. We used an unsupervised learning procedure to determine the optimal number of clusters (k) in the cgANDb matrix computed from the 86 Smc genomes analyzed. The ASW goodness of clustering statistic proposed an optimal $k = 19$, which corresponds to a gANI $= 95.5\%$. At this cutoff, 12 (instead of 13) species-like clusters are delimited within the *S. maltophilia sensu lato* clade. This unsupervised learning method therefore seems promising to define the optimal number of clusters in ANI-like matrices using a statistical procedure. However, it should be critically and extensively evaluated in other geno-taxonomic studies to better understand its properties and possible limitations, before being broadly used.

Current models of microbial speciation predict that bacterial species-like lineages should be identifiable by significantly reduced gene flow between them, even when recombination levels are high within species (Cadillo-Quiroz et al., 2012; Shapiro et al., 2012). Such lineages should also display differentiated ecological niches and phenotypes (Koeppel et al., 2008; Shapiro and Polz, 2015). In our previous comprehensive multilocus sequence analysis of species borders in the genus *Stenotrophomonas* (Ochoa-Sánchez and Vinuesa, 2017) we could show that those models fitted our data well. We found highly significant genetic differentiation and marginal gene-flow across strains from sympatric Smc1 and Smc2 lineages, as well as highly significant differences in the resistance profiles of *S. maltophilia sensu lato* isolates vs. Smc1 and Smc2 isolates. We could also show that all three lineages have different habitat preferences (Ochoa-Sánchez and Vinuesa, 2017). The genomic analyses presented in this study for five Smc1 and Smc2 strains, respectively, fully support their separate species status from a geno-taxonomic perspective. Given the recognized importance of gene gain and loss processes in bacterial speciation and ecological specialization (Richards et al., 2014; Caputo et al., 2015; Shapiro and Polz, 2015; Jeukens et al., 2017), as reported also in plants (Gordon et al., 2017), we think that the evidence gained from pan-genome phylogenies is particularly informative for microbial geno-taxonomic investigations. They should be used to validate the groupings obtained by the classical gANI cutoff-based species delimitation procedure (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009) that dominates current geno-taxonomic research. It is well documented that pan-genome-based groupings tend to better reflect ecologically relevant phenotypic differences between groups (Lukjancenko et al., 2010; Caputo et al., 2015; Jeukens et al., 2017). We recommend that future geno-taxonomic studies search for a consensus of the complementary views of genomic diversity provided by OGRIs, core- and pan-genome phylogenies, as performed herein. GET_PHYLOMARKERS is a useful and versatile tool for this task.

In summary, in this study we developed a comprehensive and powerful suite of open-source computational tools for state-of-the art phylogenomic and pan-genomic analyses. Their application to critically analyze the geno-taxonomic status of the genus *Stenotrophomonas* provided compelling evidence that the taxonomically ill-defined *S. maltophilia* complex holds many cryptic species. However, we refrain at this point from making formal taxonomic proposals for them because we have not yet performed the above-mentioned population genetic analyses to demonstrate the genetic cohesiveness of the individual species and their differentiation from closely related ones. This will be the topic of a follow-up work in preparation. We think that comparative genomic analyses designed to identify lineage-specific genetic differences that may underlie niche-differentiation of species are also powerful and objective criteria to delimit species in any taxonomic group (Vinuesa et al., 2005; Ochoa-Sánchez and Vinuesa, 2017).

AUTHOR CONTRIBUTIONS

PV designed the project, wrote the bulk of the code, assembled the genomes, performed the analyses and wrote the paper. LO-S isolated the strains sequenced in this study and performed all wet-lab experiments. BC-M was involved in the original design of the project, contributed code, and set up the docker image. All authors read and approved the final version of the manuscript.

FUNDING

We gratefully acknowledge the funding provided by DGAPA/PAPIIT-UNAM (grants IN201806-2, IN211814 and IN206318) and CONACyT-México (grants P1-60071, 179133 and FC-2105-2-879) to PV, as well as the Fundación ARAID, Consejo Superior de Investigaciones Científicas (grant 200720I038 and Spanish MINECO (AGL2013-48756-R) to BC-M.

ACKNOWLEDGMENTS

We thank Javier Rivera for excellent technical support with wet-lab experiments and José Alfredo Hernández and Víctor del Moral for support with server administration. Jason Steeel from the DNASU Sequencing Core at The Biodesign Institute, Arizona State University, is acknowledged for generating the genome sequences of our samples. Dr. Claudia Silva is thanked for her critical reading of the manuscript. We are thankful to GitHub (https://github.com/), docker (https://hub.docker.com/) and the open-source community at large, for providing great resources for software development.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00771/full#supplementary-material

REFERENCES

Adamek, M., Linke, B., and Schwartz, T. (2014). Virulence genes in clinical and environmental *Stenotrophomas maltophilia* isolates: a genome sequencing and gene expression approach. *Microb. Pathog.* 67–68, 20–30. doi: 10.1016/j.micpath.2014.02.001

Adamek, M., Overhage, J., Bathe, S., Winter, J., Fischer, R., and Schwartz, T. (2011). Genotyping of environmental and clinical *Stenotrophomonas maltophilia* isolates and their pathogenic potential. *PLoS ONE* 6:e27615. doi: 10.1371/journal.pone.0027615

Alavi, P., Starcher, M. R., Thallinger, G. G., Zachow, C., Müller, H., and Berg, G. (2014). *Stenotrophomonas* comparative genomics reveals genes and functions that differentiate beneficial and pathogenic bacteria. *BMC Genomics* 15:482. doi: 10.1186/1471-2164-15-482

Angiuoli, S. V., Gussman, A., Klimke, W., Cochrane, G., Field, D., Garrity, G., et al. (2008). Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. *OMICS* 12, 137–141. doi: 10.1089/omi.2008.0017

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 352–539. doi: 10.1080/10635150600755453

Assih, E. A., Ouattara, A. S., Thierry, S., Cayol, J.-L., Labat, M., and Macarie, H. (2002). *Stenotrophomonas acidaminiphila* sp. nov. a strictly aerobic bacterium isolated from an upflow anaerobic sludge blanket (UASB) reactor. *Int. J. Syst. Evol. Microbiol.* 52, 559–568. doi: 10.1099/00207713-52-2-559

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

Berg, G., and Martinez, J. L. (2015). Friends or foes: can we make a distinction between beneficial and harmful strains of the *Stenotrophomonas maltophilia* complex? *Front. Microbiol.* 6:241. doi: 10.3389/fmicb.2015.00241

Blow, F., Vontas, J., and Darby, A. C. (2016). Draft genome sequence of *Stenotrophomonas maltophilia* SBo1 isolated from *Bactrocera oleae*. *Genome Announc.* 4:e00905-16. doi: 10.1128/genomeA.00905-16

Bodilis, J., Youenou, B., Briolay, J., Brothier, E., Favre-Bonté, S., and Nazaret, S. (2016). Draft genome sequences of *Stenotrophomonas maltophilia* strains Sm32COP, Sm41DVV, Sm46PAILV, SmF3, SmF22, SmSOFb1, and SmCVFa1, isolated from different manures in France. *Genome Announc.* 4:e00841-16. doi: 10.1128/genomeA.00841-16

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., et al. (2015). MeDuSa: a multi-draft based scaffolder. *Bioinformatics* 31, 2443–2451. doi: 10.1093/bioinformatics/btv171

Brooke, J. S. (2012). *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin. Microbiol. Rev.* 25, 2–41. doi: 10.1128/CMR.00019-11

Bruen, T. C., Philippe, H., and Bryant, D. (2005). A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172, 2665–2681. doi: 10.1534/genetics.105.048975

Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., et al. (2012). Patterns of gene flow define species of thermophilic Archaea. *PLoS Biol.* 10:e1001265. doi: 10.1371/journal.pbio.1001265

Caputo, A., Merhej, V., Georgiades, K., Fournier, P.-E., Croce, O., Robert, C., et al. (2015). Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: the *Klebsiella* paradigm. *Biol. Direct* 10:55. doi: 10.1186/s13062-015-0085-2

Chang, Y.-T., Lin, C.-Y., Chen, Y.-H., and Hsueh, P.-R. (2015). Update on infections caused by *Stenotrophomonas maltophilia* with particular attention to resistance mechanisms and therapeutic options. *Front. Microbiol.* 6:893. doi: 10.3389/fmicb.2015.00893

Chen, C., Khaleel, S. S., Huang, H., and Wu, C. H. (2014). Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol. Med.* 9:8. doi: 10.1186/1751-0473-9-8

Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijs.0.054171-0

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287. doi: 10.1126/science.1123061

Conchillo-Solé, O., Yero, D., Coves, X., Huedo, P., Martínez-Servat, S., Daura, X., et al. (2015). Draft genome sequence of *Stenotrophomonas maltophilia* strain UV74 reveals extensive variability within its genomic group. *Genome Announc.* 3:e00611-15. doi: 10.1128/genomeA.00611-15

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Contreras-Moreira, B., Cantalapiedra, C. P., García-Pereira, M. J., Gordon, S. P., Vogel, J. P., Igartua, E., et al. (2017). Analysis of plant pan-genomes and transcriptomes with GET_HOMOLOGUES-EST, a clustering solution for sequences of the same species. *Front. Plant Sci.* 8:184. doi: 10.3389/fpls.2017.00184

Crossman, L. C., Gould, V. C., Dow, J. M., Vernikos, G. S., Okazaki, A., Sebaihia, M., et al. (2008). The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.* 9:R74. doi: 10.1186/gb-2008-9-4-r74

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A., Bik, H. M., and Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243. doi: 10.7717/peerj.243

Daubin, V., Gouy, M., and Perrière, G. (2002). A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome. Res.* 12, 1080–1090. doi: 10.1101/gr.187002

Davenport, K. W., Daligault, H. E., Minogue, T. D., Broomall, S. M., Bruce, D. C., Chain, P. S., et al. (2014). Complete genome sequence of *Stenotrophomonas*

maltophilia type strain 810-2 (ATCC 13637). *Genome Announc.* 2:e00974-14. doi: 10.1128/genomeA.00974-14

Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332–340. doi: 10.1016/j.tree.2009.01.009

Deredjian, A., Alliot, N., Blanchard, L., Brothier, E., Anane, M., Cambier, P., et al. (2016). Occurrence of *Stenotrophomonas maltophilia* in agricultural soils and antibiotic resistance properties. *Res. Microbiol.* 167, 313–324. doi: 10.1016/j.resmic.2016.01.001

Didelot, X., and Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends Microbiol.* 18, 315–322. doi: 10.1016/j.tim.2010.04.002

Dornburg, A., Townsend, J. P., and Wang, Z. (2017). "Maximizing power in phylogenetics and phylogenomics: a perspective illuminated by fungal big data," in *Advances in Genetics*, eds J. P. Townsend and Z. Wang (Cambridge, MA: Academic Press), 1–47.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359

Felsenstein, J. (2004a). *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates INC.

Felsenstein, J. (2004b). PHYLIP (Phylogeny Inference Package). Seattle, DC: University of Washington. Available online at: http://evolution.genetics.washington.edu/phylip/

Finkmann, W., Altendorf, K., Stackebrandt, E., and Lipski, A. (2000). Characterization of N(2)O-producing *Xanthomonas*-like isolates from biofilters as *Stenotrophomonas nitritireducens* sp. nov. Luteimonas mephitis gen. nov. sp. nov. and *Pseudoxanthomonas broegbernensis* gen. nov. sp. nov. *Int. J. Syst. Evol. Microbiol.* 50, 273–282. doi: 10.1099/00207713-50-1-273

Gauthier, J., Vincent, A. T., Charette, S. J., and Derome, N. (2017). Strong genomic and phenotypic heterogeneity in the *Aeromonas sobria* species complex. *Front. Microbiol.* 8:2434. doi: 10.3389/fmicb.2017.02434

Ge, S., and Ge, S. C. (2016). Simultaneous Cr(VI) reduction and Zn(II) biosorption by *Stenotrophomonas* sp. and constitutive expression of related genes. *Biotechnol. Lett.* 38, 877–884. doi: 10.1007/s10529-016-2057-8

Gomila, M., Busquets, A., Mulet, M., García-Valdés, E., and Lalucat, J. (2017). Clarification of taxonomic status within the *Pseudomonas syringae* species group based on a phylogenomic analysis. *Front. Microbiol.* 8:2422. doi: 10.3389/fmicb.2017.02422

Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., et al. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* 8:2184. doi: 10.1038/s41467-017-02292-8

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Guindon, S., Dufayard, J.-F. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Handa, Y., Tazato, N., Nagatsuka, Y., Koide, T., Kigawa, R., Sano, C., et al. (2016). *Stenotrophomonas tumulicola* sp. nov. a major contaminant of the stone chamber interior in the Takamatsuzuka Tumulus. *Int. J. Syst. Evol. Microbiol.* 66, 1119–1124. doi: 10.1099/ijsem.0.000843

Heylen, K., Vanparys, B., Peirsegaele, F., Lebbe, L., and De Vos, P. (2007). *Stenotrophomonas terrae* sp. nov. and *Stenotrophomonas humi* sp. nov. two nitrate-reducing bacteria isolated from soil. *Int. J. Syst. Evol. Microbiol.* 57, 2056–2061. doi: 10.1099/ijs.0.65044-0

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., Le, S. V., and Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1101/153916

Hughes, G. L., Raygoza Garay, J. A., Koundal, V., Rasgon, J. L., and Mwangi, M. M. (2016). Genome sequence of *Stenotrophomonas maltophilia* strain SmAs1, isolated from the asian malaria mosquito *Anopheles stephensi*. *Genome Announc.* 4:e00086-16. doi: 10.1128/genomeA.00086-16.

Jeukens, J., Freschi, L., Vincent, A. T., Emond-Rheault, J.-G., Kukavica-Ibrulj, I., Charette, S. J., et al. (2017). A pan-genomic approach to understand the basis of host adaptation in *Achromobacter*. *Genome Biol. Evol.* 9, 1030–1046. doi: 10.1093/gbe/evx061

Kaiser, S., Biehler, K., and Jonas, D. (2009). A *Stenotrophomonas maltophilia* multilocus sequence typing scheme for inferring population structure. *J. Bacteriol.* 191, 2934–2943. doi: 10.1128/JB.00892-08

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kaparullina, E., Doronina, N., Chistyakova, T., and Trotsenko, Y. (2009). *Stenotrophomonas chelatiphaga* sp. nov. a new aerobic EDTA-degrading bacterium. *Syst. Appl. Microbiol.* 32, 157–162. doi: 10.1016/j.syapm.2008.12.003

Kaufman, L., and Rousseeuw, P. J. (eds.). (1990). *Finding Groups in Data*. Hoboken, NJ: John Wiley & Sons, Inc.

Kim, H.-B., Srinivasan, S., Sathiyaraj, G., Quan, L.-H., Kim, S.-H., Bui, T. P. N., et al. (2010). *Stenotrophomonas ginsengisoli* sp. nov. isolated from a ginseng field. *Int. J. Syst. Evol. Microbiol.* 60, 1522–1526. doi: 10.1099/ijs.0.014662-0

Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., et al. (2008). Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2504–2509. doi: 10.1073/pnas.0712205105

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504–509. doi: 10.1016/j.mib.2007.08.006

Kubatko, L. S., and Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56, 17–24. doi: 10.1080/10635150601146041

Kumar, S., Filipski, A. J., Battistuzzi, F. U., Kosakovsky Pond, S. L., and Tamura, K. (2012). Statistics and truth in phylogenomics. *Mol. Biol. Evol.* 29, 457–472. doi: 10.1093/molbev/msr202

Le, Q., Sievers, F., and Higgins, D. G. (2017). Protein multiple sequence alignment benchmarking through secondary structure prediction. *Bioinformatics* 33:btw840. doi: 10.1093/bioinformatics/btw840

Lee, M., Woo, S. G., Chae, M., Shin, M. C., Jung, H.-M., and Ten, L. N. (2011). *Stenotrophomonas daejeonensis* sp. nov. isolated from sewage. *Int. J. Syst. Evol. Microbiol.* 61, 598–604. doi: 10.1099/ijs.0.017780-0

Lerat, E., Daubin, V., and Moran, N. A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol.* 1:E19. doi: 10.1371/journal.pbio.0000019

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925. doi: 10.1080/106351501753462876

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

Lira, F., Berg, G., and Martínez, J. L. (2017). Double-face meets the bacterial world: the opportunistic pathogen *Stenotrophomonas maltophilia*. *Front. Microbiol.* 8:2190. doi: 10.3389/fmicb.2017.02190

Lira, F., Hernández, A., Belda, E., Sánchez, M. B., Moya, A., Silva, F. J., et al. (2012). Whole-genome sequence of *Stenotrophomonas maltophilia* D457, A clinical isolate and a model strain. *J. Bacteriol.* 194, 3563–3564. doi: 10.1128/JB.00602-12

Lukjancenko, O., Wassenaar, T. M., and Ussery, D. W. (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* 60, 708–720. doi: 10.1007/s00248-010-9717-3

Martin, D. P. (2009). Recombination detection and analysis using RDP3. *Methods Mol. Biol.* 537, 185–205. doi: 10.1007/978-1-59745-251-9_9

Money, D., and Whelan, S. (2012). Characterizing the phylogenetic tree-search problem. *Syst. Biol.* 61:228. doi: 10.1093/sysbio/syr097

Naas, T., Cuzon, G., Villegas, M. V., Lartigue, M. F., Quinn, J. P., and Nordmann, P. (2008). Genetic structures at the origin of acquisition of the β-lactamase blaKPC gene. *Antimicrob. Agents Chemother.* 52, 1257–1263. doi: 10.1128/AAC.01451-07

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood

phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Niu, B., Paulson, J. N., Zheng, X., and Kolter, R. (2017). Simplified and representative bacterial community of maize roots. *Proc. Natl. Acad. Sci. U.S.A.* 114, E2450–E2459. doi: 10.1073/pnas.1616148114

Nourdin-Galindo, G., Sánchez, P., Molina, C. F., Espinoza-Rojas, D. A., Oliver, C., Ruiz, P., et al. (2017). Comparative pan-genome analysis of *Piscirickettsia salmonis* reveals genomic divergences within genogroups. *Front. Cell. Infect. Microbiol.* 7:459. doi: 10.3389/fcimb.2017.00459

Ochoa-Sánchez, L. E., and Vinuesa, P. (2017). Evolutionary genetic analysis uncovers multiple species with distinct habitat preferences and antibiotic resistance phenotypes in the *Stenotrophomonas maltophilia* complex. *Front. Microbiol.* 8:1548. doi: 10.3389/fmicb.2017.01548

Ormerod, K. L., George, N. M., Fraser, J. A., Wainwright, C., and Hugenholtz, P. (2015). Comparative genomics of non-pseudomonal bacterial species colonising paediatric cystic fibrosis patients. *PeerJ* 3:e1223. doi: 10.7717/peerj.1223

Ouattara, A. S., Le Mer, J., Joseph, M., and Macarie, H. (2017). Transfer of *Pseudomonas pictorum* gray and thornton 1928 to genus *Stenotrophomonas* as *Stenotrophomonas pictorum* comb. nov. and emended description of the genus *Stenotrophomonas*. *Int. J. Syst. Evol. Microbiol.* 67, 1894–1900. doi: 10.1099/ijsem.0.001880

Pak, T. R., Altman, D. R., Attie, O., Sebra, R., Hamula, C. L., Lewis, M., et al. (2015). Whole-genome sequencing identifies emergence of a quinolone resistance mutation in a case of *Stenotrophomonas maltophilia* bacteremia. *Antimicrob. Agents Chemother.* 59, 7117–7120. doi: 10.1128/AAC.01723-15

Palleroni, N. J. (2005). "Genux, IX. *Stenotrophomonas* Palleroni and Bradbury 1993," in *Bergey's Manual of Systematic Bacteriology, 2nd Edn.*, eds G. M. Garrity, D. J. Brenner, N. R. Krieg, and J. T. Staley (New York, NY: Springer), 107–115.

Palleroni, N. J., and Bradbury, J. F. (1993). *Stenotrophomonas*, a new bacterial genus for *Xanthomonas maltophilia* (Hugh 1980) Swings et al. 1983. *Int. J. Syst. Bacteriol.* 43, 606–609. doi: 10.1099/00207713-43-3-606

Pan, X., Lin, D., Zheng, Y., Zhang, Q., Yin, Y., Cai, L., et al. (2016). Biodegradation of DDT by *Stenotrophomonas* sp. DDT-1: characterization and genome functional analysis. *Sci. Rep.* 6:21332. doi: 10.1038/srep21332

Parks, M. B., Wickett, N. J., and Alverson, A. J. (2018). Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). *Mol. Biol. Evol.* 35, 80–93. doi: 10.1093/molbev/msx268

Patil, P. P., Midha, S., Kumar, S., and Patil, P. B. (2016). Genome sequence of type strains of genus *Stenotrophomonas*. *Front. Microbiol.* 7:309. doi: 10.3389/fmicb.2016.00309

Pease, J. B., and Hahn, M. W. (2013). More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* 67, 2376–2384. doi: 10.1111/evo.12118

Popescu, A.-A., Huber, K. T., and Paradis, E. (2012). ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28, 1536–1537. doi: 10.1093/bioinformatics/bts184

Posada, D., and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53, 793–808. doi: 10.1080/10635150490522304

Posada, D., and Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J. Mol. Evol.* 54, 396–402. doi: 10.1007/s00239-001-0034-9

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

Ramos, P. L., Van Trappen, S., Thompson, F. L., Rocha, R. C. S., Barbosa, H. R., De Vos, P., et al. (2011). Screening for endophytic nitrogen-fixing bacteria in Brazilian sugar cane varieties used in organic farming and description of *Stenotrophomonas pavanii* sp. nov. *Int. J. Syst. Evol. Microbiol.* 61, 926–931. doi: 10.1099/ijs.0.019372-0

Richards, V. P., Palmer, S. R., Pavinski Bitar, P. D., Qin, X., Weinstock, G. M., Highlander, S. K., et al. (2014). Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol. Evol.* 6, 741–753. doi: 10.1093/gbe/evu048

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Roach, D. J., Burton, J. N., Lee, C., Stackhouse, B., Butler-Wu, S. M., Cookson, B. T., et al. (2015). A year of infection in the intensive care unit: prospective whole genome sequencing of bacterial clinical isolates reveals cryptic transmissions and novel microbiota. *PLoS Genet.* 11:e1005413. doi: 10.1371/journal.pgen.1005413

Rosselló-Móra, R., and Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* 38, 209–216. doi: 10.1016/j.syapm.2015.02.001

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

Ryan, R. P., Monchy, S., Cardinale, M., Taghavi, S., Crossman, L., Avison, M. B., et al. (2009). The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. *Nat. Rev. Microbiol.* 7, 514–525. doi: 10.1038/nrmicro2163

Salichos, L., and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497, 327–331. doi: 10.1038/nature12130

Sánchez-Castro, I., Ruiz-Fresneda, M. A., Bakkali, M., Kämpfer, P., Glaeser, S. P., Busse, H. J., et al. (2017). *Stenotrophomonas bentonitica* sp. nov. isolated from bentonite formations. *Int. J. Syst. Evol. Microbiol.* 67, 2779–2786. doi: 10.1099/ijsem.0.002016

Sandner-Miranda, L., Vinuesa, P., Cravioto, A., and Morales-Espinosa, R. (2018). The genomic basis of intrinsic and acquired antibiotic resistance in the genus Serratia. *Front. Microbiol.* 9:828. doi: 10.3389/fmicb.2018.00828

Sangal, V., Goodfellow, M., Jones, A. L., Schwalbe, E. C., Blom, J., Hoskisson, P. A., et al. (2016). Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. *Sci. Rep.* 6:38392. doi: 10.1038/srep38392

Sassera, D., Leardini, I., Villa, L., Comandatore, F., Carta, C., Almeida, A., et al. (2013). Draft genome sequence of *Stenotrophomonas maltophilia* Strain EPM1, found in association with a culture of the human parasite Giardia duodenalis. *Genome Announc.* 1:e00182-13. doi: 10.1128/genomeA.00182-13

Savory, E. A., Fuller, S. L., Weisberg, A. J., Thomas, W. J., Gordon, M. I., Stevens, D. M., et al. (2017). Evolutionary transitions between beneficial and phytopathogenic *Rhodococcus* challenge disease management. *Elife* 6:e30925. doi: 10.7554/eLife.30925

Schierup, M. H., and Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156, 879–891. Available online at: http://www.genetics.org/content/156/2/879.long

Shapiro, B. J., and Polz, M. F. (2015). Microbial speciation. *Cold Spring Harb. Perspect. Biol.* 7:a018143. doi: 10.1101/cshperspect.a018143

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., et al. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51. doi: 10.1126/science.1218198

Shen, X.-X., Hittinger, C. T., and Rokas, A. (2017). Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* 1:126. doi: 10.1038/s41559-017-0126

Sievers, F., and Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 27, 135–145. doi: 10.1002/pro.3290

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2012). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Soubrier, J., Steel, M., Lee, M. S., Der Sarkissian, C., Guindon, S., Ho, S. Y. W., et al. (2012). The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* 29, 3345–3358. doi: 10.1093/molbev/mss140

Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44, 846–849. doi: 10.1099/00207713-44-4-846

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kampfer, P., Maiden, M. C., et al. (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047. doi: 10.1099/00207713-52-3-1043

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315

Svensson-Stadler, L. A., Mihaylova, S. A., and Moore, E. R. (2012). *Stenotrophomonas* interspecies differentiation and identification by *gyrB* sequence analysis. *FEMS Microbiol. Lett.* 327, 15–24. doi: 10.1111/j.1574-6968.2011.02452.x

Taghavi, S., Garafola, C., Monchy, S., Newman, L., Hoffman, A., Weyens, N., et al. (2009). Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl. Environ. Microbiol.* 75, 748–757. doi: 10.1128/AEM.02239-08

Tange, O. (2011). GNU parallel: the command-line power tool. *USENIX Mag.* 36, 42–47. Available online at: https://www.usenix.org/publications/login/february-2011-volume-36-number-1/gnu-parallel-command-line-power-tool

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Thompson, C. C., Chimetto, L., Edwards, R., a, Swings, J., Stackebrandt, E., and Thompson, F. L. (2013). Microbial genomic taxonomy. *BMC Genomics* 14:913. doi: 10.1186/1471-2164-14-913

Thompson, C. C., Vicente, A. C., Souza, R. C., Vasconcelos, A. T., Vesth, T., Alves, N. Jr., et al. (2009). Genomic taxonomy of Vibrios. *BMC Evol. Biol.* 9:258. doi: 10.1186/1471-2148-9-258

Timme, R. E., Pettengill, J. B., Allard, M. W., Strain, E., Barrangou, R., Wehnes, C., et al. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. enterica inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* 5, 2109–2123. doi: 10.1093/gbe/evt159

Tu, Q., and Lin, L. (2016). Gene content dissimilarity for subclassification of highly similar microbial strains. *BMC Genomics* 17:647. doi: 10.1186/s12864-016-2991-9

Turrientes, M.-C., González-Alba, J.-M., del Campo, R., Baquero, M.-R., Cantón, R., Baquero, F., et al. (2014). Recombination blurs phylogenetic groups routine assignment in *Escherichia coli*: setting the record straight. *PLoS ONE* 9:e105395. doi: 10.1371/journal.pone.0105395

Valdezate, S., Vindel, A., Martín-Dávila, P., Del Saz, B. S., Baquero, F., and Cantón, R. (2004). High genetic diversity among *Stenotrophomonas maltophilia* strains despite their originating at a single hospital. *J. Clin. Microbiol.* 42, 693–699. doi: 10.1128/JCM.42.2.693-699.2003

Vandamme, P., and Peeters, C. (2014). Time to revisit polyphasic taxonomy. *Antonie Van Leeuwenhoek* 106, 57–65. doi: 10.1007/s10482-014-0148-x

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60, 407–438.

Vasileuskaya-Schulz, Z., Kaiser, S., Maier, T., Kostrzewa, M., and Jonas, D. (2011). Delineation of *Stenotrophomonas* spp. by multi-locus sequence analysis and MALDI-TOF mass spectrometry. *Syst. Appl. Microbiol.* 34, 35–39. doi: 10.1016/j.syapm.2010.11.011

Vinuesa, P. (2010). "Multilocus sequence analysis and bacterial species phylogeny estimation," in *Molecular Phylogeny of Microorganisms*, eds A. Oren and R. T. Papke (Caister Academic Press), 41–64. Available online at: http://www.horizonpress.com/phylogeny

Vinuesa, P., and Contreras-Moreira, B. (2015). Robust identification of orthologues and paralogues for microbial pan-genomics using GET_HOMOLOGUES: a case study of pIncA/C plasmids. *Methods Mol. Biol.* 1231, 203–232. doi: 10.1007/978-1-4939-1720-4_14

Vinuesa, P., and Ochoa-Sánchez, L. E. (2015). Complete genome sequencing of *Stenotrophomonas* acidaminiphila ZAC14D2_NAIMI4_2, a multidrug-resistant strain isolated from sediments of a polluted river in Mexico, Uncovers new antibiotic resistance genes and a novel class-II lasso peptide biosynthesis gene cluster. *Genome Announc.* 10:e01433-15. doi: 10.1128/genomeA.01433-15

Vinuesa, P., Rojas-Jiménez, K., Contreras-Moreira, B., Mahna, S. K., Prasad, B. N., Moe, H., et al. (2008). Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent. *Appl. Environ. Microbiol.* 74, 6987–6996. doi: 10.1128/AEM.00875-08

Vinuesa, P., Silva, C., Werner, D., and Martínez-Romero, E. (2005). Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Mol. Phylogenet. Evol.* 34, 29–54. doi: 10.1016/j.ympev.2004.08.020

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963

Weyenberg, G., Huggins, P. M., Schardl, C. L., Howe, D. K., and Yoshida, R. (2014). kdetrees: non-parametric estimation of phylogenetic tree distributions. *Bioinformatics* 30, 2280–2287. doi: 10.1093/bioinformatics/btu258

Weyenberg, G., Yoshida, R., and Howe, D. (2017). Normalizing kernels in the billera-holmes-vogtmann treespace. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 14, 1359–1365. doi: 10.1109/TCBB.2016.2565475

Wolf, A., Fritze, A., Hagemann, M., and Berg, A. (2002). *Stenotrophomonas rhizophila* sp. nov. a novel plant-associated bacterium with antifungal properties. *Int. J. Syst. Evol. Microbiol.* 52, 1937–1944. doi: 10.1099/00207713-52-6-1937

Wu, M., and Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. doi: 10.1186/gb-2008-9-10-r151

Wu, Y., Wang, Y., Li, J., Hu, J., Chen, K., Wei, Y., et al. (2015). Draft genome sequence of *Stenotrophomonas maltophilia* strain B418, a promising agent for biocontrol of plant pathogens and root-knot nematode. *Genome Announc.* 3:e00015-15. doi: 10.1128/genomeA.00015-15

Yang, H.-C., Im, W.-T., Kang, M. S., Shin, D.-Y., and Lee, S.-T. (2006). *Stenotrophomonas koreensis* sp. nov. isolated from compost in South Korea. *Int. J. Syst. Evol. Microbiol.* 56, 81–84. doi: 10.1099/ijs.0.63826-0

Yang, Z. (1995). A space-time process model for the evolution of DNA sequences. *Genetics* 139, 993–1005.

Yi, H., Srinivasan, S., and Kim, M. K. (2010). *Stenotrophomonas panacihumi* sp. nov. isolated from soil of a ginseng field. *J. Microbiol.* 48, 30–35. doi: 10.1007/s12275-010-0006-0

Youenou, B., Favre-Bonté, S., Bodilis, J., Brothier, E., Dubost, A., Muller, D., et al. (2015). Comparative genomics of environmental and clinical *Stenotrophomonas maltophilia* strains with different antibiotic resistance profiles. *Genome Biol. Evol.* 7, 2484–2505. doi: 10.1093/gbe/evv161

Yu, D., Yin, Z., Li, B., Jin, Y., Ren, H., Zhou, J., et al. (2016). Gene flow, recombination, and positive selection in *Stenotrophomonas maltophilia*: mechanisms underlying the diversity of the widespread opportunistic pathogen. *Genome* 59, 1063–1075. doi: 10.1139/gen-2016-0073

Zhang, L., Morrison, M. O., Cuív, P., Evans, P., and Rickard, C. M. (2013). Genome sequence of *Stenotrophomonas maltophilia* Strain AU12-09, isolated from an intravascular catheter. *Genome Announc.* 1:e00195-13. doi: 10.1128/genomeA.00195-13

Zhou, X., Shen, X.-X., Hittinger, C. T., and Rokas, A. (2017). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Mol. Biol. Evol.* 35, 486–503 doi: 10.1101/142323

Zhu, B., Liu, H., Tian, W.-X., Fan, X.-Y., Li, B., Zhou, X.-P., et al. (2012). Genome sequence of *Stenotrophomonas maltophilia* RR-10, isolated as an endophyte from rice root. *J. Bacteriol.* 194, 1280–1281. doi: 10.1128/JB.06702-11

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 © 2018 Vinuesa, Ochoa-Sánchez and Contreras-Moreira. 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 June 2018
doi: 10.3389/fmicb.2018.01377

Study of Bacterial Community Composition and Correlation of Environmental Variables in Rambla Salada, a Hypersaline Environment in South-Eastern Spain

Nahid Oueriaghli[1], *David J. Castro*[1,2], *Inmaculada Llamas*[1,2], *Victoria Béjar*[1,2] *and Fernando Martínez-Checa*[1,2]*

[1] *Microbial Exopolysaccharide Research Group, Department of Microbiology, Faculty of Pharmacy, University of Granada, Granada, Spain,* [2] *Institute of Biotechnology, University of Granada, Granada, Spain*

We studied the bacterial community in Rambla Salada in three different sampling sites and in three different seasons and the effect of salinity, oxygen, and pH. All sites samples had high diversity and richness (Rr > 30). The diversity indexes and the analysis of dendrograms obtained by DGGE fingerprint after applying Pearson's and Dice's coefficient showed a strong influence of sampling season. The Pareto-Lorenz (PL) curves and *Fo* analysis indicated that the microbial communities were balanced and despite the changing environmental conditions, they can preserve their functionality. The main phyla detected by DGGE were *Bacteroidetes* (39.73%), *Proteobacteria* (28.43%), *Firmicutes* (8.23%), and *Cyanobacteria* (5.14%). The majority of the sequences corresponding to uncultured bacteria belonged to *Bacteroidetes* phylum. Within *Proteobacteria*, the main genera detected were *Halothiobacillus* and *Roseovarius*. The environmental factors which influenced the community in a higher degree were the salinity and oxygen. The bacteria belonging to *Bacteroidetes* and *Proteobacteria* were positively influenced by salinity. Nevertheless, bacteria related to *Alpha-* and *Betaproteobacteria* classes and phylum *Firmicutes* showed a positive correlation with oxygen and pH but negative with salinity. The phylum *Cyanobacteria* were less influenced by the environmental variables. The bacterial community composition of Rambla Salada was also studied by dilution-to-extinction technique. Using this method, 354 microorganisms were isolated. The 16S sequences of 61 isolates showed that the diversity was very different to those obtained by DGGE and with those obtained previously by using classic culture techniques. The taxa identified by dilution-to-extinction were *Proteobacteria* (81.92%), *Firmicutes* (11.30%), *Actinobacteria* (4.52%), and *Bacteroidetes* (2.26%) phyla with *Gammaproteobacteria* as predominant class (65.7%). The main genera were: *Marinobacter* (38.85%), *Halomonas* (20.2%), and *Bacillus* (11.2%). Nine of the 61 identified bacteria showed less than 97% sequence identity with validly described species and may well represent new taxa. The number of bacteria in different samples, locations, and seasons were calculated by CARD-FISH, ranging from 54.3 to 78.9% of the total prokaryotic population. In

OPEN ACCESS

Edited by:
Antonio Ventosa,
Universidad de Sevilla, Spain

Reviewed by:
Carmen Portillo,
Universidad Rovira i Virgili, Spain
Mohammad Ali Amoozegar,
University of Tehran, Iran

****Correspondence:***
Fernando Martínez-Checa
fmcheca@ugr.es

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

Received: *15 October 2017*
Accepted: *06 June 2018*
Published: *21 June 2018*

Citation:
Oueriaghli N, Castro DJ, Llamas I, Béjar V and Martínez-Checa F (2018) Study of Bacterial Community Composition and Correlation of Environmental Variables in Rambla Salada, a Hypersaline Environment in South-Eastern Spain. Front. Microbiol. 9:1377. doi: 10.3389/fmicb.2018.01377

conclusion, the dilution-to-extinction technique could be a complementary method to classical culture based method, but neither gets to cultivate the major taxa detected by DGGE. The bacterial community was influenced significantly by the physico-chemical parameters (specially the salinity and oxygen), the location and the season of sampling.

Keywords: biodiversity, bacteria, hypersaline habitat, Rambla Salada, DGGE, dilution-to-extinction methods

INTRODUCTION

In hypersaline environments not only, the high salt concentration limits the biodiversity that inhabits them, they also have, depending on the geographical area, low oxygen concentrations, high or low temperatures, and sometimes alkaline conditions. In addition, factors like pH, pressure, low nutrient availability, solar radiation, the presence of heavy metals, and other toxic compounds, may influence their biodiversity (Ventosa, 2004; Bell and Callaghan, 2012). These environments can be thalassohaline or athalassohaline, the first ones have a marine origin and a qualitative composition similar to sea water. In the second ones, the salt composition is similar to the composition of the surrounding geology, topography, and climatic conditions; this is particularly influenced by the dissolution of mineral deposits (Oren, 2002). Extremely and moderately halophilic microorganisms (bacteria and archaea) predominate in hypersaline environments (Ventosa, 2006; Ventosa et al., 2008; Oren, 2011).

Cultivation-based methods are widely used but often in microbial communities *in situ*, the most abundant members cannot be detected (Rappé and Giovannoni, 2003). During the 1990s and throughout 2000s, the fields of molecular ecology and metagenomics have significantly advanced our knowledge of the genetic diversity and distribution of environmental bacteria. Many new candidate divisions of bacteria and archaea are now recognized due to 16S rRNA sequence-based approaches and environmental metagenomics (Curtis et al., 2002). The numerically dominant bacteria of soils and rhizospheres are the *Alpha-*, *Beta-*, and *Gammaproteobacteria*, *Actinobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Planctomycetes*, *Bacteriodetes,* and *Firmicutes* (da Rocha et al., 2009). These findings reveal the vast disparity between the phyla now recognized by molecular methods and those with cultured representatives.

Novel cultivation strategies are addressing the problem of the uncultivable majority of bacteria and have led to resurgence in microbial cultivation. Emerging strategies roughly follow four major lines: (1) Reformulated and improved culturing media employ dilute nutrient media, non-agar matrices, alternative electron receptors and donors, increased incubation times, and modified atmospheres similar to the bacterial environment (Joseph et al., 2003; Schoenborn et al., 2004; Davis et al., 2005). (2) Diffusion chambers grow bacteria in simulated natural environments (Ferrari et al., 2005, 2008; Gavrish et al., 2008; Ferrari and Gillings, 2009). (3) Microbial signaling molecules are added to growth media that replace natural signaling molecules essential for formation of biofilms and natural microbial consortia (Bruns et al., 2003; Stevenson et al., 2004). (4) High throughput cell separation methods employ either fluorescence activated cell sorting (Zengler et al., 2002, 2005) or dilution-to-extinction (Connon and Giovannoni, 2002) to separate individual bacterial cells to initiate enrichment cultures in dilute natural media.

In order to study the microbial populations in complex habitats has been used several combinations of molecular techniques (Oren, 2003, 2007) among them, the denaturing gradient gel electrophoresis (DGGE) (Muyzer and De Waal, 1994; Muyzer et al., 1996) together with the catalyzed reporter deposition-fluorescence *in situ* hybridization (CARD-FISH) (Wagner et al., 2003; Amann and Bernhard, 2008). Regarding to the use of independent culture techniques, several authors have studied the prokaryotic diversity in athalassohaline habitats, such as the lake Tebenquiche at Salar de Atacama in Chile (Demergasso et al., 2008), lakes of mountains in the Tibetan plateau (Wu et al., 2006), hypersaline alkaline lakes, such as the Mono lake in California (Humayun et al., 2003) and Wadi An Natrun in Egypt (Mesbah et al., 2007), alkaline evaporation ponds at Sua Pan in Botswana (Gareeb and Setati, 2009), Chott El Jerid, a Tunisian hypersaline lake (Abdallah et al., 2016), saline-alkaline soil located in Ararat Plain (Armenia) (Panosyan et al., 2018) and also in different solar salterns, such as Çamalti Saltern, the biggest artificial marine solar saltern in Turkey (Mutlu and Güven, 2015).

Rambla Salada is a clear example of athalassohaline habitat located in Murcia (south-eastern Spain) with special interest by the European Union and it has been declared as a protected wildfowl zone by its regional government (BORM 10/09/1998). Rambla Salada is a course of ∼27 km that connects two areas of great ecological significance, the Natural Park of Sierra Espuña and the river Segura. It is based on an extensive area of sedimentary materials in which underground water emerged, together with the low rainfall, originates a number of wadis and streams. The salinity in Rambla Salada is due to Miocene evaporitic, gypsiferous and marly deposits (Muller and Hsü, 1987) and the most abundant of ions are Na^+ and Cl^-, followed by SO_4^{2-} and Ca^{+2} (Ramírez-Díaz et al., 1995). This habitat has been widely studied by our research group since 2005, and five novel halophilic bacterial species have been described so far: *Idiomarina ramblicola* (Martínez-Cánovas et al., 2004), *Halomonas cerina* (González-Domenech et al., 2008), *Halomonas ramblicola* (Luque et al., 2012a) *Blastomonas quesadae* (Castro et al., 2017), and recently *Roseovarius ramblicola* (Castro et al., 2018). We have also described its halophilic archaeal community by molecular (Oueriaghli et al., 2013) and classical culture techniques (Luque et al., 2012b), its cultivable halophilic bacteria (Luque et al., 2014)

and the distribution of *Halomonas* species by molecular methods (Oueriaghli et al., 2014).

In this work we used the DGGE technique, a culture-independent method, to study the diversity of the bacterial population and to establish the relationships with environmental variables such as pH, oxygen, salinity, and temperature using multivariate statistical analysis and we compare the diversity obtained by this method with those obtained using cultivation methods and incubation conditions including standard culture media (Luque et al., 2014) and dilution-to-extinction method, a technique that improves slow-growing microorganisms recovery or microorganisms that are apparently uncultivable. This unusual comparison in ecological studies enhances the importance of our study. The bacterial community by molecular methods was also quantified.

MATERIALS AND METHODS

Sampling Sites Description

The samples were taken from each of three different points in Rambla Salada at three different periods: June 2006, February, and November 2007. We took 6 samples during each sampling period, 18 samples in total: at Finca La Salina, the samples consisted of soil and watery sediment, from area next to the river (riverbed zone); watery sediment from the water-transfer conduit between the Tagus and Segura rivers (transfer zone) and watery sediment from a saline groundwater spring (upwelling zone) as shown in **Table 1**. The samples obtained from riverbed and transfer zone showed less salinity than those obtained in the upwelling zone. The salt concentration was higher in June 2006, a season characterized by low rainfall.

The watery sediments were taken from the top 15 cm of the silt deposits in each sampling point. The samples were stored in sterile polycarbonate tubes and immediately taken to the laboratory, where they were stored at 4°C until study, always within 24 h. We determined *in situ* pH, oxygen, temperature, and conductivity at each sampling site using an ECmeter (TetraConR 325), which automatically calculates salinity.

Low-Nutrient Medium

We used S3, a low-nutrient medium (Sait et al., 2002, 2006) supplemented with 3 and 15% (w/v) sea-salt solution (Rodriguez-Valera et al., 1981) and pH adjusted to 5.5, 7, and 10. The composition of the medium is the follow: sea-salt stock 30% (w/v) (Rodriguez-Valera et al., 1981), proteose peptone (0.5 g), trace element solution[†] (2 ml), vitamins solution I[‡] (2 ml), vitamins solution II[$] (6 ml), selenite/tungsten[*] solution (2 ml), purified agar (20 g), DI H_20 (1,000 ml).

[†]Trace element solution: HCl 25% (10 ml), $FeCl_2$ x $4H_2O$ (1.5 g), $CoCl_2$ x $6H_2O$ (190 mg), $MnCl_2$ x $4H_2O$ (100 mg), $ZnCl_2$ (70 mg), H_3BO_3 (6 mg), Na_2MoO_4 x $2H_2O$ (36 mg), $NiCl_2$ x $6H_2O$ (24 mg), $CuCl_2$ x $2H_2O$ (2 mg), DI H_2O (1,000 ml).

[‡]Vitamins solution I: 4-aminobenzoate (40 mg), Biotin (10 mg), hemicalcium D+/–pantothenate (100 mg), pyridoxamine hydrochloride (50 mg), thiamin hydrochloride (150 mg), cyanocobalamin (100 mg), DI H_2O (1,000 ml).

[$]Vitamins solution II: nicotinic acid (33 mg), DL-6,8-thioctic acid (10 mg), riboflavin (10 mg), folic acid (4 mg), DI H_2O (1,000 ml).

[*]Selenite/tungstate solution: NaOH (0.5 g), Na_2SeO_3 x $5H_2O$ (3 mg), Na_2WO_4 x $2H_2O$ (4 mg), DI H_2O (1,000 ml).

Dilution-to-Extinction Method

The cultivation method used in this work was based on the dilution-to-extinction approach (Button et al., 1993; Connon and Giovannoni, 2002; Bruns et al., 2003; Koch et al., 2008). For this purpose, serial dilutions from 1 g of soil and/or 1 ml of watery sediment were prepared; soil samples were previously sonicated for 30 s, in 10 ml of S3 medium. The number of microorganisms in each dilution was determined using a Petroff Hausser counting chamber using methylene blue as contrast. A 48-well microtiter plate, containing 490 μl of supplemented S3 medium, was inoculated with 10 μl of the dilution which containing 100 bacteria per milliliter (∼1 bacterium per well) and incubated at 25°C for 30 days. The bacteria grown in the wells were then isolated in Difco™ R2A agar medium plates (Reasoner and Geldreich, 1985) supplemented with 3 and 15% (w/v) sea-salt solution (Rodriguez-Valera et al., 1981).

DNA Extraction and Partial Bacterial 16S rRNA Gen Amplification

Total DNA was extracted, within 24 h since the samples were taken, from 10 g of each of samples using the PowerMax™ Kit for Soil (MO BIO Laboratories) according to the manufacturer's instructions. The primers used for the variable region amplification V1–V3 (∼500 bp) of the 16S rRNA gene of domain *Bacteria* were Bact-8F (5′ AGAGTTTGATCCTGGCTCAG 3′) (Edwards et al., 1989) and the reverse primer Bact-518R (5′ ATTACCGCGGCTGCTGG 3′) (Muyzer et al., 1993). A 40-bp-long GC clamp (5′CGC CCG CCG CGC CCC GCG CCC GTC CCG CCG CCC CCG CCC G-3′) was attached to the 5′ end of the forward primer to obtain PCR fragments adequate for DGGE analysis (Muyzer et al., 1996). PCR reactions were carried out as described by Oueriaghli et al. (2013). An electroforesis in a 1.5 % w/v agarose gel in TBE 1× buffer was used to separate the PCR products (5 μl). Then, the DNA bands were concentrated using Amicon Ultra 0.5 ml 100 K centrifugal filters (Eppendorf, Hamburg, Germany).

DNA from pure culture strains isolated by dilution-to-extinction methods were extracted with X-DNA Extraction Kit from XtremBiotech S.L. (www.xtrembiotech.com) according to the protocol provided by the company. In this case, PCR amplification of 16S rRNA gene was performed as described elsewhere (Castro et al., 2017).

DGGE

A universal mutation detection system, Bio-Rad DCODE™, was used to carry out the denaturing gel gradient electrophoresis. "A 45 to 60% (w/v) (7 M urea and 40% deionized formamide) in a gel with 8% w/v polyacrylamide (37.5:1 acrylamide/bisacrylamide) was used. Each sample (1,000–1,600 ng of PCR products) was loaded onto the gel and run for 20 min at 200 V and again at 100 V at 60°C for 16 h in 1× TAE buffer. The DGGE gel was stained

TABLE 1 | Physico-chemical parameters at the three sites and sampling seasons in which samples were taken.

Sites	Co-ordinates	Sample*	Season	Physico-chemical parameters		
				O_2 (mg L^{-1})	Salinity (g L^{-1})	pH
Site 1: Riverbed zone	38° 07′ 34.44″ N; 1° 07′ 11.13″ W	S1A, S2A	June, 2006	10.2	44.4	6.3
		S4A		4.5	83	6.8
		S1B, S2B	February, 2007	17.2	18.6	8.2
		S4B		14.3	20.2	8.1
		S1C, S2C	November, 2007	13	11.8	8.3
		S4C		7.9	22.2	8.0
Site 2: Transfer zone	38° 07′ 30.23″ N; 1° 07′ 42.22″ W	S3A	June, 2006	1.5	62.1	8.7
		S3B	February, 2007	10.5	34	8.1
		S3C	November, 2007	20.6	29	8.3
Site 3: Upwelling zone	38° 07′ 29.09″ N; 1° 07′ 42.15″ W	S7A, S8A	June, 2006	1	140	7.1
		S7B, S8B	February, 2007	0.6	157.6	6.7
		S7C, S8C	November, 2007	4	151.2	7.2

Data from Luque et al. (2012b); Oueriaghli et al. (2013, 2014).
**Type of sample: S1 and S2, soil sample; S3, S4, S7, and S8 watery sediments.*

with a 1:10,000 dilution of a stock solution of Syber® Gold (Invitrogen-Molecular Probes) for 45 min. DNA bands were visualized with an UV transilluminator (Molecular Imager®, Gel Doc™ XR System, Bio-Rad). All the type bands were excised from the gel, but among those at the same level, only three were chosen at random for cutting. They were resuspended in 10 µl of Milli-Q water and kept overnight at 4 °C. An aliquot of 2 µl of the supernatant was reamplified using the original set of primers (Bact-8F without the GC clamp and Bact-518R) under the conditions described above. The PCR products were purified with Illustra® GFX DNA before being sequenced with an ABI PRISM dye-terminator, cycle-sequencing, ready-reaction kit (Perking-Elmer), and an ABI PRISM 377 sequencer (Perking-Elmer) according to the manufacturer's instructions" (Oueriaghli et al., 2013).

Phylogenetic Study

The variable V1–V3 region sequences of the 16S rRNA gene obtained from DGGE and almost complete sequences of the 16S rRNA gene from dilution-to-extinction isolates were compared using BLASTN program (Altschul et al., 1997) against the GenBank/EMBL/DDBJ database to determine their phylogenetic affiliations. The sequences were then aligned using ClustalW included in MEGA 7 software (Kumar et al., 2016). The phylogenetic relationships of sequences obtained from the DGGE bands with those from the databases (more than 90 % identity) was studied applying neighbor-joining (NJ), maximum likelihood (ML), and maximum parsimony (MP) methods using MEGA 7. The phylogenetic trees and their robustness were tested by bootstrap analysis with 1000 replicates. *Aquifex pyrophilus* Kol5a^T (M83548) was used as out group.

DGGE Fingerprint Analysis

FPquest v.5.101 software (Bio-Rad®) was used to standardize and compare the DGGE band patterns and clustering analysis was performed by determining the Pearson and Dice coefficients. Pearson coefficient takes into consideration the intensity of each band, and the Dice coefficient is based on the presence or absence of bands. Dendrograms relating band-pattern similarities were automatically calculated with UPGMA algorithms (unweighted pair-group method with arithmetic mean). The significance of UPGMA clustering was estimated by calculating the cophenetic correlation coefficients (Sokal and Rohlf, 1962).

Diversity Indexes

Data derived from the presence/absence of bands and from their intensity were exported from the FPquest program to determine the corresponding indexes. We calculated the Shannon-Weaver H' (diversity) and Simpson SI' (dominance) indexes (Shannon and Weaver, 1963; Magurran, 1996) for each DGGE lane using the following equations:

$$H' = \sum_{i=1}^{s} pi \, ln \, pi \qquad SI' = \sum (pi)^2$$

S corresponds to the total number of bands in a DGGE lane and *pi* is calculated as $pi = ni / N$; *ni* is the intensity of each individual band and *N* the sum of the intensities of all the bands in the analized DGGE lane.

Using the equation $Rr = (N2 \times Dg)$ we estimated range-weighted richness index (*Rr*) where *N* represents the total number of bands in each DGGE pattern and *Dg* is the denaturing gradient between the first and last band of each

pattern (Marzorati et al., 2008). As described in our previous work (Oueriaghli et al., 2013), "the evenness of the bacterial community was represented graphically by using the Pareto-Lorenz (*PL*) distribution curves on the basis of the DGGE fingerprints (Marzorati et al., 2008) and the bands in each DGGE lane were assorted according to their intensity. The cumulative normalized numbers of bands are represented along the *x*-axis and their respective cumulative normalized intensities are represented along the *y*-axis. The 45° diagonal represents the perfect evenness of a community in which all the species are equally abundant. To interpret the *PL* curves numerically and calculate the functional organization index of evenness (*Fo*) we plotted the *y*-axis with the vertical 20% *x*-axis line (Marzorati et al., 2008). The resulting values indicate the percentage of total band intensities that constitute 20% of the population. STATGRAPHICS® *plus* v. 3.2 (STSC, Rockville, MD, USA) was used for the analyses of variance (ANOVA). A significance level of 95% ($p < 0.05$) was chosen."

Multivariate Statistical Analysis

The influence of environmental variables upon bacterial diversity was evaluated by applying the detrended correspondence analysis (DCA) (Lepš and Šmilauer, 2003). Also, according to Oueriaghli et al. (2013) "we applied a CCA analysis using CANOCO 4.5 (Biometris, Wageningen, Netherlands). A Monte Carlo test was used to determinate the significance of each axis and to evaluate the influence of the environmental variables upon the overall distribution of bacterial species and their distribution at each sites and sampling seasons. The significance of the CCA axes was tested by means of 999 unrestricted permutations in order to check the null hypothesis that the bacterial profiles were not related to the environmental variables. The effect of any determined environmental variable was chosen according to its significance level ($p < 0.05$) (Salles et al., 2004; Sapp et al., 2007). Ordination biplots are used to represent the effect of environmental variables on bacterial community structure. The environmental factors are represented as arrows: the length of the arrows indicates the relative importance of that environmental factor in explaining the variation in the bacterial communities, whilst the angle between each arrow and the nearest axis indicates the closeness of the relationship between each other."

CARD-FISH

Catalyzed reporter deposition (also known as tyramide signal amplification) *in situ* hybridization (CARD-FISH) was conducted to stain bacterial cells selectively (Pernthaler et al., 2002) and was carry out according to the protocol described by our group in a previous work (Oueriaghli et al., 2013). "The preparation of samples was done by suspending 0.5 g of soil or watery sediments or 0.5 ml of water in 10 ml of PBS buffer 1X. In the case of soil or watery sediment, the resulting suspension was sonicated using the Ultrasonic (Sonorex Digitec) system for 20 min. Supernatant (1 ml of each) were fixed overnight with paraformaldehyde (2%) at 4°C. Cells were filtered and immobilized on 0.2-μm pore-size filters (GTTP, Millipore, Eschobron, Germany), embedded in 0.1% w/v agarose and permeabilized by treatment with 10 mg/ml lysozyme in 50 mM EDTA and 100 mM Tris/HCl for 1 h at 37°C (Pernthaler et al., 2002). Filter sections were cut and hybridized with a mixture of 50 ng/μl of horseradish-peroxide-labeled oligonucleotide probe Eub338 (Amann et al., 1995) (2:20 for each section) and buffer hybridization (Pernthaler et al., 2002) for 2.5 h at 35°C. For signal amplification, we used fluorochrome-labeled tyramide (1 mg/ml; FITC) (Pernthaler et al., 2002). All the microbial cells were counterstained with 4′,6′diamidino-2-phenylindole (DAPI) at a final concentration of 1 mg/ml (Snaidr et al., 1997). For microscopy, filters were first embedded in Citifluor™ (Citifluor Ltd., London, UK), after which the cells were studied under a Leica TCS-SP5 confocal laser scanning microscope (CLSM). Controls with the antisense probe HRP-Non915 were always negative. CARD-FISH stained cells were counted in 20 randomly selected frames using ImageJ software (http://rsb.info.nih.gov/ij/) (Rhasband, 2010)."

RESULTS

Dilution-to-Extinction Approach

Dilution-to-extinction culturing yielded 182 positives microtiter plate wells from a total of 4,800 inoculated wells from which we obtained 354 isolates after re-isolation in R2A medium plates. BLAST searches of the sequences in GenBank revealed that the strains belonged to *Proteobacteria* (81.92%), *Firmicutes* (11.30%), *Actinobacteria* (4.52%), and *Bacteroidetes* (2.26%) phyla with *Gammaproteobacteria* as predominant class (65.7%). Detailed phylogenetic analysis revealed that the isolates shared 94 to 100% 16S rRNA gene sequence identity with the most closely related validly described species. A total of 61 different species were isolated (**Table 2**). The most abundant taxa were *Marinobacter* (38.85%), *Halomonas* (20.2%), and *Bacillus* (11.2%). Nine of the 61 identified bacteria showed less than 97% sequence identity with validly described species and may well represent new taxa.

Analysis of the Bacterial Communities by DGGE Fingerprinting

DGGE fingerprint of samples were compared by using FPquest software. According to the intensity of the bands, the Pearson's coefficient based dendrogram (**Figure 1A**) showed two clusters with a 15% similarity level between them, indicating a low relationship between the two groups of bacteria. The first cluster includes the samples corresponding to June 2006 and November 2007 and the second cluster includes all the samples from February 2007.

Nevertheless, we obtained different results by using the Dice's coefficient, which is based on the presence or absence of bands. As shown in **Figure 1B**, we found two clusters with 36% similarity, including the June 2006 samples in the first one and the February and November 2007 samples in the second one. Samples from the upwelling zone (S7 and S8) in the Dice's dendrogram were grouped together during all three seasons studied.

Phylogenetic Analysis of the DNA Sequences of the DGGE Bands

A total of 67 DGGE bands were successfully reamplified and sequenced (around 500 bp each) from the 90 band classes detected. The identification of phylogenetic neighbors was

TABLE 2 | Taxa isolated by dilution-to-extinction, identified by comparison of their 16S rRNA gene sequences using BLAST.

Isolate	GenBank acc. No.	Closet relatives and acc. No.
PL20	MH266125	*Acinetobacter albensis* ANC 4874 (NR_145641)
25-D-8	MH266126	*Alcanivorax jadensis* T9 (NR_025271)
M35R	MH266127	*Alkalibacillus silvisoli* BM2 (NR_041482)
SR54	MH266128	*Bacillus enclensis* SGD-1123 (KF265350)
D16_1	MH266129	*Bacillus halotolerans* ATCC 25096 (LPVF01000003)
D28_1	MH266130	*Bacillus safensis* kv2 (MH200636)
16SRL	MH266131	*Bacillus siamensis* KCTC 13613 (AJVF01000043)
R63L	MH266132	*Bacillus tequilensis* FJAT-40022 (MG905894)
SR43	MH266133	*Bacillus thioparans* KmS3200909 (MG011570)
M5-FL	MH266134	*Bacillus velezensis* CR-502 (AY603658)
D22-913	MH266135	*Blastomonas quesadae* 912 (KX990274)
R25L	MH266136	*Citreimonas salinaria* CL-SP20 (NR_043303)
SR37F	MH266137	*Erythrobacter litoralis* HMF8222 (KY047411)
D11	MH266138	*Erythrobacter longus* DSM 6997 (NR_041889)
2-C-1	MH266139	*Erythrobacter marinus* HWDM-33 (NR_109054)
D40-857	MH266140	*Flavobacterium jumunjinense* HME7102 (NR_109367)
M35(1)RL	MH266141	*Halobacillus alkaliphilus* SP22 (KX885464)
ML35R	MH266142	*Halobacillus halophilus* 3 (NR_075035)
M45	MH266143	*Halomonas alimentaria* YKJ-16 (NR_025054)
F-5-2	MH266144	*Halomonas fontilapidosi* HMF4436 (KT984005)
SR33	MH266145	*Halomonas gomseomensis* M12 (AM229314)
SR1	MH266146	*Halomonas janggokensis* FMH54 (KX821765)
28-C-6	MH266147	*Halomonas stenophila* N8 (MG563245)
A-10-2	MH266148	*Halomonas ventosae* NRS2HaP1 (LT221212)
L30	MH266149	*Idiomarina abyssalis* MCCC:1A05090 (KM407705)
L30-F	MH266150	*Idiomarina homiensis* MCCC:1A05917 (KM407707.1)
26-F-6	MH266151	*Idiomarina loihiensis* GSL 199 (CP005964)
26f-6	MH266152	*Idiomarina ramblicola* R22 (NR_025806)
L30B	MH266153	*Idiomarina salinarum* MCCC:1A02680 (KM407738)
D14-37	MH266154	*Lysobacter concretionis* Ko07 (NR_041003)
G-6-2	MH266155	*Marinobacter adhaerens* NIOSSK56#15(KY604871)
38CL	MH266156	*Marinobacter algicola* VSW110 (KC534310)
14Df	MH266157	*Marinobacter aquaticus* M6-53 (LT714149)
D-4-4	MH266158	*Marinobacter flavimaris* D6028 (FJ161304)
17-F-15	MH266159	*Marinobacter guineae* M3B (AM503093)
24fL	MH266160	*Marinobacter nanhaiticus* EAR19 (KU320883)
28-C-8	MH266161	*Marinobacter pelagius* KJ-W13 (JQ799110)
27-E-1	MH266162	*Marinobacter persicus* M9B (NR_109110)
D17-1	MH266163	*Marinobacter salsuginis* 11WSA1 (LN794817)
F-7	MH266164	*Marinobacter sediminum* UDC408 (HM031994)
709	MH266165	*Marinobacterium lutimaris* AN9 (NR_116590)
16SF	MH266166	*Microbulbifer salipaludis* SM-1 (NR_025232)
1-5-E	MH266167	*Microbulbifer taiwanensis* CC-LN1-12 (NR_108519)
D26	MH266168	*Modestobacter marinus* AL27 (KU258224)
R14	MH266169	*Palleronia marisminoris* 221-F2 (KJ638254)
SR14FL	MH266170	*Pararhodobacter aggregans* D1-19 (NR_115018)
18RL	MH266171	*Planococcus maritimus* Y67 (KU601234)
M28-R	MH266172	*Planococcus rifietoensis* M8 (CP013659)
L43-F	MH266173	*Pseudoalteromonas issachenkonii* KMM 3549 (CP011030)
L11-R	MH266174	*Pseudomonas psychrotolerans* Pp1 (MH233970)
23-B-5	MH266175	*Pseudoruegeria aquimaris* SW-255 (NR_043932)
M6RL	MH266176	*Rheinheimera pacifica* NBRC 103167 (NR_114230)
6FL	MH266177	*Rheinheimera tangshanensis* RB-213 (JQ361154)
13-B-7	MH266178	*Roseovarius mucosus* DFL-24 (NR_042159)
D-6-6	MH266179	*Roseovarius nubinhibens* SM25 (LT600603)
12-D-7	MH266180	*Roseovarius pacificus* 81-2 (NR_043564)
6-F-3	MH266181	*Roseovarius tolerans* EL-164 (KP723471)
20-B-2	MH266182	*Sediminimonas qiaohouensis* YIM B025 (EU878004)
4-C-4	MH266183	*Thalassospira profundimaris* WP0211 (NR_042766)
6-D-7	MH266184	*Wenzhouxiangella marina* 4S-CH-S3-s2 (MG264256)
M34FL	MH266185	*Winogradskyella arenosi* R60 (NR_041689)

carried out by the BLASTN (Altschul et al., 1997) program against the GenBank/EMBL/DDBJ database containing type strains with validly published prokaryotic names and representatives of uncultured phylotypes. The results are shown in **Table 3**. Clustering was determined using the neighbor-joining, maximum-parsimony, and maximum-likelihood algorithms giving the three, similar topologies and bootstrap values. The neighbor-joining phylogenetic tree (**Figure 2**) shows four main clusters, corresponding to the phyla *Bacteroidetes, Proteobacteria, Firmicutes, and Cyanobacteria.* Within the phyllum *Proteobacteria,* we found four groups, corresponding to *Alpha-*, *Beta-,* and *Gammaproteobacteria* classes. **Figure 3** shows the bacterial diversity found by molecular techniques (DGGE) (a) in comparison to that identified by us in a previous work (Luque et al., 2014) in the same habitats, using classical culture techniques (b) and by a dilution-to-extinction approach (c). The relative abundance of identified sequences during the three seasons in Rambla Salada indicated that *Bacteroidetes* (39.73%) was the most abundant phylum with 46.07, 37.42, and 35.72% of relative abundance percentage in June 2006 and February and November 2007, respectively, followed by *Proteobacteria* (28.43%), showing higher proportion in samples taken in November 2006 (30.23%), and *Firmicutes* (8.23%) and *Cyanobacteria* (5.14%) with the highest proportion in samples taken in February 2007 (9.73 and 9.13%, respectively). A group of unidentified bacteria was also detected. All the sequences related with *Bacteroidetes* were identified as uncultured bacteria, while *Proteobacteria* included sequences related with species belonging to different genera, such as, *Idiomarina, Alteromonas, Halothiobacillus, Ectothiorhodospira, Caulobacter, Porphyrobacter, Azospirillum, Rhodovibrio, Azoarcus, Comamonas, Methylibium, Alkalilimnicola, Salipiger, Roseivivax, Oceanicola, Paracoccus,* and *Roseovarius*.

Analysis of Diversity Indexes

The average number of bands obtained in the DGGE per sample were 25, with a minimum of 14 bands in one sample from the upwelling zone (S8) taken in June 2006 and a maximum of 36

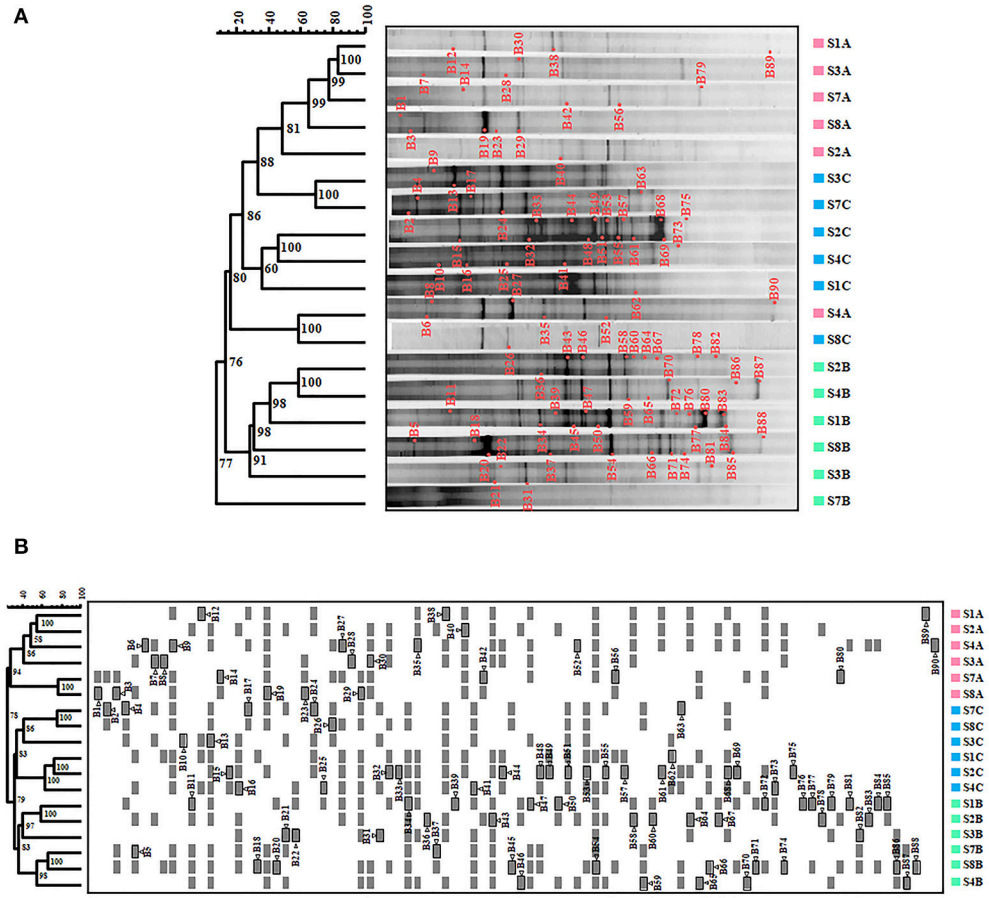


FIGURE 1 | Bacteria in Rambla Salada analyzed by DGGE. **(A)** Pearson coefficient-based analysis. **(B)** Dice coefficient-based analysis. The *scale bar* indicates the percentage of similarity. *Numbers in nodes* represent the cophenetic correlation coefficient values. *Triangles followed by gray shaded squares* bands re-amplified and sequenced to conduct the phylogenetic study. *S1A* to *S8A*, sampling sites in June 2006; *S1B* to *S8B*, sampling sites in February 2007, and *S1C* to *S8C*, sampling site in November 2007.

bands from the same zone (S8) taken in February 2007. The diversity and richness of the bacterial communities depended on the sampling season and showed average *Rr* index values ranging from 41.36 ± 25.95 to 74.14 ± 38.31 (**Table 4**) with high average richness-index values ($Rr > 30$) in the three seasons.

The Shannon-Weaver index values (*H'*) were 2.22 ± 0.34, 2.60 ± 0.36, and 2.63 ± 0.24 for the June 2006, February 2007, and November 2007 samples respectively. The Simpson index (*SI'*), which represents dominance and is inversely proportional to the Shannon-Weaver index, were 0.13 ± 0.07, 0.11 ± 0.06, and 0.09 ± 0.02 respectively. ANOVA analysis ($p < 0.05$) revealed that there were significant differences in the diversity indexes from one season to another (**Table 4**).

The functional organization of the bacterial communities was carried out using the *Fo* index (**Table 4**) and the Pareto-Lorenz distribution curve (**Figure 4**). In the samples corresponding to June 2006, 20% of the bands showed *Fo* values of 60.50, 55.04, and 63.02% in the riverbed, river-transfer and upwelling zones respectively, and the average value of the cumulative band intensities being 59.52% (**Figure 4A**). Twenty per cent of the bands detected in the February 2007 samples represented 60.55, 35.77, and 65.13% (53.81% on average) of the accumulative band intensities (**Figure 4B**), and another 20% of the bands represented 60.55, 52.29, and 55.96% of the accumulative band intensities (52.26% on average) in the November 2007 samples (**Figure 4C**).

Relationships Between the Composition of Bacterial Communities and Environmental Variables

DCA analysis were carried out to determine whether our data were unimodal or linear. DCA analysis showed that the data (2.841) exhibited an unimodal or lineal response to the environmental variables (Lepš and Šmilauer, 2003), so we decided to apply a CCA analysis. **Table 5** shows the eigen values, the cumulative percentage variance in species data and the cumulative variance in the species-environment relationship along the three axes of the CCA analysis.

TABLE 3 | 16S gene sequences obtained from the bands in DGGE and percentages of identity with their closest relatives.

Band no.	GenBank acc. no.	Closet relatives and acc. no.	Identity (%)
B1	MH252018	*Halothiobacillus* sp. XI15 (KT716396)	99
B2	MH252029	*Comamonas* sp. PHD-10 (DQ301787)	97
B4	MH252015	*Halothiobacillus* sp. XI15 (KT716396)	100
B7	MH252009	Uncultured Bacterium EJ97 (FM210936)	100
B10	MH252010	Uncultured *Cyanobacteria* 2P352 (EF106449)	98
B11	MH252011	*Phormidium* sp. HBC9 (EU249125)	99
B12	MH252006	*Halanaerobium alcaliphilum* DSM8275^T (HE582777)	99
B13	MH252004	*Halanaerobium* sp. AN-BI5B (AM157647)	99
B14	MH252005	*Halanaerobium* sp. AN-BI5B (AM157647)	95
B15	MH252001	*Halanaerobium* sp. AN-BI5B (AM157647)	94
B16	MH251989	Uncultured *Bacteroidetes* Ppss_CK72 (JF421218)	96
B17	MH251994	Uncultured *Bacteroidetes* NdSurf156 (FJ753206)	99
B19	MH251998	Uncultured *Cytophagales* LA7-B21N (AF513957)	99
B20	MH251996	Uncultured *Bacteroidetes* BPS_L224 (HQ857722)	98
B21	MH252019	*Halothiobacillus* sp. XI15 (KT716396)	100
B22	MH252016	*Halothiobacillus* sp. XI15 (KT716396)	99
B23	MH252036	*Salipiger mucosus* A3^T (NR_029116)	99
B24	MH252037	*Roseivivax* sp. Y5 (EF177677)	95
B25	MH252017	*Halothiobacillus* sp. HL27 (DQ469574)	100
B26	MH252014	*Halothiobacillus* sp. BI24 (AM157650)	94
B27	MH252013	*Alteromonas* sp. Disko Bay 11 (FJ581928)	99
B29	MH251997	Uncultured *Bacteroidetes* ONGS183 (JX240981)	99
B30	MH252002	*Halanaerobium* sp. AN-BI5B (AM157647)	94
B31	MH252000	*Halanaerobium* sp. AN-BI5B (AM157647)	99
B32	MH251999	*Halanaerobium* sp. AN-BI5B (AM157647)	94
B33	MH252007	Uncultured *Cyanobacteria* 2_RT_2.4.10_A02-T7 (JQ310470)	94
B34	MH252003	*Halanaerobium praevalens* GSL^T (NR_074859)	95
B35	MH252024	*Caulobacter* sp. ARUP UnID 196 (JQ259392)	94
B36	MH252025	*Porphyrobacter* sp. PBW-150 (KC216725)	94
B37	MH252041	*Roseovarius* sp. 2S5-2 (AB114422)	98
B38	MH252040	*Roseovarius* sp. M1S-127 (GU808820)	99
B39	MH252039	*Paracoccus* sp. SS14.12 (KC160783)	94
B40	MH252043	*Roseovarius tolerans* Ekho Lake-172^T (Y11551)	99
B41	MH252012	*Idiomarina* sp. SP96 (FJ404759)	98
B42	MH252038	*Oceanicola* sp. ONGS129 (JX240931)	95
B43	MH251980	Uncultured *Bacteroidetes* Ppss_Ma285 (JF421166)	99
B44	MH251993	Uncultured *Bacteroidetes* SL149 (JX240580)	92
B46	MH251981	Uncultured Bacterium SN135 (EU735690)	97
B48	MH252042	*Roseovarius* sp. 2S5-2 (AB114422)	99
B49	MH252030	*Methylibium fulvum* S32403 (AB649013)	96
B50	MH252028	*Azoarcus* sp. AgN-18 (JN083453)	98
B51	MH252020	*Ectothiorhodospira imhoffii* JA319^T (AM902494)	99
B52	MH252021	*Ectothiorhodospira imhoffii* JA319^T (AM902494)	99
B53	MH252008	Uncultured *Cyanobacteria* FII-TR126 (Q579921)	97
B54	MH251990	Uncultured *Bacteroidetes* JU55122(R4) (HQ706416)	98
B55	MH251987	Uncultured *Bacteroidetes* JU55122(R4) (HQ706416)	99
B56	MH251988	Uncultured *Bacteroidetes* Ppss_CK72 (JF421218)	97
B57	MH251986	Uncultured *Bacteroidetes* Ppss_CK72 (JF421218)	95
B58	MH251982	Uncultured Bacterium SN135 (EU735690)	99
B60	MH252034	Uncultured *Rhodobacteraceae* TDNP_Bbc97_73_2_136 (FJ516820)	99
B61	MH252023	*Azospirillum* sp. LH-CAB12 (HQ717395)	96

(Continued)

TABLE 3 | Continued

Band no.	GenBank acc. no.	Closet relatives and acc. no.	Identity (%)
B62	MH252035	Uncultured *Rhodobacteraceae* TDNP_Bbc97_73_2_136 (FJ516820)	99
B63	MH252031	*Methylibium fulvum* S32403 (AB649013)	94
B64	MH252022	Uncultured *Alfaproteobacteria* MS036A1_A03 (EF701606)	96
B68	MH251992	Uncultured *Bacteroidetes* JU55122(R4) (HQ706416)	99
B69	MH252027	*Rhodovibrio sodomensis* DSM9895^T (FR733704)	99
B71	MH252032	*Alkalilimnicola ehrlichii* MLHE-1 (NR_074775)	94
B72	MH252033	*Alkalilimnicola ehrlichii* MLHE-1 (NR_074775)	94
B73	MH251978	Uncultured Bacterium SN151 (EU735696)	99
B74	MH251977	Uncultured Bacterium SN151 (EU735696)	99
B75	MH251995	Uncultured *Bacteroidetes* HAHS13.3 (HQ396933)	99
B76	MH251983	Uncultured Bacterium SN135 (EU735690)	99
B77	MH251991	Uncultured *Bacteroidetes* JU55122(R4) (HQ706416)	97
B78	MH251984	Uncultured Bacterium SN135 (EU735690)	98
B79	MH251985	Uncultured *Bacteroidetes* BPS_H527 (HQ857687)	98
B80	MH251979	Uncultured *Bacteroidetes* JU5578(R13) (HQ706422)	94
B85	MH252026	*Rhodovibrio salinarum* JA281 (FM177506)	98

Based on the 5% level in a partial Monte Carlo permutation test, the value for oxygen and salinity were significant ($P < 0.05$), providing 75 and 41.2%, respectively, of the total CCA explanatory power. Therefore, the data concerning the environmental factors contributing to the model were ranked in the following order: oxygen, salinity, and finally pH. Species environment correlation for the three axes was more than 0.93, suggesting that bacterial community were strongly correlated with these environmental factors. **Figure 5** shows the influence of the environmental variables upon the diversity of bacterial community in the three seasons studied (**Figure 5A**) and also in relation to the sampling site (**Figure 5B**). **Figure 5A**, in which each environmental variable is represented by an arrow, the projection of any given taxon along an axis shows the level of the variable where the taxon is most abundant. CCA analysis showed a positive correlation with salinity on members belonging to the phylum *Bacteroidetes*, as well as *Gammaproteobacteria* class. Most of the uncultured bacteria also correlated positively with this environmental factor. Nevertheless, all the bacteria related to *Alpha-* and *Betaproteobacteria* class and phylum *Firmicutes* showed a positive correlation with oxygen and pH and negative with salinity. Finally, the phylum *Cyanobacteria* were less influenced by the environmental variables.

As seen in **Figure 5B**, in relation with the location of samples there were two different groups: group A, that included the upwelling zone samples (S7 and S8), showing a positive correlation with salinity and negative correlation with oxygen and pH, and group B, that included riverbed zone and transfer zone samples, showing a positive correlation with pH and oxygen. S3A sample, from the river-transfer conduit, taken in June 2006, is completely different from all the other samples analyzed and thus has not been taken into consideration in our interpretation of the results.

Quantitative Analysis of the Microbial Community as a Whole and its Bacterial Component

Total microbial cells in Rambla Salada detected with DAPI staining were 6.1×10^8, 6.7×10^8, and 7.1×10^8 cells/ml in February 2007, November 2007, and June 2006, respectively (CFU/ml) (**Figure 6B**). CARD-FISH, using universal probes for the *Bacteria* domain, allowed us to know that the bacteria population in Rambla Salada were 3.9×10^8 in June 2006, 4.3×10^8 in November 2007, and 4.8×10^8 cells/ml in February 2007. **Figure 6A** shows a photograph of bacterial cells hybridized with universal bacterial probe (Eub338-HRP-FITC) in samples from riverbed zone in February 2007.

DISCUSSION

Microbial diversity studies has progressed increasingly since the inclusion of molecular techniques that allow to obtain the fingerprinting based on the 16S ribosomal RNA gene analysis, such as the Denaturing Gradient Gel Electrophoresis (DGGE). DGGE fingerprinting is a useful molecular tool due to number of bands and their intensities are related to diversity (Muyzer et al., 1993). Additionally, clustering and ordination methods can include environmental parameters that help to evaluate the impact of different factors on the community composition and structure (Besemer et al., 2005). Nevertheless, in spite of DGGE has significant limitations that should be considered like the possible bias introduced through the DNA extraction, PCR amplification, selection of universal primers and different number of rRNA gene copies (Neufeled and Mohn, 2006) the technique is still used to get an idea of the predominant bacterial population in a habitat (Tang et al., 2016; Yin et al., 2016; Garofalo et al., 2017; Huang et al., 2017; Panosyan et al., 2018). Moreover,

FIGURE 2 | Neighbor-joining phylogenetic tree showing the relationships between the 67 bacterial sequences from the DGGE bands and the most similar sequences retrieved from the GenBank/EMBL/DDBJ database. The scale bar indicates 0.01% divergence. Bootstrap values over 50% are shown in nodes.

DGGE technique permit use the FPquest software to determine several diversity indexes as we explained below.

In this work we studied the bacterial community of Rambla Salada by DGGE and by dilution-to-extintion culture and we compare these data with those obtained by our group with classical culture media (Luque et al., 2012b). Moreover, we determine the correlation of geographical location, season and physic-chemical parameters (oxygen, salinity, and pH) in the bacterial community.

The results from the DGGE fingerprint analyzed by FPquest software revealed time differences (sampling period) in the bacterial community. Dendrogram showed in **Figure 1A**, based on band intensities (Pearson's coefficient) clustered together samples gathered in June 2006 and November 2007 while samples taken in February 2007 were not related. In contrast, Dice's coefficient (**Figure 1B**) showed a clear partition of the three sampling periods in the bacterial community in Rambla Salada. As shown in **Figure 1A**, the diversity profile of bacteria obtained in June 2006 differed significantly from the ones taken in February and June 2007, however, the presence-absence and intensity of the bands were similar for all of these periods. Thus, UPGMA group method carried out with Pearson's coefficient was more sensitive to intensity relative variations of the bands (Huys and Swings, 1999). More conservative rates, as Dice's similarity coefficient (Schwalbach et al., 2005; Hewson et al., 2006a,b) are recommended for genetic fingerprinting like DGGE technique. In spite of the use of a variety of similarity rates, different results in the samples clustering, Pearson and Dice coefficients revealed differences in the bacterial community, depending on the sampling time period. Nevertheless, no relationship could be established between the sample type and a specific taxon.

We can conclude that sample type (soil or watery sediment) or sampling area does not seem to affect the bacterial clustering, with the exception of samples from the upwelling area (S7 and S8) that were clustered together in the different sampling period, as showed in the dendrogram obtained by the Dice index. The upwelling zone is a sulfurous saline water pool with particular characteristics due to the high amount of chlorides and the great sulfate proportion from plaster. This water pool stays constant along the year and it has a regular salt concentration, which might be the reason that the biodiversity found in this area was analogous but different of the other sampling sites (**Table 1**).

The statistical study carried out over the different sampling areas and periods using various diversity indexes, confirmed the relation between the microbial communities found and the sampling periods, being June 2006 community the one with greatest difference. Regarding *Rr*, February and November 2007 had the highest richness values (74.14 ± 38.31 and 58.73 ± 26.86, respectively (**Table 4**). *Rr* values above 30 are typical in microbial diverse environments, according to Marzorati et al. (2008).

When applying Shannon-Waver (H') diversity index (Shannon and Weaver, 1963), which take the number of species present in the study area (species richness) and the number of individual of every species into account, similar results were reported. In this case, February and November 2007 were the seasons with the highest diversity values

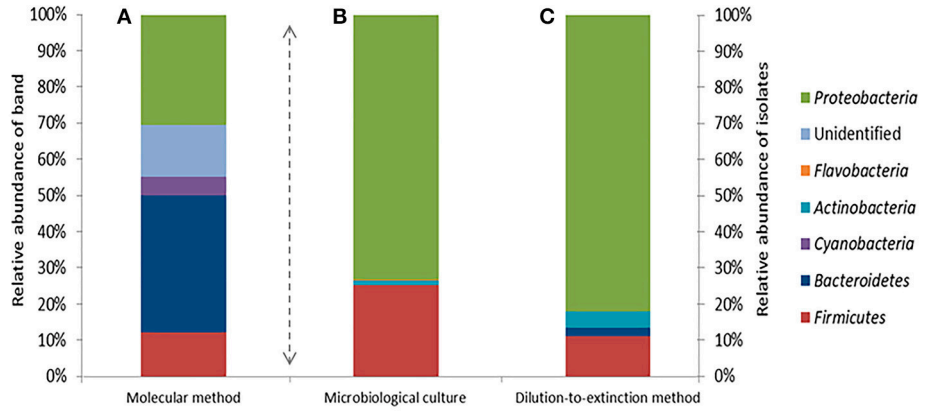


FIGURE 3 | Bacterial diversity determined by analysis of the DGGE bands **(A)**, standard microbiological culture-dependent methods **(B)** (Data from Luque et al., 2014) and dilution-to-extinction method **(C)**.

TABLE 4 | Diversity indexes of bacterial communities in Rambla Salada at the three sites and sampling seasons.

Sampling season	Range-weighted richness (*Rr*)	Functional organization (*Fo*)	Shannon-Weaver (*H'*)	Simpson's (*SI'*)
June-06	41.36 ± 25.95*	59.52%*	2.22 ± 0.34*	0.13 ± 0.07*
February-07	74.14 ± 38.31	53.81%	2.60 ± 0.36	0.11 ± 0.06
November-07	58.73 ± 26.86	52.26%	2.63 ± 0.24	0.09 ± 0.02
LSD‡	0.0365	0.0128	0.0385	0.0251

**Significant differences among the three sampling seasons.*
‡Least-significant difference at $p < 0.05$.

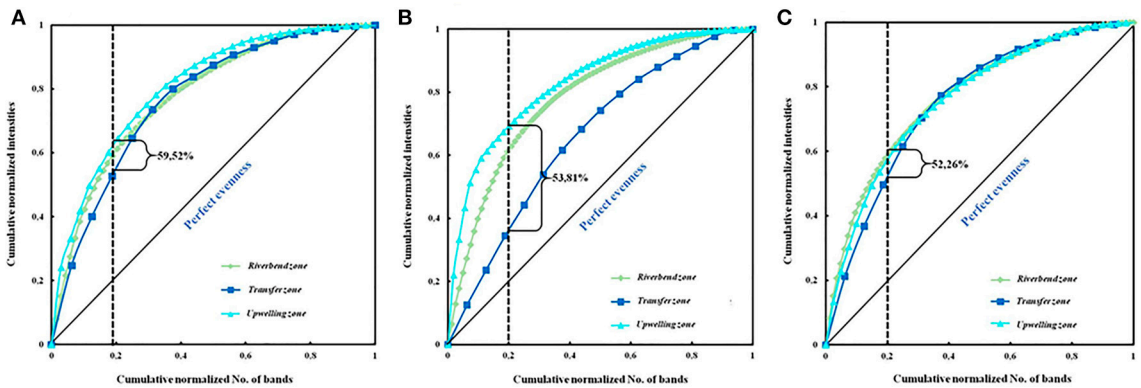


FIGURE 4 | Pareto-Lorenz distribution curves based on the DGGE fingerprints of the bacterial community in the different sampling zones in June 2006 **(A)**, February 2007 **(B),** and November 2007 **(C)**. The *vertical lines* at the 0.2 *x*-axis are plotted to determine the Pareto values (Fo).

(2.60 ± 0.36 and 2.63 ± 0.24, respectively); these values were above 2.5, indicating that a high bacterial diversity was present in Rambla Salada. Luque et al. (2014) obtained similar results analyzing the bacterial diversity using traditional culture-dependent methods in the same areas and time periods.

Simpson index (*SI*) was applied to determine the species dominance within a sample, is based on the number of species and their abundance. Dominance was the lowest in November 2007 (0.09 ± 0.02) and the highest in June 2006 (0.13 ± 0.07). These results confirm that dominance decreases concomitantly with an increase in diversity (Magurran, 1996).

The highest richness and diversity values were detected in samples taken from riverbed zone in February and November 2007 with a low salt and high oxygen concentrations. The high richness and diversity values were probably related to these two environmental factors as confirmed by using CANOCO. In fact, in this analysis, salinity and oxygen were the most significant environmental factors that affected the distribution and composition of the bacterial community.

TABLE 5 | Summary of CANOCO results.

Axes inertia	1	2	3
Eigenvalues	0.277	0.277	0.168
Specie-environment correlations	0.943	0.973	0.930
Cumulative percentage variance of species data	9.7	17.7	23.6
Cumulative percentage variance of species-environment relation	41.2	75.0	100.0

Jiang et al. (2007) demonstrated that the salinity was the dominant factor influencing the composition and community structure of the bacteria population in a hypersaline lake located in Tibet, where the sample with highest salt concentration exhibited the least diversity. Similar results have been previously reported by different authors (Benlloch et al., 2002).

Bacterial community uniformity was determined by Pareto-Lorenz (PL) curves, showing the majority of the analyzed samples, medium *Fo* values (ranging from 40 to 60%) (**Figure 4** and **Table 4**), indicating that the communities were balanced and could, therefore, potentially deal with changing environmental conditions and preserve their functionality (Marzorati et al., 2008). Samples taken in June 2006 and February 2007, however, seem to have similar biogeographical patterns and *Fo* index above 60%, indicating that the bacterial community in this zone and time period was more specialized, due to having a few dominant species, whilst the others were only represented by a few cells (Marzorati et al., 2008).

Ninety DGGE band classes in total were detected, and 67 of them were sequenced. Identified sequences were affiliated to phyla *Bacteroidetes*, *Proteobacteria* (*Alfa-*, *Beta-*, and *Gammaproteobacteria* classes), *Firmicutes,* and *Cyanobacteria* (**Figure 2** and **Table 3**). Different authors have reported that in saline and hypersaline habitats, the most abundant groups of cultivable bacteria are affiliated to *Bacteroidetes*, *Proteobacteria,* and *Firmicutes* phyla (Benlloch et al., 2002; Mouné et al., 2003; Dong et al., 2006; Jiang et al., 2006, 2007; Maturrano et al., 2006; Mesbah et al., 2007; Mutlu et al., 2008; Hollister et al., 2010; Makhdoumi-Kakhki et al., 2012). The cultured bacteria present in soils and sediments, taken around the world from saline and hypersaline environments, mainly belong to the phylum *Proteobacteria* (Hollister et al., 2010; López-López et al., 2010; Swan et al., 2010; Nemergut et al., 2011; Luque et al., 2014). Nevertheless, our results showed differences between the bacterial community determined by molecular techniques and those obtained by culture-dependent methods (dilution-to-extinction and classical methods) (Luque et al., 2014). Maturrano et al. (2006) and (Panosyan et al., 2018) also showed differences between the biodiversity studied by molecular methods and classical methods in different hypersaline habitats.

In this study, 39.73% of the relative abundance of the total bacterial community was affiliated to uncultured taxa belonging to phylum *Bacteroidetes*, indicating that it was the dominant group. These features agreed with the results previously obtained by Makhdoumi-Kakhki et al. (2012), who reported that 59% of the identified sequences in Aran-Bidgol Lake (Iran) were affiliated to phylum *Bacteroidetes*, corresponding the 40% to *Salinibacter* genus. *Salinibacter* has been described as the most abundant taxon in different solar salterns located in Mallorca and Alicante, Spain (Antón et al., 1999, 2000; Rosselló-Mora et al., 2003) and in Çamalti saltern, the biggest artificial marine solar saltern in Turkey (Mutlu and Güven, 2015), as well as in different hypersaline lakes (Maturrano et al., 2006; Mesbah et al., 2007; Mutlu et al., 2008; Makhdoumi-Kakhki et al., 2012). Nevertheless, we could not detect *Salinibacter* in soil or watery sediments in Rambla Salada. All the sequences affiliated to *Bacteroidetes* were identified as non-cultivated species. Therefore, salt concentrations of Rambla Salada (1.1–15.8%, w/v) may not be suitable for *Salinibacter* to growth. Phylum *Bacteroidetes* was widely spread in every sampling area and season, even in low, medium and high salt concentrations samples. These results agreed with those obtained using CANOCO software, in which most of the sequences affiliated to *Bacteroidetes* were in salinity axis (**Figure 5A**). *Bacteroidetes* dominance in saline and hypersaline environments has been previously reported (Antón et al., 1999, 2000; Rosselló-Mora et al., 2003; Makhdoumi-Kakhki et al., 2012) and its presence increases with the increase of salinity (Benlloch et al., 2002; Demergasso et al., 2004, 2008; Jiang et al., 2006).

Proteobacteria was the second most abundant phylum, including members of *Alpha-, Beta-, and Gammaproteobacteria* classes. Wu et al. (2006) demonstrated that when salinity increase, the relative abundance of *Betaproteobacteria* class members decreases, but the relative abundance of *Apha-* and *Gammaproteobacteria* taxa increases. These results agreed with different studies carried out in continental waters (Böckelmanna et al., 2000; Brümmer et al., 2000), estuaries (del Giorgio and Bouvier, 2002; Kirchman et al., 2005; Henriques et al., 2006; Zhang et al., 2006), solar salterns (Benlloch et al., 2002), and a DGGE study in the soda saline crater lake from Isabel island, in the eastern tropical Pacific coast of Mexico (Aguirre-Garrido et al., 2016). In our study, the sequences affiliated to class *Alpha-* and *Betaproteobacteria* showed a negative correlation with salinity and positive with oxygen and pH, whilst class *Gammaproteobacteria* was positively related to salinity. So, the correlation with salinity is the same that above studies in the case of sequences affiliated to *Beta-* and *Gammaproteobacteria,* however the *Alphaproteobacteria* showed a different correlation with salinity. Nevertheless, Langenheder et al. (2003) demonstrated that *Alpha-*, *Beta-,* and *Gammaproteobacteria* classes were more abundant in low saline conditions. Jiang et al. (2007) reported that *Betaproteobacteria* members were the most abundant class, within the *Proteobacteria* phylum, in different lakes located in Tibet, northeast China, with high salt concentration, and no variations of the relative abundance of *Alpha-* and *Gammaproteobacteria* members along the salinity gradients were found.

The relative abundance of phylum *Firmicutes* represents 8.23% of the total bacterial community in Rambla Salada. Different authors reported that *Firmicutes* taxa only represents 11 to 25% of the phylotypes detected in most of the saline and alkaline lakes (Scholten et al., 2005; Jiang et al., 2006; Mesbah et al., 2007). In a study carried out in Aran-Bidgol

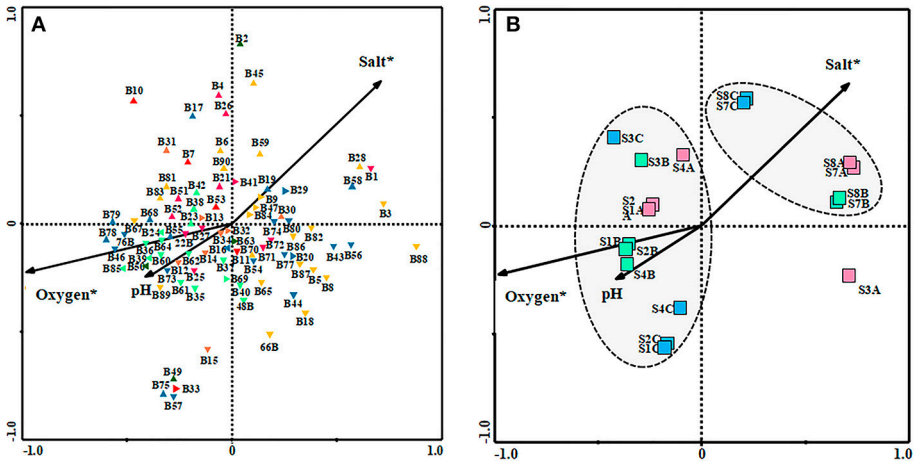


FIGURE 5 | Canonical correspondence analysis (CCA) ordination diagram (biplot) of bacterial species with environmental variables [salt, pH and oxygen] **(A)** and of them with the site and sampling season [*SA* (June 2006), *SB* (February 2007), *SC* (November 2007). (*S1*, *S2*, and *S4*: Riverbend zone; *S3*: River-transfer zone; *S7* and *S8*: Upwelling zone)] **(B)**. Environmental variables are indicated as *arrows*. Environmental variables marked with asterisks are significant ($p < 0.05$). *Bacteroidetes*; *Gammaproteobacteria*; *Alfaproteobacteria*; *Betaproteobacteria*; *Firmicutes*; *Cyanobacteria*; unidentified Bacteria.

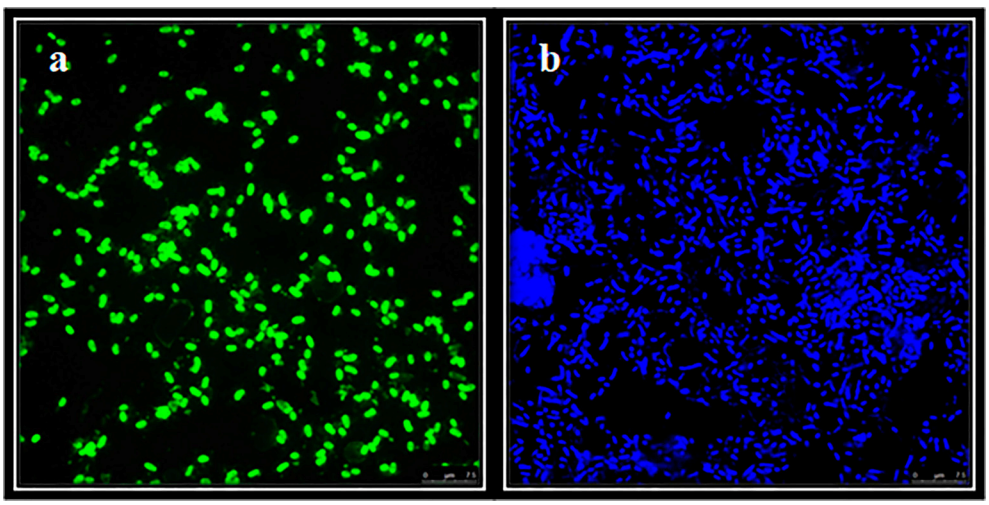


FIGURE 6 | CARD-FISH of samples from Rambla Salada. FITC (total bacteria) counts from riverbed zone in February 2007 **(A)** and DAPI (total population) **(B)**. Image of bacterial cells visualized with a Leica TCS-SP5 CLSM. *Scale bars*, 7.5 μm.

Lake (Iran), 6% of the relative abundance of taxa belonged to phylum *Firmicutes* (Makhdoumi-Kakhki et al., 2012). Nevertheless, in a study carried out in a saline-alkaline soils in Ararat Plain (Armenia) that combined DGGE and culture-dependent method a dominance of *Firmicutes* populations inhabited by moderately halophilic bacilli belonging to the genera *Halobacillus*, *Piscibacillus*, *Bacillus,* and *Virgibacillus* was found (Panosyan et al., 2018). In Rambla Salada, the DGGE bands identified as *Firmicutes* members were negatively related with salinity, except the B30 band (94% of identity with *Halanaerobium*). Jiang et al. (2007) corroborated our results since they reported that the abundance of *Firmicutes* taxa was high in low salt concentration areas but low in high concentration areas.

The relative abundance of phylum *Cyanobacteria* in Rambla Salada was 5.14%. This percentage was also agree with Makhdoumi-Kakhki et al. (2012) who reported that 8% of clones retrieved from Aran-Bidgol Lake (Iran) were affiliated to this phylum.

On the other hand, by using classic culture methods, we determined that the bacterial population in Rambla Salada was mainly affiliated to phylum *Proteobacteria* and *Firmicutes* (72.5 and 25.8%, respectively) (Luque et al., 2014). The genus *Halomonas* and *Marinobacter* within phylum *Proteobacteria* were the predominant (40 and 13% respectively), whilst *Bacillus* genus was the most prevalent within phylum *Firmicutes*. In this work using DGGE, the *Halomonas* genus was not detected; instead this taxon is the easiest isolated in hypersaline

environments. However, *Halomonas* was detected by DGGE in Rambla Salada using specific primers (Oueriaghli et al., 2014).

Extinction culturing methods have evolved from most probable number (MPN) techniques, where highest positive dilutions of MPN or dilution series provide enrichments, or even pure cultures, of abundant but fastidious bacteria that are often undetected by conventional culturing methods (Yang et al., 2016). Nutrient composition also determines the profile of the microbes that can be recovered on artificial media. Excessive nutrient concentrations are atypical in natural environments and can inhibit growth of bacteria adapted to oligotrophic conditions when they are transferred to high nutrient concentrations (Sait et al., 2002; Janssen, 2009). Adaptations to low substrate concentrations in natural environments likely have been a major factor contributing to bacterial unculturability. To address this limitation, novel media modifying the key parameters that simulate the environment (e.g., amount of nutrients, nutrient composition, the presence of trace elements, and pH) have been designed for optimized bacterial growth. Furthermore, nutrient-rich media favoring colonies of fast-growing bacteria can inhibit colony formation of slower growing species and negatively affects the recovery of difficult-to-culture organisms. Therefore, low nutrient media, combined with long incubation periods at relatively low temperatures and pH adjustments can allow bacteria with slow growth rates to form colonies; a great range of bacteria obtained with this approach have proved to be novel organisms (Kenters et al., 2011; Yang et al., 2016).

Dilution-to-extinction approach in combination with S3 low-nutrient-medium (Sait et al., 2002, 2006) and long incubation periods at 25°C, allowed us to obtain 354 isolates, after re-isolation the positives wells in R2A medium plates. The results showed an increment in the cultivability percentages (**Figure 3**) compared to those obtained by classical isolation techniques. By the dilution-to-extinction technique we obtained isolates belonging to *Proteobacteria* (81.9 %), *Firmicutes* (11.3%), *Actinobacteria* (4.5%), and *Bacteroidetes* (2.2%), while Luque et al. (2014) found the same phyla but different percentages (72.5, 25.8, 1.4, and 0.3, respectively). Using dilution-to-extinction technique, we were able to isolate more bacteria belonging to phyla *Actinobacteria* and *Bacteroidetes* than those obtained by classical culture media. However with both techniques the main isolated genera were the same, *Halomonas* and *Marinobacter*. *Cyanobacteria* members, detected by DGGE, were not isolated by any of both methods, probably due to culture conditions used.

Using dilution-to-extinction method, we obtained 9 isolates that showed less than 97% 16S sequence identity and may well represent new taxa. Recently, one of this species belonging to the genus *Blastomonas*, *B. quesadae* has been characterized (Castro et al., 2017).

To determine the number of bacteria at Rambla Salada, we carried out a direct count using CARD-FISH and universal bacterial probe. The results showed the highest counts in February 2007 (4.8×10^8 cells/ml). These results agreed with the ones obtained by DGGE and diversity indexes. The percentage of bacteria found in the different sampling areas and seasons ranged from 54.3 to 78.9% of the total prokaryotic population. These results were similar to that found in a hypersaline deposit in the Canadian High Arctic (Niederberger et al., 2010). The cultivable bacteria counted in Rambla Salada ranged from 10^6 to 10^7 CFU/ml (Luque et al., 2014), which only represents 1% of the total bacterial population detected by molecular methods.

In conclusion, the methods combination used in this study allow us to demonstrate a reliable description of the bacterial populations in the different sampling areas at Rambla Salada, and to find new uncultured taxa so far. Moreover, we shown the correlation of environmental variables with the dominance of several phylum and we demonstrate that the predominant taxa found by DGGE aren't correlated with those isolated by dilution to extinction techniques. Our study highlights and confirms the relevance of this habitat as a diversity reservoir described previously using culture-dependent approach (Luque et al., 2014).

AUTHOR CONTRIBUTIONS

NO performed the experimental DGGE techniques and statistical analysis. DC performed the experimental dilution-to-extinction techniques. IL performed the design of dilution-to-extinction technique and comparative study with culturable methods. VB performed the design of DGGE study and analysis of DGGE results. FM-C performed the design of DGGE study and dilution-to extinction technique and executed the analysis of the results and drafting of the manuscript.

FUNDING

This research was supported by grants from the Dirección General de Investigación Científica y Técnica (CGL2005-05947; CGL2008-02399; CGL2011-25748), Ministerio de Economía y Competitividad and from the Plan Andaluz de Invesigacion (P07-CVI-03150; CVI06226), Spain.

ACKNOWLEDGMENTS

The authors are very grateful to Kadiya Calderón (UGR) for her valuable assistance with the statistical analyses. We thank the Centro de Instrumentacion Cientifica of the University of Granada for their microscopy (CLSM) service. We also thank David Porcel for his suggestions about the instructions concerning the Microscopy (CLSM).

REFERENCES

Abdallah M. B., Karray, F., Mhiri, N., Mei, N., Quéméneur, M., Cayol, J. L., et al. (2016). Prokaryotic diversity in a Tunisian hypersaline lake, Chott El Jerid. *Extremophiles* 20, 125–138. doi: 10.1007/s00792-015-0805-7

Aguirre-Garrido, J. F., Ramírez-Saad, H. C., Toro, N., and Martínez-Abarco, F. (2016). Bacterial diversity in the soda saline crater lake from Isabel island, Mexico. *Microb. Ecol.* 71, 68–77. doi: 10.1007/s00248-015-0676-6

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation

of protein data base search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Amann, R., and Bernhard, M. (2008). Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nat. Microbiol.* 6, 339–348. doi: 10.1038/nrmicro1888

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

Antón, J., Llobet-Brossa, E., Rodríguez-Valera, F., and Amann, R. (1999). Fluorescence*in situ* hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. *Environ. Microbiol.* 1, 517–523. doi: 10.1046/j.1462-2920.1999.00065.x

Antón, J., Rosselló-Mora, R., Rodríguez-Valera, F., and Amann, R. (2000). Extremely halophilic Bacteria in crystallizer ponds from solar salterns. *Appl. Environ. Microbiol.* 66, 3052–3057. doi: 10.1128/AEM.66.7.3052-3057.2000

Bell, E. M., and Callaghan, T. V. (2012). "What are extreme environments and what lives in them?," in *Life at Extremes: Environments, Organisms and Strategies for Survival*, ed E. M. Bell (Wallingford: CAB International), 1–12.

Benlloch, S., López-López, A., Casamayor, E. O., Ovreas, L., Goddard, V., Daae, F. L., et al. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ. Microbiol.* 4, 349–360. doi: 10.1046/j.1462-2920.2002.00306.x

Besemer, K., Moeseneder, M. M., Arrieta, J. M., Herndl, G. J., and Peduzzi, P. (2005). Complexity of bacterial communities in a river-floodplain system (Dnube, Austria). *Appl. Environ. Microbiol.* 71, 609–620. doi: 10.1128/AEM.71.2.609-620.2005

Böckelmanna, U., Manz, W., Neu, T. R., and Szewzyk, U. (2000). Characterization of the microbial community of lotic organic aggregates ('river snow') in the Elbe river of Germany by cultivation and molecular methods. *FEMS Microbiol. Ecol.* 33, 157–170. doi: 10.1111/j.1574-6941.2000.tb00738.x

Brümmer, I. H. M., Fehr, W., and Wagner-Dobler, I. (2000). Biofilm community structure in polluted rivers: abundance of dominant phylogenetic groups over a complete annual cycle. *Appl. Environ. Microbiol.* 66, 3078–3082. doi: 10.1128/AEM.66.7.3078-3082.2000

Bruns, A., Nubel, U., Cypionka, H., and Overmann, J. (2003). Effect of signal compounds and incubation conditions on the culturability of freshwater bacterioplankton. *Appl. Environ. Microbiol.* 69, 1980–1989. doi: 10.1128/AEM.69.4.1980-1989.2003

Button, D. K., Schut, F., Quang, P., Martin, R., and Robertson, B. R. (1993). Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. *Appl. Environ. Microbiol.* 59, 881–891.

Castro, D. J., Cerezo, I., Sampedro, I., and Martínez-Checa, F. (2018). *Roseovarius ramblicola* sp. nov., a moderately halophilic bacterium isolated from saline soil in Spain. *Int. J. Syst. Evol. Microbiol.* 68, 1851–1856. doi: 10.1099/ijsem.0.002744

Castro, D. J., Llamas, I., Béjar, V., and Martínez-Checa, F. (2017). *Blastomonas quesadae* sp. nov., isolated from a saline soil by dilution-to-extinction cultivation. *Int. J. Syst. Evol. Microbiol.* 67, 2001–2007. doi: 10.1099/ijsem.0.001902

Connon, S. A., and Giovannoni, S. J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* 68, 3878–3885. doi: 10.1128/AEM.68.8.3878-3885.2002

Curtis, T. P., Sloan, W. T., and Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10494–10499. doi: 10.1073/pnas.142680199

da Rocha, U. N., van Overbeek, L., and van Elsas, J. D. (2009). Exploration of hitherto-uncultured bacteria from the rhizosphere. *FEMS Microbiol. Ecol.* 69, 313–328. doi: 10.1111/j.1574-6941.2009.00702.x

Davis, K. E., Joseph, S. J., and Janssen, P. H. (2005). Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Appl. Environ. Microbiol.* 71, 826–834. doi: 10.1128/AEM.71.2.826-834.2005

del Giorgio, P. A., and Bouvier, T. C. (2002). Linking the physiologic and phylogenetic successions in free-living bacterial communities along an estuarine salinity gradient. *Limnol. Oceanogr.* 47, 471–486. doi: 10.4319/lo.2002.47.2.0471

Demergasso, C., Casamayor, E. O., Chong, G., Galleguillos, P., Escudero, L., and Pedrós-Alió, C. (2004). Distribution of prokaryotic genetic diversity in athalassohaline lakes of the Atacama Desert, Northern Chile. *FEMS Microbiol. Ecol.* 48, 57–69. doi: 10.1016/j.femsec.2003.12.013

Demergasso, C., Escudero, L., Casamayor, E. O., Chong, G., Balague, V., and Perdrós-Alió, C. (2008). Novelty and spatio-temporal heterogeneity in the bacterial diversity of hypersaline Lake Tebenquiche (Salar de Atacama). *Extremophiles* 12, 491–504. doi: 10.1007/s00792-008-0153-y

Dong, H., Zhang, G., Jiang, H., Yu, B., Leah, R. C., Courtney, R. L., et al. (2006). Microbial diversity in sediment of saline Qinghai Lake, China: linking geochemical controls to microbial ecology. *Microbiol. Ecol.* 51, 65–82. doi: 10.1007/s00248-005-0228-6

Edwards, U., Rogall, T., Blocker, H., Emde, M., and Bottger, 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. doi: 10.1093/nar/17.19.7843

Ferrari, B. C., Binnerup, S. J., and Gillings, M. (2005). Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. *Appl. Environ. Microbiol.* 71, 8714–8720. doi: 10.1128/AEM.71.12.8714-8720.2005

Ferrari, B. C., and Gillings, M. R. (2009). Cultivation of fastidious bacteria by viability staining and micromanipulation in a soil substrate membrane system. *Appl. Environ. Microbiol.* 75, 3352–3354. doi: 10.1128/AEM.02407-08

Ferrari, B. C., Winsley, T., Gillings, M., and Binnerup, S. (2008). Cultivating previously uncultured soil bacteria using a soil substrate membrane system. *Nat. Protoc.* 3, 1261–1269. doi: 10.1038/nprot.2008.102

Gareeb, A. P., and Setati, M. E. (2009). Assessment of alkaliphilic haloarchaeal diversity in Sua pan evaporator ponds in Botswana. *Afr. J. Biotechnol.* 8, 259–267. doi: 10.5897/AJB2009.000-9046

Garofalo, C., Bancalari, E., Milanović, V., Cardinali, F., Osimani, A., Sardaro, M. L. S., et al. (2017). Study of the bacterial diversity of foods: PCR-DGGE versus LH-PCR. *Int. J. Food Microbiol.* 242, 24–36. doi: 10.1016/j.ijfoodmicro.2016.11.008

Gavrish, E., Bollmann, A., Epstein, S., and Lewis, K. (2008). A trap for*in situ* cultivation of filamentous actinobacteria. *J. Microbiol. Methods* 72, 257–262. doi: 10.1016/j.mimet.2007.12.009

González-Domenech, C. M., Martínez-Checa, F., Quesada, E., and Béjar, V. (2008). Halomonas cerina sp. nov., a moderately halophilic, denitrifying, exopolysaccharide-producing bacterium. *Int. J. Syst. Evol. Microbiol.* 58, 803–809. doi: 10.1099/ijs.0.65322-0

Henriques, I. S., Alves, A., Tacao, M., Almeida, A., Cunha, A., and Correia, A. (2006). Seasonal and spatial variability of free-living bacterial community composition along an estuarine gradient (Ria de Aveiro, Portugal). *Estuar. Coast. Shelf Sci.* 68, 139–148. doi: 10.1016/j.ecss.2006.01.015

Hewson, I., Steele, J. A., Capone, D. G., and Fuhrman, J. A. (2006a). Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Mar. Ecol. Prog. Ser.* 311, 67–77. doi: 10.3354/meps311067

Hewson, I., Steele, J. A., Capone, D. G., and Fuhrman, J. A. (2006b). Remarkable heterogeneity in meso and bathypelagic bacterioplankton assemblage composition. *Limnol. Oceanogr.* 51, 1274–1283. doi: 10.4319/lo.2006.51.3.1274

Hollister, E. B., Amanda, S. E., Hammett, A. J. M., Provin, T. L., Wilkinson, H. H., and Gentry, T. J. (2010). Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *ISME J.* 4, 829–838. doi: 10.1038/ismej.2010.3

Huang, W. C., Tsai, H. C., Tao, C. W., and Chen, J. S., Shih, Y. J., Kao, et al. (2017). Approach to determine the diversity of *Legionella* species by nested PCR-DGGE in aquatic environments. *PLoS ONE* 12:e0170992. doi: 10.1371/journal.pone.0170992

Humayun, S. B., Bano, N., and Hollibaugh, J. T. (2003). Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. *Appl. Environ. Microbiol.* 69, 1030–1042. doi: 10.1128/AEM.69.2.1030-1042.2003

Huys, G., and Swings, J. (1999). Evaluation of a fluorescent amplified fragment length polymorphism (FAFLP) methodology for the genotypic discrimination of *Aeromonas* taxa. *FEMS Microbiol. Lett.* 177, 83–92. doi: 10.1111/j.1574-6968.1999.tb13717.x

Janssen, P. H. (2009). Dormant microbes: scouting ahead or plodding along?. *Nature* 458, 831–831. doi: 10.1038/458831a

Jiang, H., Dong, H., Yu, B., Li, Y., Ji, S., Liu, X., 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

Jiang, H., Dong, H., Zhang, G., Yu, B., Chapman, L. R., and Fields, M. W. (2006). Microbial diversity in water and sediment of Lake Chaka: an athalassohaline lake in northwestern China. *Appl. Environ. Microbiol.* 72, 3832–3845. doi: 10.1128/AEM.02869-05

Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A., and Janssen, P. H. (2003). Laboratory cultivation of widespread and previously uncultured soil bacteria. *Appl. Environ. Microbiol.* 69, 7210–7215. doi: 10.1128/AEM.69.12.7210-7215.2003

Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S., and Janssen, P. H. (2011). Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *J. Microbiol. Methods* 84, 52–60. doi: 10.1016/j.mimet.2010.10.011

Kirchman, D. L., Dittel, A. I., Malmstrom, R. R., and Cottrell, M. T. (2005). Biogeography of major bacterial groups in the Delaware Estuary. *Limnol. Oceanogr*. 50, 1697–1706. doi: 10.4319/lo.2005.50.5.1697

Koch, I. H., Gich, F., Dunfield, P. F., and Overmann, J. (2008). Edaphobacter modestus gen. nov., sp nov., and *Edaphobacter aggregans* sp nov., acidobacteria isolated from alpine and forest soils. *Int. J. Syst. Evol. Microbiol*. 58, 1114–1122. doi: 10.1099/ijs.0.65303-0

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

Langenheder, S., Kisand, V., Wikner, J., and Tranvik, L. J. (2003). Salinity as a structuring factor for the composition and performance of bacterioplankton degrading riverine DOC. *FEMS Microbiol. Ecol.* 45, 189–202. doi: 10.1016/S0168-6496(03)00149-1

Lepš, J., and Šmilauer, P. (2003). *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge: Cambridge University Press.

López-López, A., Yarza, P., Richter, M., Suárez-Suárez, A., Antón, J., Niemann, H., et al. (2010). Extremely halophilic microbial communities in anaerobic sediments from a solar saltern. *Environ. Microbiol. Rep.* 2, 258–271. doi: 10.1111/j.1758-2229.2009.00108.x

Luque, R., Béjar, V., Quesada, E., and Llamas, I. (2014). Diversity of halophilic bacteria isolated from Rambla Salada, Murcia (Spain). *Can. J. Microbiol*. 60, 839–846. doi: 10.1139/cjm-2014-0342

Luque, R., Béjar, V., Quesada, E., Martínez-Checa, F., and Llamas, I. (2012a). *Halomonas ramblicola* sp. nov., a moderately halophilic bacterium from Rambla Salada, a mediterranean hypersaline rambla in south-east Spain. *Int. J. Syst. Evol. Microbiol*. 62, 2903–2909. doi: 10.1099/ijs.0.039453-0

Luque, R., González-Domenech, C. M., Llamas, I., Quesada, E., and Béjar, V. (2012b). Diversity of culturable halophilic archaea isolated from Rambla Salada, Murcia (Spain). *Extremophiles* 16, 205–213. doi: 10.1007/s00792-011-0420-1

Magurran, A. E. (1996). *Ecological Diversity and its Measurement*. London: Chapman and Hall.

Makhdoumi-Kakhki, A., Amoozegar, M. A., and Ventosa, A. (2012). *Salinibacter iranicus* sp. nov. and Salinibacter luteus sp. nov., isolated from a salt lake, and emended descriptions of the genus *Salinibacter* and of *Salinibacter ruber*. *Int. J. Syst. Evol. Microbiol*. 62, 1521–1527. doi: 10.1099/ijs.0.031971-0

Martínez-Cánovas, M. J., Béjar, V., Martínez-Checa, F., Páez, R., and Quesada, E. (2004). *Idiomarina fontislapidosi* sp. nov. and *Idiomarina ramblicola* sp. nov., isolated from inland hypersaline habitats in Spain. *Int. J. Syst. Evol. Microbiol.* 54, 1793–1797. doi: 10.1099/ijs.0.63172-0

Marzorati, M., Wittebolle, L., Boon, N., Daffonchio, D., and Verstraete, W. (2008). How to get more out of molecular fingerprints, practical tools for microbial ecology. *Environ. Microbiol.* 10, 1571–1581. doi: 10.1111/j.1462-2920.2008.01572.x

Maturrano, L., Santos, F., Rosselo-Mora, R., and Antón, J. (2006). Microbial Diversity in Maras salterns, a hypersaline environment in the Peruvian Andes. *Appl. Environ. Microbiol.* 72, 3887–3895. doi: 10.1128/AEM.02214-05

Mesbah, N. M., Abou-El-Ela, S. H., and Wiegel, J. (2007). Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. *Microb. Ecol.* 54, 598–617. doi: 10.1007/s00248-006-9193-y

Mouné, S., Caumette, P., Matheron, R., and Willison, J. C. (2003). Molecular sequence analysis of prokaryotic diversity in the anoxic sediments underlying cyanobacterial mats of two hypersaline ponds in Mediterranean salterns. *FEMS Microbiol. Ecol*. 44, 117–130. doi: 10.1016/S0168-6496(03)00017-5

Muller, D. W., and Hsü, K. J. (1987). Event stratigraphy and paleoceanography in the Fortuna basin (Southeast Spain): a scenario for the Messinian salinity crisis. *Paleoceanogr* 2, 679–696. doi: 10.1029/PA002i006p00679

Mutlu, M. B., and Güven, K. (2015). Bacterial diversity in Çamalti Saltern, Turkey. *Pol. J. Microbiol.* 64, 37–45.

Mutlu, M. B., Martínez-García, M., Santos, F., Peña, A., Guven, K., and Antón, J. (2008). Prokaryotic diversity in Tuz Lake, a hypersaline environment in Inland Turkey. *FEMS Microbiol. Ecol.* 65, 474–483. doi: 10.1111/j.1574-6941.2008.00510.x

Muyzer, G., and De Waal, E. C. (1994). "Determination of the genetic diversity of microbial communities using DGGE analysis of PCR-amplified 16S rDNA," in *Microbial Mats: Structure, Development and Environmental Significance*, ed L. J. Stal and P. Caumette (Heidelberg: Springer-Verlag), 207–214.

Muyzer, G., De Waal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by 16 denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59, 695–700.

Muyzer, G., Hottentra-Ger, S., Teske, A., and Wawer, C. (1996). "Denaturing gradient gel electrophoresis of PCR-amplified 16S rDNA a new molecular approach to analyse the genetic diversity of mixed microbial communities," in *Molecular Microbial Ecology Manual*, ed A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (Dordrecht: Kluwer Academic Publishers), 1–23.

Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., et al. (2011). Global patterns in the biogeography of bacterial taxa. *Environ. Microbiol.* 13, 135–144. doi: 10.1111/j.1462-2920.2010.02315.x

Neufeled, J. D., and Mohn, W. W. (2006). "Assessment of microbial phylogenetic diversity based on environmental nucleic acid," in *Molecular Identification Systematic and Population Structure of Prokaryotes*, ed E. Stackebrandt (Berlin: Springer-Verlag), 220–259.

Niederberger, T. D., Perreault, N. N., Tille, S., Lollar, B. S., Lacrampe-Couloume, G., Andersen, D., et al. (2010). Microbial characterization of a subzero, hypersaline methane seep in the Canadian High Arctic. *Int. J. Syst. Evol. Microbiol.* 4, 1326–1339. doi: 10.1038/ismej.2010.57

Oren, A. (2002). *Halophilic Microorganisms and Their Environments*. Dordrecht: Kluwer Academic.

Oren, A. (2003). Molecular ecology of extremely halophilic *Archaea* and *Bacteria*. *FEMS Microbiol. Ecol.* 39, 1–7. doi: 10.1111/j.1574-6941.2002.tb00900.x

Oren, A. (2007). "Biodiversity in highly saline environments," in *Physiology and Biochemistry of Extremophiles*, eds C. Gerday and N Glansdorff (Washington, DC: ASM Press), 223–231.

Oren, A. (2011). Thermodynamic limits to microbial life at high salt concentrations. *Environ. Microbiol.* 13, 1908–1923. doi: 10.1111/j.1462-2920.2010.02365.x

Oueriaghli, N., Béjar, V., Quesada, E., and Martínez-Checa, F. (2013). Molecular ecology techniques reveal both spatial and temporal variations in the diversity of archaeal communities within the athalassohaline environment of Rambla Salada, Spain. *Microl. Ecol*. 66, 297–311. doi: 10.1007/s00248-013-0176-5

Oueriaghli, N., González-Domenech, C. M., Martínez-Checa, F., Muyzer, G., Ventosa, A., Quesada, E., et al. (2014). Diversity and distribution of *Halomonas* in Rambla Salada, a hypersaline environment in the southeast of Spain. *FEMS Microbiol. Ecol*. 87, 460–474. doi: 10.1111/1574-6941.12237

Panosyan, H., Hakobyan, A., Birkeland, N. K., and Trchounian, A. (2018). Bacilli community of saline-alkaline soils from the Ararat Plain (Armenia) assessed by molecular and culture-based methods. *Syst. Appl. Microbiol.* 41, 232–240. doi: 10.1016/j.syapm.2017.12.002

Pernthaler, A., Pernthaler, J., and Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* 68, 3094–3101. doi: 10.1128/AEM.68.6.3094-3101.2002

Ramírez-Díaz, L., Vidal-Abarca, M. R., Calvo, J. F., Suarez, M. L., Palazón, J. A., Esteve, M. A., et al. (1995). *Bases Ecológicas Para la Delimitación, Ordenación y Gestión del Paisaje Protegido de las Ramblas de Ajauque y Salada, Vol 1 and 2*, Murcia: Consejería de Medio Ambiente, Comunidad Autónoma de la Región de Murcia, Spain.

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

Reasoner, D. J., and Geldreich, E. E. (1985). A new medium for the enumeration and subculture of bacteria from potable water. *Appl. Environ. Microbiol.* 49, 1–7.

Rhasband, W. (2010). *The ImageJ (Image processing and Analysis in Java) Version 1.43u*. National Institutes of Health, USA. Available online at: http://rsb.info.nih.gov/ij/

Rodriguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1981). Characteristics of heterotrophic bacterial populations in hypersaline environments of different salt concentrations. *Microb. Ecol.* 7, 235–243. doi: 10.1007/BF02010306

Rosselló-Mora, R., Lee, N., Antón, J., and Wagner, M. (2003). Substrate uptake in extremely halophilic microbial communities revealed by microautoradiography and fluorescence*in situ* hybridization. *Extremophiles* 7, 409–413. doi: 10.1007/s00792-003-0336-5

Sait, M., Davis, K. E., and Janssen, P. H. (2006). Effect of pH on isolation and distribution of members of subdivision 1 of the phylum *Acidobacteria* occurring in soil. *Appl. Environ. Microbiol.* 72, 1852–1857. doi: 10.1128/AEM.72.3.1852-1857.2006

Sait, M., Hugenholtz, P., and Janssen, P. H. (2002). Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ. Microbiol.* 4, 654–666. doi: 10.1046/j.1462-2920.2002.00352.x

Salles, J. F., Van Veen, J. A., and Van Elsas, J. D. (2004). Multivariate analyses of *Burkholderia* species in soil: effect of crop and land use history. *Appl. Environ. Microbiol.* 70, 4012–4020. doi: 10.1128/AEM.70.7.4012-4020.2004

Sapp, M., Wichels, A., Wiltshire, K., and Gerdts, G. (2007). Bacterial community dynamics during the winter-spring transition in the North Sea. *FEMS Microbiol. Ecol.* 59, 622–637. doi: 10.1111/j.1574-6941.2006.00238.x

Schoenborn, L., Yates, P. S., Grinton, B. E., Hugenholtz, P., and Janssen, P. H. (2004). Liquid serial dilution is inferior to solid media for isolation of cultures representative of the phylum-level diversity of soil bacteria. *Appl. Environ. Microbiol.* 70, 4363–4366. doi: 10.1128/AEM.70.7.4363-4366.2004

Scholten, J. C., Joye, S. B., Hollibaugh, J. T., and Murrell, J. C. (2005). Molecular analysis of the sulfate reducing and archaeal community in a meromictic soda lake (Mono Lake, California) by targeting 16S rRNA, mcrA, apsA, and dsrAB genes. *Microbiol. Ecol.* 50, 29–39. doi: 10.1007/s00248-004-0085-8

Schwalbach, M., Brown, M., and Fuhrman, J. (2005). Impact of light on marine bacterioplankton community structure. *Aquat. Microbiol. Ecol.* 39, 235–245. doi: 10.3354/ame039235

Shannon, C. E., and Weaver,W. (1963). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Snaidr, J., Amann, R., Huber, I., Ludwig, W., and Schleifer, K. H. (1997). Phylogenetic analysis and*in situ* identification of bacteria in activated sludge. *Appl. Environ. Microbiol.* 63, 2884–2896.

Sokal, R. R., and Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon* 11, 33–40. doi: 10.2307/1217208

Stevenson, B. S., Eichorst, S. A., Wertz, J. T., Schmidt, T. M., and Breznak, J. A. (2004). New strategies for cultivation and detection of previously uncultured microbes. *Appl. Environ. Microbiol.* 70, 4748–4755. doi: 10.1128/AEM.70.8.4748-4755.2004

Swan, B. K., Ehrhardt, C. J., Reifel, K. M., Moreno, L. I., and Valentine, D. L. (2010). Archaeal and bacterial communities respond differently to environmental gradients in anoxic sediments of a California hypersaline lake, the Salton Sea. *Appl. Environ. Microbiol.* 76, 757–768. doi: 10.1128/AEM.02409-09

Tang, W., Gong, J., Wu, L., Li, Y., Zhang, M., and Zeng, X. (2016). DGGE diversity of manganese mine samples and isolation of a *Lysinibacillus* sp. efficient in removal of high Mn (II) concentrations. *Chemosphere* 165, 277–283. doi: 10.1016/j.chemosphere.2016.08.134

Ventosa, A. (2004). *Halophilic Microorganisms*. Berlin; Heidelberg: Germany: Springer-Verlag.

Ventosa, A. (2006). "Unusual micro-organisms from unusual habitats: hypersaline environments," in *Prokaryotic Diversity Mechanism and Significance*, eds N. A. Logan, H. M. Lappin-Scott, and P. C. F. Oyston (Cambridge; London: Cambridge University Press), 223–254.

Ventosa, A., Mellado, E., Sanchez-Porrro, C., and Marquez, M. C. (2008). "Halophilic and halotolerant microorganisms from soils," in *Microbiology of Extreme Soils*, eds P. Dion and C. S. Nautiyal (Berlin: Springer-Verlag), 87–115.

Wagner, M., Horn, M., and Daims, H. (2003). Fluorescence*in situ* hybridisation for the identification and characterisation of prokaryotes. *Curr. Opin. Microbiol.* 6, 302–309. doi: 10.1016/S1369-5274(03)00054-7

Wu, Q. L., Zwart, G., Schauer, M., Kamst-van Agterveled, M. P., and Hahn, H. V. (2006). Bacterioplankton community composition along a salinity gradient of sixteen high-mountain Lakes located on the Tibetan plateau, China. *Appl. Environ. Microbiol.* 72, 5478–5485. doi: 10.1128/AEM.00767-06

Yang, S. J., Kang, I., and Cho, J. C. (2016). Expansion of cultured bacterial diversity by large-scale dilution-to-extinction culturing from a single seawater sample. *Microb. Ecol.* 71, 29–43. doi: 10.1007/s00248-015-0695-3

Yin, Y., Song, W., Gu, J., Zhang, K., Qian, X., Zhang, X., et al. (2016). Effects of copper on the abundance and diversity of ammonia oxidizers during dairy cattle manure composting. *Bioresour. Technol.* 221, 181–187. doi: 10.1016/j.biortech.2016.09.016

Zengler, K., Toledo, G., Rappé, M., Elkins, J., Mathur, E. J., Short, J. M., et al. (2002). Cultivating the uncultured. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15681–15686. doi: 10.1073/pnas.252630999

Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., et al. (2005). High-throughput cultivation of microorganisms using microcapsules. *Methods Enzymol.* 397, 124–130. doi: 10.1016/S0076-6879(05)97007-9

Zhang, Y., Jiao, N. Z., Cottrell, M. T., and Kirchman, D. L. (2006). Contribution of major bacterial groups to bacterial biomass production along a salinity gradient in the South China Sea. *Aquat. Microb. Ecol.* 43, 233–241. doi: 10.3354/ame043233

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 Oueriaghli, Castro, Llamas, Béjar and Martínez-Checa. 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: 23 January 2018
doi: 10.3389/fmicb.2018.00022

Biogeographic Differences in the Microbiome and Pathobiome of the Coral *Cladocora caespitosa* in the Western Mediterranean Sea

Esther Rubio-Portillo[1], Diego K. Kersting[2,3], Cristina Linares[3], Alfonso A. Ramos-Esplá[4] and Josefa Antón[1]**

[1] Department of Physiology, Genetics and Microbiology, University of Alicante, Alicante, Spain, [2] Working Group on Geobiology and Anthropocene Research, Institute of Geological Sciences, Freie Universität Berlin, Berlin, Germany, [3] Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain, [4] Marine Research Centre of Santa Pola, University of Alicante, Alicante, Spain

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Sarah Annalise Gignoux-Wolfsohn, Rutgers, The State University of New Jersey, United States
Christina A. Kellogg, United States Geological Survey, United States

***Correspondence:**
Esther Rubio-Portillo
esther.portillo@ua.es
Josefa Antón
anton@ua.es

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

Received: *07 October 2017*
Accepted: *05 January 2018*
Published: *23 January 2018*

Citation:
Rubio-Portillo E, Kersting DK, Linares C, Ramos-Esplá AA and Antón J (2018) Biogeographic Differences in the Microbiome and Pathobiome of the Coral Cladocora caespitosa in the Western Mediterranean Sea. Front. Microbiol. 9:22. doi: 10.3389/fmicb.2018.00022

The endemic Mediterranean zooxanthellate scleractinian reef-builder *Cladocora caespitosa* is among the organisms most affected by warming-related mass mortality events in the Mediterranean Sea. Corals are known to contain a diverse microbiota that plays a key role in their physiology and health. Here we report the first study that examines the microbiome and pathobiome associated with *C. caespitosa* in three different Mediterranean locations (i.e., Genova, Columbretes Islands, and Tabarca Island). The microbial communities associated with this species showed biogeographical differences, but shared a common core microbiome that probably plays a key role in the coral holobiont. The putatively pathogenic microbial assemblage (i.e., pathobiome) of *C. caespitosa* also seemed to depend on geographic location and the human footprint. In locations near the coast and with higher human influence, the pathobiome was entirely constituted by *Vibrio* species, including the well-known coral pathogens *Vibrio coralliilyticus* and *V. mediterranei*. However, in the Columbretes Islands, located off the coast and the most pristine of the analyzed locations, no changes among microbial communities associated to healthy and necrosed samples were detected. Hence, our results provide new insights into the microbiome of the temperate corals and its role in coral health status, highlighting its dependence on the local environmental conditions and the human footprint.

Keywords: *Cladocora caespitosa*, necrosis, microbiome, pathobiome, Mediterranean Sea, amplicon sequencing, microbial diversity

INTRODUCTION

Mass mortality events of benthic invertebrates from different phyla (sponges, cnidarians, molluscs, ascidians, and bryozoans) have increased in frequency in the last two decades in the temperate Mediterranean Sea, with catastrophic effects in benthic communities (Cerrano et al., 2000; Linares et al., 2005; Garrabou et al., 2009; Crisci et al., 2011; Kersting et al., 2013; Kružić and Popijač, 2015; Jiménez et al., 2016; Rubio-Portillo et al., 2016a). Although the direct causes of these events remain unknown, there is scientific evidence to confirm that sea surface temperature anomalies linked to global warming are among the primary triggering factors (Garrabou et al., 2009; Kersting et al., 2013; Rivetti et al., 2014), together with energetic constraints (Coma et al., 2009) and the

potential occurrence of thermodependent pathogens (Bally and Garrabou, 2007; Vezzulli et al., 2010; Rubio-Portillo et al., 2014).

Cladocora caespitosa is the only endemic zooxanthellate scleractinian reef-building coral in the Mediterranean Sea and is one of the invertebrates repeatedly affected by mass mortalities (Rodolfo-Metalpa et al., 2005; Garrabou et al., 2009; Kersting et al., 2013; Rubio-Portillo et al., 2016a). The occurrence of the recurrent mortality events in *C. caespitosa* has shown a significant association to positive thermal anomalies (Kersting et al., 2013) and the events were characterized by partial or total colony death due to polyp tissue necrosis (Rodolfo-Metalpa et al., 2005; Kersting et al., 2013). Extensive bioconstructions of this coral are very rare at the present time (Peirano et al., 1998) and a significant effort has been made to assess its response to different global change-related impacts, such as the increase of sea water temperature (Rodolfo-Metalpa et al., 2006a,b; Kersting et al., 2013). However, to better understand the response of this temperate coral to environmental changes it should be also taken into account that scleractinian corals form a great collaborative consortium with a wide range of different microbial partners (Rohwer et al., 2002), which play key functional roles and contribute to coral survival (Ritchie, 2006; Glasl et al., 2016) and heat tolerance (Ziegler et al., 2017). Indeed, it has been recently shown that the coral microbiome is one of the most complex microbial habitats studied to date (Blackall et al., 2015). The term microbiome describes the assemblage of microorganisms, active or inactive, associated with a habitat (Lederberg and McCray, 2001) and the core microbiome is comprised of the organisms that are common across the microbiomes from different habitats and likely play a key role within the habitat (Turnbaugh et al., 2007). Conversely, the term pathobiome is used to describe the consortium of microbes within the microbiome that play a direct role in the causation of disease (Vayssier-Taussat et al., 2014).

Meron et al. (2012) is the only work focused on examining the response of *C. caespitosa* associated microbial communities to environmental changes, in particular to decreased pH conditions by clone libraries. Here we provide the first deep sequencing taxonomic characterization of the *C. caespitosa* microbiome and its relation with geographic location and coral health status. As an initial step in understanding microbial community variability in *C. caespitosa* and its potential association to the occurrence of coral tissue necrosis, we have also identified the *C. caespitosa* core microbiome and pathobiome.

MATERIALS AND METHODS

Sample Collection

Thirty *C. caespitosa* samples were collected between 5 and 15 m depth in three different locations in the Western Mediterranean Sea: (1) Pietra Ligure (44°08′50.17″N, 08°17′04.02″E, Italy), a location on the Genova coast with an estimated human population of around 9,000 habitants; (2) Columbretes Islands (39°53′49.5″N, 00°41′12.8″E, Spain), a Marine Protected Area without permanent human population and situated at 30 nautical miles off the coast; and (3) Tabarca Island (38°09′59″N, 00°28′56″E, Spain) a Marine Protected Area 2 nautical miles off the coast which has a permanent population of 50–60 habitants, but with a highly seasonal tourism activity with more than 3,000 persons visiting the island each day in the summer (**Figure 1**). A fragment of six different colonies, three of them visually healthy and three with necrosis signs, were collected at Genova and Tabarca in 2012, at Columbretes in 2014 and at Tabarca and Columbretes in 2015.

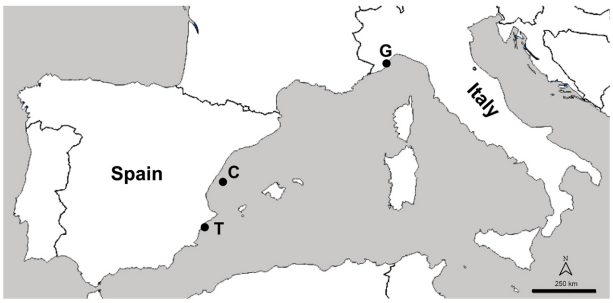


FIGURE 1 | Map of the samplings location: G (Genova), C (Columbretes Islands), and T (Tabarca).

DNA Extraction and Polymerase Chain Reaction Amplification of 16S rRNA Genes

DNA was extracted from coral tissues using the UltraClean Soil DNA Kit (Mo Bio; Carlsbad, CA, United States) following the manufacturer's instructions for maximum yield. The extracted genomic DNA was used for PCR amplifications of V3–V4 region of the 16S rRNA gene by using the following universal primers: Pro341F (CCTACGGGNBGCASCAG) (Takahashi et al., 2014) and Bact805R (GACTACHVGGGTATCTAATCC) (Herlemann et al., 2011). Each PCR mixture contained 5 μl of 10x PCR reaction buffer (Invitrogen), 1.5 μl of 50 mM MgCl2, 1 μl 10 mM dNTP mixture, 1 μl of 100 μM of each primer, 1 units of Taq polymerase, 3 μl of BSA (New England BioLabs), sterile MilliQ water up to 50 μl and 10 ng of DNA. Negative controls (with no template DNA) were included to assess potential contamination of reagents. The amplification products were purified with the GeneJET PCR purification kit (Fermentas, EU), quantified using the Qubit Kit (Invitrogen), and the quality (integrity and presence of a unique band) was confirmed by 1% agarose gel electrophoresis.

Illumina High-Throughput 16S rRNA Gene Sequencing and Bioinformatic Analyses

The QIIME 1.8.0 pipeline (Caporaso et al., 2010) was used for data processing. Paired-end MiSeq sequencing of the 30 samples generated 2,481,629 reads, which were deposited in the NCBI Sequence Read Archive (SRA) database under BioProject PRJNA407809. Forward and reverse reads were merged using SeqPrep and classified to their respective samples

according to their barcodes and then sequences were screened by quality and size, and de-replicated with the split_libraries.py script. The resulting file was checked for chimeric sequences with identify_chimeric_seqs.py script, against the SILVA_123 database[1] (Quast et al., 2013) using UCHIME (Edgar et al., 2011). Operational taxonomic units (OTUs) were defined at the level of 99% similarity, close to the threshold used to distinguish species (98.7% similarity in the whole 16S rRNA gene, Stackebrandt and Ebers, 2006), followed by taxonomy assignments against the SILVA reference database (version 123) using the UCLUST algorithm (Edgar, 2010), using pick_open_reference_otus.py script. Singletons, OTUs with less than 0.05% of abundance, and OTUs classified as chloroplast or mitochondria were removed from the dataset. Due to the great difference in library size among samples, the OTU table was rarefied to 1,066 reads with single_rarefaction.py script (the lowest number of the post-assembly and filtered sequences in a sample) for comparisons across samples (Weiss et al., 2015). A non-rarefied dataset was also analyzed to confirm the sensitivity of our results after rarefaction and elimination of a portion of available data (data not shown). Alpha diversity metrics (total observed number of OTUs, and Shannon-Wiener diversity) were generated from the rarefied OTU table using alpha_diversity.py script. The similarity among different microbial communities was assessed using phylogenetic information using jackknifed UPGMA (weighted pair group method with arithmetic mean) clustering based on the weighted UniFrac (Lozupone and Knight, 2005) distances between samples implemented in the QIIME pipeline with jackknifed_beta_diversity.py script. The statistical significance of the cluster among samples was tested using PERMANOVA analysis and differences in diversity indexes using ANOVA analysis. All statistical analyses were performed in R with the 'vegan' package (Oksanen, 2011).

The core microbiome and pathobiome were analyzed using QIIME with the minimum fraction of samples set at 85 and 100%, lowest percentage at which core OTU abundance was stable across healthy and necrosed samples, respectively. For functional prediction, PICRUSt software package[2] (Langille et al., 2013) was applied, which predicts the gene content of a microbial community from the information inferred from 16S RNA genes using an existing database of microbial genomes which predicts the tentative function of microbial communities. Metabolic predictions were made based on copy-number normalized OTUs and using healthy and samples showing tissue necrosis separately.

RESULTS AND DISCUSSION

Assessment of the Bacterial Diversity and Changes in Community Composition in *Cladocora caespitosa*

A dataset of 2,145,856 high-quality partial 16S rRNA gene sequences (length 447.6 ± 17.7) was generated after merging, quality trimming and chimera detection from paired-end Illumina reads. We clustered reads into 33,707 OTUs at 99% similarity. 445 OTUs had an overall relative abundance over 0.05%. The results related to the richness indexes, including number of observed OTUs and Shannon's diversity indexes are summarized in **Table 1**. Diversity differences were detected among locations (ANOVA, $F = 11.478$ and $p < 0.001$), being significantly higher in Columbretes than in Tabarca and Genova.

PERMANOVA analysis (factors: location and health status) showed that the geographic location was the determining factor explaining differences in coral bacterial community composition ($F = 0.4269$, $p < 0.001$), but there was also a significant interaction between location and health status ($F = 0.528$, $p < 0.001$). In order to assess temporal variation in each location, we analyzed data from Tabarca and Columbretes, where samples were collected in two different years (factors: year and health status). While in Tabarca significant differences in community structure were found between years ($F = 0.2729$, $p < 0.001$) and coral health status ($F = 0.1651$, $p < 0.01$), in Columbretes only significant differences between years were found ($F = 0.25503$, $p < 0.01$). Therefore, the response or involvement of the microbial community in coral tissue necrosis seemed to be different in Columbretes than in Genova or Tabarca, where differences in microbial communities associated to coral health status were detected. These differences between geographic locations could be related to local environmental conditions, such as distance to the coast and human impact. It is well known that the diversity and abundance of *Vibrio* species, including coral pathogens, is higher in zones closer to the coast and with higher anthropic impact (Vezzulli et al., 2013; Rubio-Portillo et al., 2014), which, in turn, could contribute to geographic differences in coral susceptibility to bacterial infection.

Cladocora caespitosa Bacterial Community Depends on Geographic Location

Principal coordinate analysis using weighted UniFrac distances (Lozupone and Knight, 2005) clearly separated the samples by geographic location (**Figure 2**). Although Illumina 16S rRNA gene sequencing is not a suitable technique for absolute quantification purposes, as we have shown by analyzing mock communities (Rubio-Portillo et al., 2016b), it can be used to compare relative abundances among samples. The relative abundance at the phylum level (**Figure 3**) showed that the proportion of each phylum varied among corals from different locations. Differences in coral microbial communities depending on geographic location have been previously observed in *Acropora* species (Littman et al., 2009) and in *Seriatopora hystrix* (Pantos et al., 2015). Nonetheless, we found the phylum *Proteobacteria* (mainly Alpha and Gammaproteobacteria classes) to be ubiquitous and dominant in all *C. caespitosa* samples (20–90%), as previously found in other coral species, such as *Porites astreoides* (Wegley et al., 2007) and *Oculina patagonica* (Rubio-Portillo et al., 2016b). Coral samples also comprised other dominant phyla depending on location: *Fusobacteria* was more abundant in Genova (28–56%), *Chloroflexi* in

[1] http://www.arb-silva.de

[2] http://picrust.github.com/picrust/

TABLE 1 | Alpha diversity analysis of bacterial community associated with *Cladocora caespitosa*.

Location	Year	Health status	Replicates	Postfiltering reads	N° OTUs	Shannon diversity
Genova	2012	Healthy	1	9215	23	2.97
			2	9624	28	2.58
			3	216439	23	3.15
		Necrosed	1	8647	30	3.39
			2	4126	17	2.73
			3	5765	16	2.81
Columbretes	2014	Healthy	1	28129	29	2.50
			2	34848	59	4.20
			3	64999	45	3.87
		Necrosed	1	9171	67	4.75
			2	101108	51	3.26
			3	13549	43	3.85
	2015	Healthy	1	19508	38	2.64
			2	31399	45	2.82
			3	21020	61	3.86
		Necrosed	1	64999	38	2.64
			2	86562	30	2.45
			3	45963	50	2.44
Tabarca	2012	Healthy	1	4060	42	3.10
			2	180521	29	1.09
			3	40661	62	3.97
		Necrosed	1	14046	32	2.74
			2	1253	35	2.70
			3	1066	57	3.92
	2015	Healthy	1	2607	55	3.56
			2	121073	43	3.78
			3	44250	68	3.76
		Necrosed	1	97135	65	4.14
			2	93380	77	5.11
			3	97876	75	4.73

OTUS and Shannon diversity index were calculated from normalized sequences with the lowest number of the post-assembly and filtered sequences in a sample (without singletons, OTUs with less than 0.05% of abundance and OTUs classified as chloroplast or mitochondria).

Columbertes (1–64%) and *Bacteroidetes* in Tabarca (3–69%), although these last two phyla showed a great variability among samples within each location. The only study to date that has addressed microbiota associated with *C. caespitosa*, in Gulf of Naples, also found that *Proteobacteria* together with Bacteridetes were the dominant phyla in coral tissue (Meron et al., 2012), hence these phyla seem to be ubiquitous and dominant in *C. caespitosa* tissues regardless of geographic location.

Dissimilarity percentages calculated using SIMPER confirmed that the average dissimilarity among locations was high: 97.98% (Genova-Columbretes), 88.13% (Columbretes-Tabarca), and 84.20% (Genova-Tabarca). The OTUs primarily responsible for the biogeographical differences showed similarity percentages with known species below 97%, which could be indicative of the occurrence of new species. The largest representative sequences of each of these OTUs were inserted into the non-redundant SILVA SSURef_NR99_123 default tree (Quast et al., 2013) to identify the closest available sequences and then a phylogenetic reconstruction with the reference database LTP s123 (Munoz et al., 2011) was carried out using the ARB software package (Ludwig et al., 2004) (Supplementary Figure 1). In Genova the dominant sequence cluster was affiliated with *Propionigenium* lineage with 93.7–98.2% partial 16S rRNA gene sequence identity. Species of this genus are anaerobic bacteria that contribute to organic matter degradation in tidal sediments (Graue et al., 2012) and were previously associated with the gut environment of marine invertebrates (Dishaw et al., 2014). OTUs from Tabarca were loosely affiliated to *Maritimimonas* and *Tenacibaculum* genus with identity values of 93 and 89%, respectively. *Maritimimonas* genus is comprised of only one species *Maritimimonas rapanae*, which was isolated from gut microflora of a mollusk (Park et al., 2009) and *Tenacibaculum* genus has also been recently related to the gut microbiome of jelly fishes (Viver et al., 2017). Further studies are needed to assess if these taxa belong to the coral gut microbiota, but our results seem to indicate that corals living in different habitats with different nutritional resources could host different gut microbiota, which is consistent with previous studies in cold corals (Meistertzheim et al., 2016).

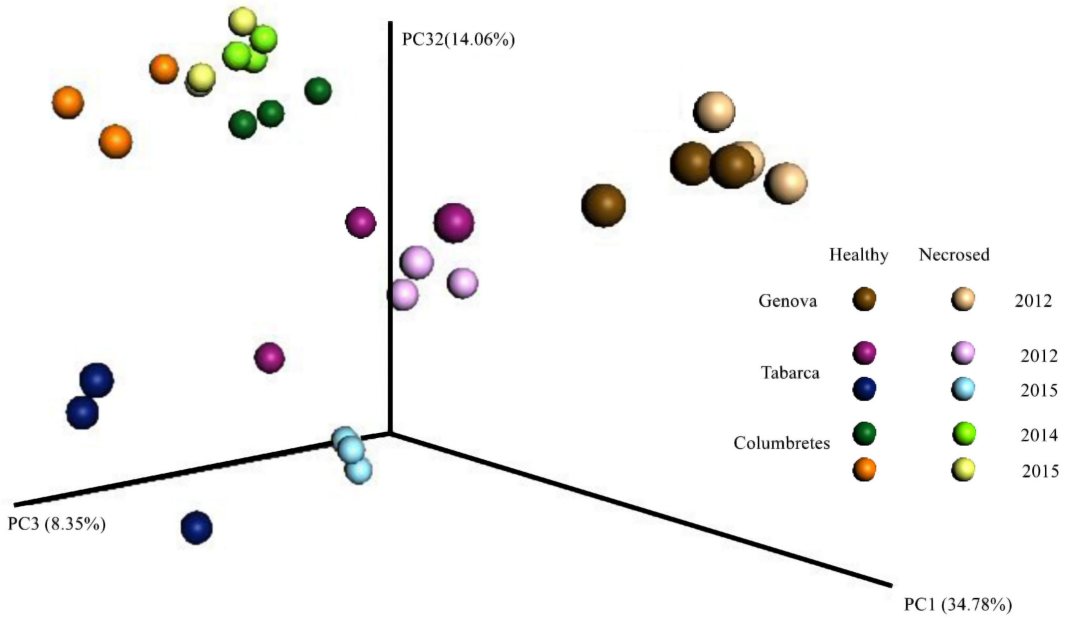


FIGURE 2 | Bacterial communities associated with *Cladocora caespitosa* tissues clustered using coordinated analysis of the weighed UniFrac distance matrix. Each circle corresponds to a coral sample.

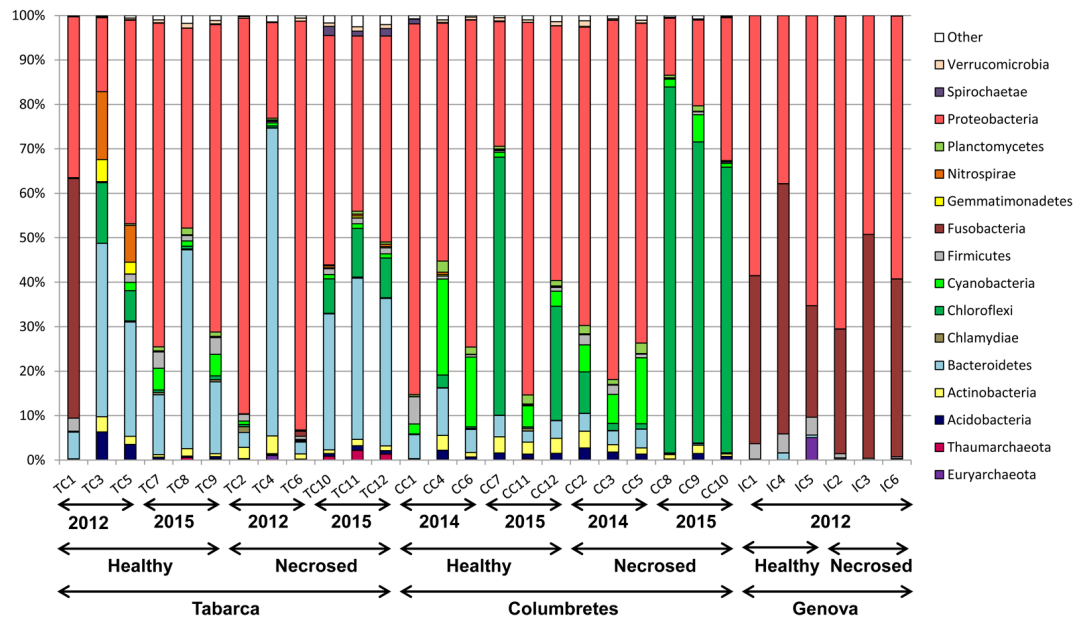


FIGURE 3 | Taxonomic classification at the phylum level of bacterial communities associated with *C. caespitosa*.

Changes in the Bacterial Community Related to Coral Health Status

Bacterial community composition differed in relation to coral health status in Genova and Tabarca (**Figure 2**). In general terms, the microbial communities associated with colonies with tissue necrosis signs were more similar to each other than in healthy colonies in the two locations (**Table 2**, see similarity percentages). This fact has been previously observed in the coral *O. patagonica*, which also inhabits the Mediterranean Sea (Rubio-Portillo et al., 2016a). In addition to the OTUs belonging to *Propionigenium* and *Maritimimonas,* which could be part of the coral gut microbiome (see above), SIMPER analysis indicated that OTUs responsible for the observed differences (**Table 2** and Supplementary Figure 2) were *Thalassospira*

TABLE 2 | Percentage of contribution of main OTUs to bacterial community structure (based on SIMPER analysis), indicating the average contribution to the similarity (S) and dissimilarity (D) between healthy and unhealthy corals in each location: **(A)** Genova and **(B)** Tabarca.

(A) Genova			
OTU	**Healthy** $S = 46.66$	**Unhealthy** $S = 60.79$	**Healthy/ unhealthy** $D = 58.84$
Vibrio scophthalmi	5.66	12.92	14.25
Arcobacter sp.	0.38	6.16	8.97
Thalassospira sp.	5.77	0.19	6.41
Propionigenium sp.1	44.31	28.49	5.33
Propionigenium sp. 2	17.83	12.77	5.26
(B) Tabarca			
OTU	**Healthy** $S = 28.43$	**Unhealthy** $S = 41.02$	**Healthy/ unhealthy** $D = 87.53$
Pseudovibrio sp.	17.06	1.82	28.64
Vibrio alginolyticus	0.17	9.80	11.72
Vibrio sp. *FPLN2*	0.22	8.37	10.54
Maritimimonas sp.	2.94	5.28	5.09

and *Pseudovibrio* species, which were mainly associated with healthy corals in Genova and Tabarca, respectively. Species belonging to these two genera have been detected in microbial communities associated with other coral species and previous studies have emphasized their possible role in the coral holobiont. *Thalassospira* could be involved in carbon and phosphorus cycles (Thomas et al., 2010) and *Pseudovibrio* in the nitrogen cycle (Bondarev et al., 2013), as well as in the coral protection by inhibiting pathogens growth (Nissimov et al., 2009; Rypien et al., 2010). Nevertheless, *Vibrio* spp. were the OTUs mainly related to necrosed corals from both locations (**Tables 2A,B**), in agreement with previous findings that reported an increase of this bacteria genus in corals (Rubio-Portillo et al., 2014, 2016b) and gorgonians (Vezzulli et al., 2010) with disease signs.

Conversely, with the level of resolution provided by 16S rRNA gene profiling, no significant differences were found in the microbial community between healthy and necrosed corals in Columbretes, which may suggest that bacteria are not directly involved in the development of tissue necrosis in this location. Furthermore, the coral pathogen *Vibrio coralliilyticus* was detected both in necrosed and healthy corals in Columbretes (see below), which adds additional uncertainty to the role that microorganisms might play in *C. caespitosa* necrosis in this location. It is well known that interactions among coral pathogens, environmental stresses and physiological and immune status of the coral host are complex (Bourne et al., 2009) and it would be necessary to carry out further studies to better understand the role of microorganisms in the necrosis events. Therefore, *Vibrio* species could be transient members that can vary in response to geographic location and other environmental factors.

TABLE 3 | Percentage of contribution of main OTUs to bacterial community structure (based on SIMPER analysis), indicating the average contribution to the similarity (S) and dissimilarity (D) between corals collected in different years in **(A)** Columbretes and **(B)** Tabarca.

(A) Columbretes			
OTU	**2014** $S = 12.5$	**2015** $S = 42.76$	**2014/2015** $D = 90.30$
Uncultured Chloroflexi bacterium	34.67	468	23.02
Uncultured *Cohaesibacter*	1.33	98.67	5.17
Uncultured *Rhodothermaceae*	89.93	0	4.67
Uncultured *Anaerolineaceae*	9.67	90.50	4.64
(B) Tabarca			
OTU	**2012** $S = 8.39$	**2015** $S = 41.67$	**2012/2015** $D = 90.58$
Ruegeria sp.	15.83	153.5	8.36
Pseudovibrio sp.	142.5	57.83	8.22
Vibrio sp.	8.00	97.83	7.91
Uncultured *Blattabacteriaceae*	93.83	1.33	5.03

Temporal Changes in *Cladocora caespitosa* Bacterial Community

SIMPER analysis showed that in 2015, when an exceptional heat wave was recorded in Tabarca (Rubio-Portillo et al., 2016a) and in Columbretes (unpublished results), *C. caespitosa* bacterial community associated with corals with apparently different health status were more similar to each other than to those collected in previous years (**Table 3**, see similarity percentages) in the same locations. Therefore, coral microbial communities' changes due to heat stress were detected in both healthy and necrosed corals. In Columbretes an increase in species belonging to *Chloroflexi* phylum was detected, while in Tabarca a decrease in *Pseudovibrio* genus and an increase of *Vibrio* and *Ruegeria* genus was detected (**Table 3**). Species belonging to *Vibrio* genus are well known coral pathogens but *Ruegeria* genus has also been previously linked to different coral diseases, such as Black Band Disease in the Caribbean Sea (Sekar et al., 2008), Yellow Band Disease in the Red Sea (Apprill et al., 2013) or White Patch Syndrome in the Indian Ocean (Séré et al., 2013), although it has never been demonstrated to be the etiological agent of these coral diseases. Therefore, in Tabarca heat stress seemed to be accompanied by an increase of potential coral pathogens in coral microbial communities. Similarly, the microbial community associated with this coral species also showed some changes in bacterial group composition under different pH conditions (Meron et al., 2012), so environmental stress related to climate change could influence *C. caespitosa* microbial community.

The Core Microbiome of *Cladocora caespitosa*

Using Venn diagrams (**Figure 4A**), we have identified a small group of bacteria belonging to the genera *Vallitaela*, *Vibrio*,

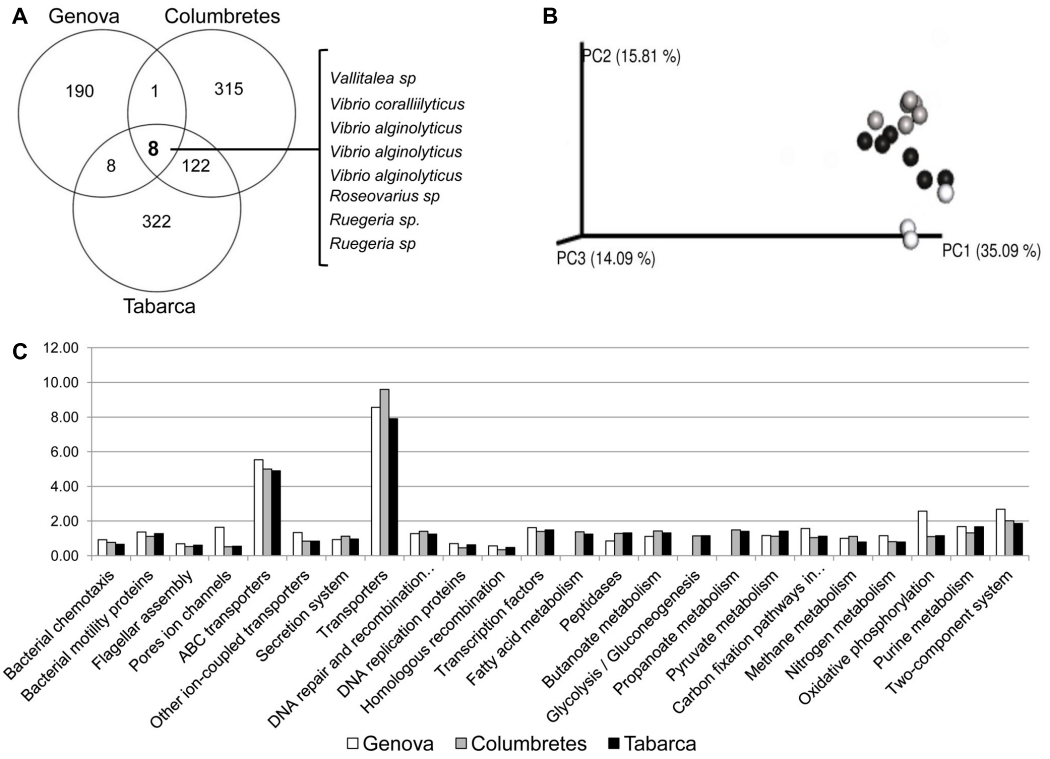


FIGURE 4 | Analysis of the core microbiome associated with *C. caespitosa*. **(A)** Venn Diagrams showing OTUs shared by core microbiomes; **(B)** phylogenetic cluster of samples using coordinated analysis of the weighed UniFrac distance matrix, each circle corresponds to a coral sample (in white Genova, in gray Columbretes and in black Tabarca); and **(C)** most abundant predicted functions from the core microbiome using PICRUSt and copy-number normalized OTUs from healthy corals samples.

Roseovarius, and *Ruegeria* that are ubiquitously associated with *C. caespitosa* regardless of local environmental factors, and could thus be considered as the coral core microbiome. These bacterial phylotypes, which had less than 4% relative abundance within the microbiome, are rare but conserved among geographic locations and time and they probably play key roles in the coral holobiont, as suggested by Ainsworth et al. (2015). Indeed, some of these OTUs like *Vibrio* spp. and *Roseovarius* sp., have been previously related to nitrogen (Chimetto et al., 2008) and sulfur cycling (Raina et al., 2009) in corals. Noticeably, the coral pathogen *V. coralliilyticus*, previously related to diseases in the red gorgonian *Paramuricea clavata* (Bally and Garrabou, 2007; Vezzulli et al., 2010) and the coral *O. patagonica* (Rubio-Portillo et al., 2014), forms part of the *C. caespitosa* core microbiome. This is in agreement with our recent results that showed that healthy and necrosed samples from *C. caespitosa* harbor different clonal types of *V. coralliilyticus*, which could, therefore, have different invasive-disease potential (Rubio-Portillo et al., 2017). This pathogen was also detected in apparently healthy tissue of the coral *Montastrea annularis* (Rypien et al., 2010), which is in good agreement with our findings. The role of the pathogen *V. coralliilyticus* in *C. caespitosa* microbiome is not clear and requires further research, but this species seems to be a native member of *C. caespitosa* coral microbial community, which in particular conditions (e.g., heat stress) could produce tissue necrosis.

Venn Diagrams only show differences among samples based on presence/absence of OTUs, but different OTUs that belong to the same lineage could perform similar functions within the coral holobiont (Shade and Handelsman, 2012). Therefore, UniFrac distances analysis was used to assess differences in phylogenetic diversity across samples. This analysis did not show differences in the core microbiome of samples collected at the three different locations (**Figure 4B**) and putative functions, predicted from 16S RNA gene information with PICRUSt, were also very similar (**Figure 4C**). Most of these functions were linked to energy metabolism, such as the nitrogen cycle or carbon fixation, or related to transport of sugars and ions, i.e., to metabolic exchange between the coral host and bacterial microbiota. Therefore, OTUs that constituted the core microbiome were different among locations but they were phylogenetically closed and they carry out similar functions in the coral holobiont.

The Pathobiome of *Cladocora caespitosa*

The *C. caespitosa* pathobiome was assessed only from necrosed samples collected in Tabarca and Genova, the two locations where the coral bacterial associated community was different among healthy and necrosed corals. Nineteen OTUs (**Figure 5A**)

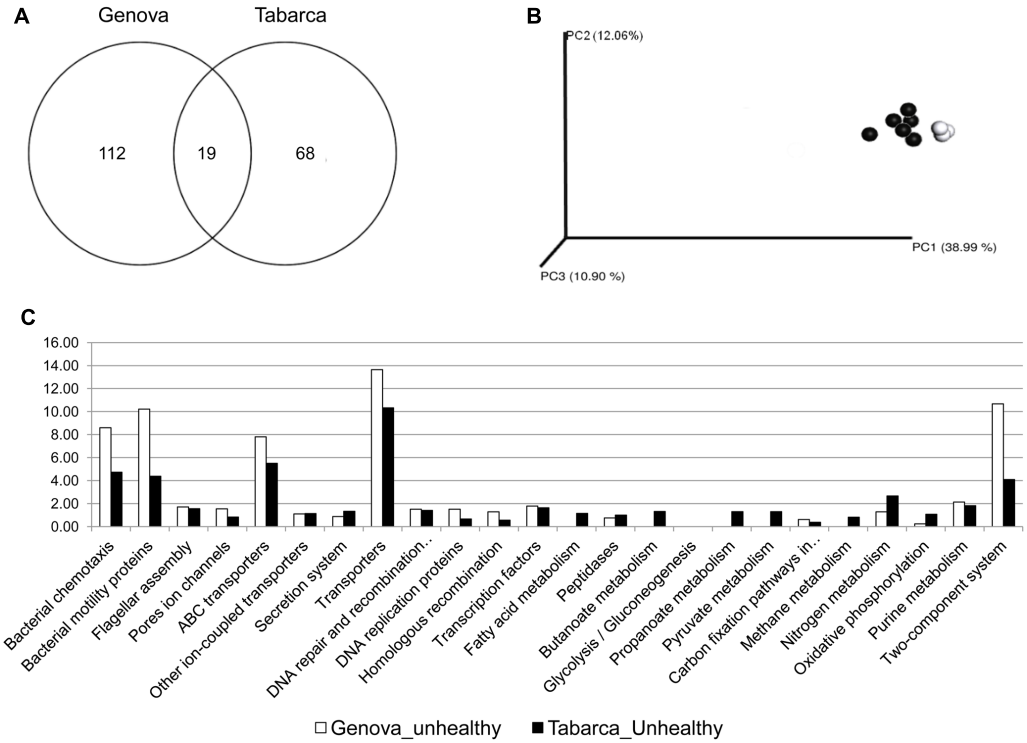


FIGURE 5 | Analysis of pathobiome associated with *C. caespitosa*. **(A)** Venn Diagrams showing OTUs shared by pathobiomes; **(B)** phylogenetic cluster of samples using coordinated analysis of the weighed UniFrac distance matrix, each circle corresponds to a coral sample (in white Genovan and in black Tabarca); and **(C)** most abundant predicted functions from the pathobiome using PICRUSt and copy-number normalized OTUs from corals samples with tissue necrosis signs.

constitute the *C. caespitosa* pathobiome and 15 of them belong to *Vibrio* genus, including the two known coral pathogens *V. mediterranei* and *V. corallilyticus*. This is worth noting since a previous study suggested a synergistic effect on the virulence of these two pathogens (Rubio-Portillo et al., 2014). UniFrac distances analysis did not show differences in the core pathobiome of samples collected in these two locations (**Figure 5B**). Furthermore, virulence-associated factors, involved in motility, chemotaxis and two-component regulatory signal transductions systems, were the functions predicted for the species present in these pathobiomes (**Figure 5C**), which also suggested that *Vibrio* species could be related to the increase of virulence factors in corals with tissue necrosis signs, and could thus be responsible for this process in Genova and Tabarca.

CONCLUSION

Recently, Hernandez-Agreda et al. (2017) suggested that the coral microbiome is divided into three main components: (i) a ubiquitous core microbiome; (ii) a dynamic site and/or species-specific community; and (iii) a highly variable community reflective of biotic and abiotic fluctuations. Here, we show that in *C. caespitosa* the ubiquitous core microbiome is constituted by rare conserved OTUs, including the coral pathogen *V. corallilyticus*, which could be involved in nutrient cycling in the coral holobiont. The site-specific community is composed of OTUs that may be related to the gut microbiome and hence depends on nutritional resources and environmental conditions. The pathobiome of *C. caespitosa* was constituted entirely by *Vibrio* species, including the pathogens *V. coralliilyticus* and *V. mediterranei*. This fact, together with the increase of virulence factors predicted in the pathobiomes, suggests that these pathogens could be involved in *C. caespitosa* tissue necrosis in Tabarca and Genova, but not in Columbretes, the most pristine of the three analyzed locations. Accordingly, as previously hinted (Kersting et al., 2013, 2015), *C. caespitosa* tissue necrosis likely has a multi factorial origin, with the increase of sea water temperature as the main triggering factor (Kersting et al., 2013), while the role of the coral microbial community in the necrosis events could be highly dependent on local environmental conditions and the human footprint.

ETHICS STATEMENT

The sampling in the Mediterranean Sea was performed in accordance with Spanish and Italian laws. In particular, for sampling in the Marine protected Areas of Tabarca and Columbretes permissions were granted by the Secretary-General for Fisheries of the Spanish Ministry of Agriculture, Food and Environment (permission numbers are 03/12 and 09/15 for

Tabarca and 04/14 and 06/15 for Columbretes). For sampling in Pietra Ligure local competent authorities (Coast Guard and Municipality of Pietra Ligure) were informed and allowed the sampling of biological material for research purposes only. Italian laws do not ask for permissions for sampling in case of scientific purposes.

AUTHOR CONTRIBUTIONS

DKK and ER-P collected the samples. ER-P performed the DNA extractions, analysis of results. All authors contributed to the manuscript and participated in the writing and editing process.

FUNDING

This work was funded by the European Union's framework program Horizon 2020 (LEIT - BIO - 2015 - 685474, Metafluidics, to JA) and the Ministerio de Economía y Competitividad Smart project (CGL2012-32194, to CL).

ACKNOWLEDGMENTS

The authors gratefully thank the staff of the Department of Marine Sciences and Applied Biology and the Marine Research Centre of Santa Pola (CIMAR). They also greatly appreciate the friendly cooperation of the Secretary-General for Fisheries of the Spanish Ministry of Agriculture, Food and Environment, the Marine Reserve of Tabarca wardens (particularly Felio Lozano) and the Columbretes Islands Marine Reserve staff.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00022/full#supplementary-material

REFERENCES

Ainsworth, T., Krause, L., Bridge, T., Torda, G., Raina, J.-B., Zakrzewski, M., et al. (2015). The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J.* 9, 2261–2274. doi: 10.1038/ismej.2015.39

Apprill, A., Hughen, K., and Mincer, T. (2013). Major similarities in the bacterial communities associated with lesioned and healthy Fungiidae corals. *Environ. Microbiol.* 15, 2063–2072. doi: 10.1111/1462-2920.12107

Bally, M., and Garrabou, J. (2007). Thermodependent bacterial pathogens and mass mortalities in temperate benthic communities: a new case of emerging disease linked to climate change. *Glob. Change Biol.* 13, 2078–2088. doi: 10.1111/j.1365-2486.2007.01423.x

Blackall, L. L., Wilson, B., and van Oppen, M. J. (2015). Coral-the world's most diverse symbiotic ecosystem. *Mol. Ecol.* 24, 5330–5347. doi: 10.1111/mec.13400

Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A., and Schulz-Vogt, H. N. (2013). The genus *Pseudovibrio* contains metabolically versatile bacteria adapted for symbiosis. *Environ. Microbiol.* 15, 2095–2113. doi: 10.1111/1462-2920.12123

Bourne, D. G., Garren, M., Work, T. M., Rosenberg, E., Smith, G. W., and Harvell, C. D. (2009). Microbial disease and the coral holobiont. *Trends Microbial.* 17, 554–562. doi: 10.1016/j.tim.2009.09.004

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

Cerrano, C., Bavestrello, G., Bianchi, C. N., Cattaneo-Vietti, R., Bava, S., Morganti, C., et al. (2000). A catastrophic mass-mortality episode of gorgonians and other organisms in the Ligurian Sea (North-western Mediterranean), summer 1999. *Ecol. Lett.* 3, 284–293. doi: 10.1046/j.1461-0248.2000.00152.x

Chimetto, L. A., Brocchi, M., Thompson, C. C., Martins, R. C. R., Ramos, H. R., and Thompson, F. L. (2008). Vibrios dominate as culturable nitrogen-fixing bacteria of the Brazilian coral *Mussismilia hispida*. *Syst. Appl. Microbiol.* 31, 312–319. doi: 10.1016/j.syapm.2008.06.001

Coma, R., Ribes, M. M., Serrano, E., Jiménez, E., Salat, J., and Pascual, J. (2009). Global warming-enhanced stratification and mass mortality events in the Mediterranean. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6176–6181. doi: 10.1073/pnas.0805801106

Crisci, C., Bensoussan, N., Romano, J. C., and Garrabou, J. (2011). Temperature anomalies and mortality events in marine communities: insights on factors behind differential mortality impacts in the NW Mediterranean. *PLOS ONE* 6:e23814. doi: 10.1371/journal.pone.0023814

Dishaw, L. J., Flores-Torres, J., Lax, S., Gemayel, K., Leigh, B., Melillo, D., et al. (2014). The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLOS ONE* 9:e93386. doi: 10.1371/journal.pone.0093386

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

Garrabou, J., Coma, R., Bensoussan, N., Bally, M., Chevaldonné, P., Cigliano, M., et al. (2009). Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave. *Glob. Change Biol.* 15, 1090–1103. doi: 10.1111/j.1365-2486.2008.01823.x

Glasl, B., Herndl, G. J., and Frade, P. R. (2016). The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *ISME J.* 10, 2280–2292. doi: 10.1038/ismej.2016.9

Graue, J., Engelen, B., and Cypionka, H. (2012). Degradation of cyanobacterial biomass in anoxic tidal-flat sediments: a microcosm study of metabolic processes and community changes. *ISME J.* 6, 660–669. doi: 10.1038/ismej.2011.120

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41

Hernandez-Agreda, A., Gates, R. D., and Ainsworth, T. D. (2017). Defining the core microbiome in corals' microbial soup. *Trends Microbiol.* 25, 125–140. doi: 10.1016/j.tim.2016.11.003

Jiménez, C., Hadjioannou, L., Petrou, A., Nikolaidis, A., Evriviadou, M., and Lange, M. A. (2016). Mortality of the scleractinian coral *Cladocora caespitosa* during a warming event in the Levantine Sea (Cyprus). *Reg. Environ. Change* 16, 1963–1973. doi: 10.1007/s10113-014-0729-2

Kersting, D. K., Bensoussan, N., and Linares, C. (2013). Long-term responses of the endemic reef-builder *Cladocora caespitosa* to Mediterranean warming. *PLOS ONE* 8:e70820. doi: 10.1371/journal.pone.0070820

Kersting, D. K., Cebrian, E., Casado, C., Teixidó, N., Garrabou, J., and Linares, C. (2015). Experimental evidence of the synergistic effects of warming and invasive algae on a temperate reef-builder coral. *Sci. Rep.* 5:18635. doi: 10.1038/srep18635

Kružić, P., and Popijač, A. (2015). Mass mortality events of the coral *Balanophyllia europaea* (Scleractinia, Dendrophylliidae) in the Mljet National Park (Eastern

Adriatic Sea) caused by sea temperature anomalies. *Coral Reefs* 34, 109–118. doi: 10.1007/s00338-014-1231-5

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

Lederberg, J., and McCray, A. T. (2001). Ome sweet 'omics – a genealogical treasury of words. *Scientist* 15:8.

Linares, C., Coma, R., Diaz, D., Zabala, M., Hereu, B., and Dantart, L. (2005). Immediate and delayed effects of a mass mortality event on gorgonian population dynamics and benthic community structure in the NW Mediterranean Sea. *Mar. Ecol. Prog. Ser.* 305, 127–137. doi: 10.3354/meps305127

Littman, R. A., Willis, B. L., Pfeffer, C., and Bourne, D. G. (2009). Diversities of coral-associated bacteria differ with location, but not species, for three acroporid corals on the Great Barrier Reef. *FEMS Microbiol. Ecol.* 68, 152–163. doi: 10.1111/j.1574-6941.2009.00666.x

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

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar Buchner, A., et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

Meistertzheim, A. L., Lartaud, F., Arnaud-Haond, S., Kalenitchenko, D., Bessalam, M., Le Bris, N., et al. (2016). Patterns of bacteria-host associations suggest different ecological strategies between two reef building cold-water coral species. *Deep Sea Res. I* 114, 12–22. doi: 10.1016/j.dsr.2016.04.013

Meron, D., Rodolfo-Metalpa, R., Cunning, R., Baker, A. C., Fine, M., and Banin, E. (2012). Changes in coral microbial communities in response to a natural pH gradient. *ISME J.* 6, 1775–1785. doi: 10.1038/ismej.2012.19

Munoz, R., Yarza, P., Ludwig, W., Euze, J., Amann, R., Schleifer, K. H., et al. (2011). Release LTPs104 of the all-species living tree. *Syst. Appl. Microbiol.* 34, 169–170. doi: 10.1016/j.syapm.2011.03.001

Nissimov, J., Rosenberg, E., and Munn, C. B. (2009). Antimicrobial properties of resident coral mucus bacteria of *Oculina patagonica*. *FEMS Microbiol. Lett.* 292, 210–215. doi: 10.1111/j.1574-6968.2009.01490.x

Oksanen, J. (2011). *Multivariate Analysis of Ecological Communities in R: Vegan Tutorial*. Available at: http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf

Pantos, O., Bongaerts, P., Dennis, P. G., Tyson, G. W., and Hoegh-Guldberg, O. (2015). Habitat-specific environmental conditions primarily control the microbiomes of the coral *Seriatopora hystrix*. *ISME J.* 9, 1916–1927. doi: 10.1038/ismej.2015.3

Park, S. C., Baik, K. S., Kim, D., and Seong, C. N. (2009). *Maritimimonas rapanae* gen. nov., sp. nov., isolated from gut microflora of the veined rapa whelk, *Rapana venosa*. *Int. J. Syst. Evol. Microbiol.* 59(Pt 11), 2824–2829. doi: 10.1099/ijs.0.010504-0

Peirano, A., Morri, C., Mastronuzzi, G., and Bianchi, C. N. (1998). The coral *Cladocora caespitosa* (Anthozoa, Scleractinia) as a bioherm builder in the Mediterranean Sea. *Mem. Descr. Carta Geol. Ital.* 52, 59–74.

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, 590–596. doi: 10.1093/nar/gks1219

Raina, J. B., Tapiolas, D., Willis, B. L., and Bourne, D. G. (2009). Coral-associated bacteria and their role in the biogeochemical cycling of sulfur. *Appl. Environ. Microbiol.* 75, 3492–3501. doi: 10.1128/AEM.02567-08

Ritchie, K. B. (2006). Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. *Mar. Ecol. Prog. Ser.* 322, 1–14. doi: 10.3354/meps322001

Rivetti, I., Fraschetti, S., Lionello, P., Zambianchi, E., and Boero, F. (2014). Global warming and mass mortalities of benthic invertebrates in the Mediterranean Sea. *PLOS ONE* 9:e115655. doi: 10.1371/journal.pone.0115655

Rodolfo-Metalpa, R., Bianchi, C. N., Peirano, A., and Morri, C. (2005). Tissue necrosis and mortality of the temperate coral *Cladocora caespitosa*. *Ital. J. Zool.* 72, 271–276. doi: 10.1371/journal.pone.0070820

Rodolfo-Metalpa, R., Richard, C., Allemand, D., Bianchi, C. N., Morri, C., and Ferrier-Pages, C. (2006a). Response of zooxanthellae in symbiosis with the Mediterranean corals *Cladocora caespitosa* and *Oculina patagonica* to elevated temperatures. *Mar. Biol.* 150, 45–55.

Rodolfo-Metalpa, R., Richard, C., Allemand, D., and Ferrier-Pages, C. (2006b). Growth and photosynthesis of two Mediterranean corals, *Cladocora caespitosa* and *Oculina patagonica*, under normal and elevated temperatures. *J. Exp. Biol.* 209, 4546–4556.

Rohwer, F., Seguritan, V., Azam, F., and Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Mar. Ecol. Prog. Ser.* 243, 1–10. doi: 10.3354/meps243001

Rubio-Portillo, E., Gago, J. F., Martínez-García, M., Vezzulli, L., Rosselló-Mora, R., Antón, J., et al. (2017). *Vibrio* communities in scleractinian corals differ according to health status and geographic location in the Mediterranean Sea. *Syst. Appl. Microbiol.* (in press). doi: 10.1016/j.syapm.2017.11.007

Rubio-Portillo, E., Izquierdo-Muñoz, A., Gago, J. F., Rosselló-Mora, R., Antón, J., and Ramos-Esplá, A. A. (2016a). Effects of the 2015 heat wave on benthic invertebrates in the Tabarca Marine Protected Area (southeast Spain). *Mar. Environ. Res.* 122, 135–142. doi: 10.1016/j.marenvres.2016.10.004

Rubio-Portillo, E., Santos, F., Martínez-García, M., de los Ríos, A., Ascaso, C., Souza-Egipsy, V., et al. (2016b). Structure and temporal dynamics of the bacterial communities associated to microhabitats of the coral *Oculina patagonica*. *Environ. Microbiol.* 18, 4564–4578. doi: 10.1111/1462-2920.13548

Rubio-Portillo, E., Yarza, P., Penalver, C., Ramos-Esplá, A. A., and Anton, J. (2014). New insights into *Oculina patagonica* coral diseases and their associated *Vibrio* spp. communities. *ISME J.* 8, 1794–1807. doi: 10.1038/ismej.2014.33

Rypien, K. L., Ward, J. R., and Azam, F. (2010). Antagonistic interactions among coral-associated bacteria. *Environ. Microbiol.* 12, 28–39. doi: 10.1111/j.1462-2920.2009.02027.x

Sekar, R., Kaczmarsky, L. T., and Richardson, L. L. (2008). Microbial community composition of black band disease on the coral host *Siderastrea siderea* from three regions of the wider Caribbean. *Mar. Ecol. Prog. Ser.* 362, 85–98. doi: 10.3354/meps07496

Séré, M. G., Tortosa, P., Chabanet, P., Turquet, J., Quod, J. P., and Schleyer, M. H. (2013). Bacterial communities associated with *Porites* white patch syndrome (PWPS) on three Western Indian Ocean (WIO) coral reefs. *PLOS ONE* 8:e83746. doi: 10.1371/journal.pone.0083746

Shade, A., and Handelsman, J. (2012). Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* 14, 4–12. doi: 10.1111/j.1462-2920.2011.02585.x

Stackebrandt, E., and Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today* 33, 152–155.

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., and Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of *Bacteria* and *Archaea* using next-generation sequencing. *PLOS ONE* 9:e105592. doi: 10.1371/journal.pone.0105592

Thomas, S., Burdett, H., Temperton, B., Wick, R., Snelling, D., McGrath, J. W., et al. (2010). Evidence for phosphonate usage in the coral holobiont. *ISME J.* 4, 459–461. doi: 10.1038/ismej.2009.129

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, 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

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

Vezzulli, L., Colwell, R. R., and Pruzzo, C. (2013). Ocean warming and spread of pathogenic vibrios in the aquatic environment. *Microbiol. Ecol.* 65, 817–825. doi: 10.1007/s00248-012-0163-2

Vezzulli, L., Previati, M., Pruzzo, C., Marchese, A., Bourne, D. G., and Cerrano, C. (2010). *Vibrio* infections triggering mass mortality events in a warming Mediterranean Sea. *Environ. Microbiol.* 12, 2007–2019. doi: 10.1111/j.1462-2920.2010.02209.x

Viver, T., Orellana, L. H., Hatt, J. K., Urdiain, M., Díaz, S., Richter, M., et al. (2017). The low diverse gastric microbiome of the jellyfish *Cotylorhiza tuberculata* is dominated by four novel taxa. *Environ. Microbiol.* 19, 3039–3058. doi: 10.1111/1462-2920.13763

Wegley, L., Edwards, R. A., Rodriguez-Brito, B., Hong, L., 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

Weiss, S. J., Xu, Z., Amir, A., Peddada, S., Bittinger, K., Gonzalez, A., et al. (2015). Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. *PeerJ Prepr.* 3:e1408.

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., and Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nat. Commun.* 8:14213. doi: 10.1038/ncomms14213

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 Rubio-Portillo, Kersting, Linares, Ramos-Esplá and Antón. 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: 22 December 2017
doi: 10.3389/fmicb.2017.02619

Functional Stability and Community Dynamics during Spring and Autumn Seasons Over 3 Years in Camargue Microbial Mats

Mercedes Berlanga[1], Montserrat Palau[1] and Ricardo Guerrero[2,3]*

[1] Department of Biology, Environment and Health, Section Microbiology, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain, [2] Laboratory of Molecular Microbiology and Antimicrobials, Department of Pathology and Experimental Therapeutics, Faculty of Medicine, University of Barcelona – Institut d'Investigació Biomédica de Bellvitge, Barcelona, Spain, [3] Academia Europaea-Barcelona Knowledge Hub, Barcelona, Spain

Microbial mats are complex biofilms in which the major element cycles are represented at a millimeter scale. In this study, community variability within microbial mats from the Camargue wetlands (Rhone Delta, southern France) were analyzed over 3 years during two different seasons (spring and autumn) and at different layers of the mat (0–2, 2–4, and 4–6 mm). To assess bacterial diversity in the mats, amplicons of the V1–V2 region of the 16S rRNA gene were sequenced. The community's functionality was characterized using two approaches: (i) inferred functionality through 16S rRNA amplicons genes according to PICRUSt, and (ii) a shotgun metagenomic analysis. Based on the reads distinguished, microbial communities were dominated by Bacteria (∼94%), followed by Archaea (∼4%) and Eukarya (∼1%). The major phyla of Bacteria were Proteobacteria, Bacteroidetes, Spirochaetes, Actinobacteria, Firmicutes, and Cyanobacteria, which together represented 70–80% of the total population detected. The phylum Euryarchaeota represented ∼80% of the Archaea identified. These results showed that the total bacterial diversity from the Camargue microbial mats was not significantly affected by seasonal changes at the studied location; however, there were differences among layers, especially between the 0–2 mm layer and the other two layers. PICRUSt and shotgun metagenomic analyses revealed similar general biological processes in all samples analyzed, by season and depth, indicating that different layers were functionally stable, although some taxa changed during the spring and autumn seasons over the 3 years. Several gene families and pathways were tracked with the oxic-anoxic gradient of the layers. Genes directly involved in photosynthesis (KO, KEGG Orthology) were significantly more abundant in the top layer (0–2 mm) than in the lower layers (2–4 and 4–6 mm). In the anoxic layers, the presence of ferredoxins likely reflected the variation of redox reactions required for anaerobic respiration. Sulfatase genes had the highest relative abundance below 2 mm. Finally, chemotaxis signature genes peaked sharply at the oxic/photic and transitional oxic-anoxic boundary. This functional differentiation reflected the taxonomic diversity of the different layers of the mat.

Keywords: Camargue microbial mats, 16S rRNA amplicon sequencing, shotgun metagenome, diversity, functionality

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Asim K. Bej, University of Alabama at Birmingham, United States
Brendan Paul Burns, University of New South Wales, Australia

****Correspondence:***
Mercedes Berlanga
mberlanga@ub.edu

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

Received: *10 September 2017*
Accepted: *15 December 2017*
Published: *22 December 2017*

Citation:
Berlanga M, Palau M and Guerrero R (2017) Functional Stability and Community Dynamics during Spring and Autumn Seasons Over 3 Years in Camargue Microbial Mats. Front. Microbiol. 8:2619. doi: 10.3389/fmicb.2017.02619

INTRODUCTION

Extant microbial mats are valid equivalents of some of the Earth earliest Archaean ecosystems, which form lithified (e.g., Shark Bay, Western Australia) as well as non-lithified (e.g., Ebro Delta, north-eastern Spain) structures (Dupraz and Visscher, 2005; Wierzchos et al., 2006; Ruvindy et al., 2016). Among their features is their visible lamination, result of physicochemical gradients (e.g., light, oxygen and sulfide) along the vertical axis that allows the creation of microenvironments at a millimeter scale within the mat and justify its taxonomically and functionally heterogeneity (Guerrero and Berlanga, 2013; Harris et al., 2013; Wong et al., 2015; Saghaï et al., 2017). Microbial mats contain diverse groups of microorganisms, such as producers (e.g., photosynthetic bacteria), heterotrophs (e.g., aerobic/anaerobic respirators, especially sulfate-reducing species, and fermenters), and chemolithotrophs (notably, sulfur-oxidizing species) (Bolhuis et al., 2014).

Microorganisms do not exist in isolation (as axenic culture) but form complex ecological interaction webs, such as food-webs, by combining metabolic pathways flows (Faust and Raes, 2012; Sachs and Hollowell, 2012; Guerrero and Berlanga, 2016). Microbial mats are an extraordinary example of microbial interaction, where all possible connections (commensalism, mutualism, competition, predation, or parasitism) among microorganisms may be possible. Elucidating competitive and cooperative relationships is a challenge in describing a microbial interaction network, and interpretation of such networks is not straightforward. Population interactions, such as metabolic, physical, or signaling regulations, may determine temporal changes in the composition, function, or spatial organization of the microbial community (Widder et al., 2016). Modeled networking is a versatile tool for predicting relationships that can be due to genes (Christian et al., 2007; Großkopf and Soyer, 2014) or OTUs' presence/absence and abundance (Weiss et al., 2016). Those models can generate hypotheses on what interactions could be biologically relevant. In addition, interactions may be studied through laboratory experimental work. For instance, Long et al. (2012) tested antagonistic interaction between heterotrophic bacteria isolates from microbial mats as regulators of the community structure.

According to Liebig's law of the minimum, growth is regulated by the amount of the scarcest nutritional element available; thus, among biotic conditions, the availability of food regulates microorganismal biomass (Guerrero and Berlanga, 2006). According to Shelford's law of tolerance, each organism requires certain abiotic conditions to survive and develop (Guerrero and Berlanga, 2006). The abiotic factors influencing the distribution and function of microbial populations are principally the diel fluctuations in the concentrations of oxygen, sulfide, and other chemical nutrients and the cyclic seasonal fluctuations of inundating and desiccation (Bolhuis et al., 2014). During the day, in microbial mats can be distinguished three main chemical zones: the oxic/photic (∼0–2 mm depth) zone, the low-sulfide or transitional oxic-anoxic zone (∼2–4 mm depth), and the high-sulfide/anoxic zone (∼5 mm and deeper). At night, however, the mats become anoxic and high in hydrogen sulfide concentration, as a result of continuing sulfate reduction in the absence of oxygenic photosynthesis (Ley et al., 2006; Villanueva et al., 2007; Nielsen et al., 2015; Guerrero and Berlanga, 2016).

Microbial mats are present in several habitats such as coastal zones (e.g., Guerrero Negro, Baja California, Mexico), athalassic wetlands (e.g., Salar de Atacama, north of Chile), diverse geothermal environments (hot springs), and in polar regions. Camargue and Ebro Delta microbial mats are coastal estuarine not lithified mats from the Western Mediterranean. Camargue microbial mats were usually permanently flooded and contained more salinity concentration than Ebro Delta mats, although season temperature and latitude were similar in both microbial mats (Berlanga et al., 2008). The microbial mats in the area of Salins-de-Giraud, in the Camargue (04° 11′ E to 04° 57′ E; 43° 40′ N to 44° 40′ N), are located inside commercial salterns, which are being mined for salt. These salterns are a succession of water concentration ponds at the final part of the main mouth of the Rhone River. In the first series of ponds, seawater is concentrated to a total salinity of 50–130‰. This pond has a depth of the water column that never exceed 20 cm. In the second series, water is concentrated to salinities in the range of 130–300‰, while in the final series of ponds the salinity is increased to 340–350‰ (Fourçans et al., 2004; Guerrero and Berlanga, 2013). The vertical structure and temporal variation of microbial mats from the Camargue were previously revealed by combining molecular approaches, lipid analyses, and microscopy (Fourçans et al., 2004, 2008; Villanueva et al., 2007; Berlanga et al., 2008).

The aim of this study was to decipher the phylogenetic composition of the Camargue microbial mat community and to interpret its functional potential complexity using next-generation sequencing (NGS) methods at temporal level through three consecutive years (two season) at the same sampling place. The NGS studies used in this work included amplicon sequencing (for variant identification and phylogenetic surveys) and random-genome shotgun sequencing (for metagenomics analysis). Core samples of microbial mats from the Camargue were analyzed in detail over 3 years (2011–2013), during two different seasons, spring and autumn, and at different layers (0–2 mm, 2–4 mm, and 4–6 mm) to study the community variability of the mat.

Winter and summer seasons in the area of Salins-de-Giraud, in the Camargue had "extreme" temperature conditions, colder and warmer respectively, when compared to spring and autumn seasons. We supposed that spring and autumn had "transitional" conditions respect to temperature between those extremes. Indeed, temperature between spring and autumn in analyzed years was similar. Salinity in Camargue microbial mats was similar in all seasons through years analyzed (55–65‰). During winter, ambient temperatures are lower and daily temperature variations (day–night) are less pronounced than in summer, so less pronounced daily temperature variation in winter may have favored the adaptation of the microbial population to lower temperatures. We speculated that if there were changes in the microbial composition by temperature (cold or warm), it would be interesting to study if microbial communities could reach similar populations in spring and autumn seasons, although the initial population may be different from populations "cold

adapted" in winter and populations "warm adapted" in summer. Results could reflect the "capacity of resilience" of the Camargue microbial mat system after a perturbation such as a cold period (observed on spring samples) and after a warm period (autumn samples). The adaptation of populations to different temperatures may help to provide homeostasis within a mat community (Ward et al., 2006; Wieland and Kühl, 2006; Berlanga et al., 2008). In addition, the results will shed light on how shifts in community taxonomy may affect the relationship between biodiversity and ecosystem function. As such, they significantly enhance our understanding of the community structure of the Camargue microbial mats, their contributions to element cycling and other fundamental processes that are ongoing within the mat that are critical to the function of this ecosystem.

MATERIALS AND METHODS

Sample Collection

Samples analyzed in this study were collected at noon (12.00 h) in May (spring, SP) and November (autumn, AU) during three consecutive years (2011–2013). Environmental temperatures in May and November ranged from 15–18°C and 13.5–15°C, respectively. The mats in all cases were flooded. The salinity of the water covering the mats was 58–62‰ in May and 55–65‰ in November. Mat samples were collected in cores (1 cm × 3 cm) and frozen in liquid nitrogen immediately. Then, cores were stored in the lab at −80°C until DNA extraction. We collected three cores as in previous works (Armitage et al., 2012; Harris et al., 2013), separated by 10 cm each, for each year and season. Our samples were taken each year at the same location. Dillon et al. (2009) sampled cores across 1 km. They observed that population structure diverged with increasing distance between sample sites, but positional replicates were highly similar among samples < 1 m distance. We pooled the extracted DNA for the three samples corresponding to year/season and layer to obtain a representative sample for each year/season/layer. We expected that if there were differences in microbial composition it could be due to seasonal environmental variables and not to location of sampling.

DNA Extraction and Amplification

The frozen cores were sliced with a sterile blade in aseptic conditions horizontally in 2-mm increments (from the top to a depth of 6 mm): 0–2 mm (layer 1, oxic/photic layer), 2–4 mm (layer 2, oxic-anoxic transition layer), and 4–6 mm (layer 3, anoxic layer). A fresh blade was used at each interface. Then, pieces of microbial mat of approximately 3 mm^3 were cut from each slice and suspended in 100 μl of TE buffer in 2.0-ml vials containing a capful of 0.1-mm glass beads. The mixture was homogenized for 1 min in a Mini-BeadBeater-8 (Biospec Products, Bartlesville, OK, United States), and centrifuged at high speed for 2 min. While avoiding transfer of the beads, ∼500 μl from each sample was pipetted into sterile 1.5-ml Eppendorf tubes. DNA was extracted using a phenol-chloroform mixture and precipitated in the cold using 95% ethanol. Three DNA extractions corresponding to every year, season and layer were performed. The DNAs obtained were mixed to correct for potential local heterogeneity effects to obtain a representative sample for each year, season and layer.

For years 1, 2, and 3, we performed amplicon sequencing of the bacterial 16S rDNA gene. The primers used for multiplex Roche 454 GS FLX pyrosequencing, contained a 25 nucleotide sequence adapter, 10-base-pair molecular barcode (multiplex identifier), and the universal bacterial sequence for the region V1–V2, 8F-338R (5′-AGAGTTTGATCCTGGCTCAG-3′ and 5′-TGCTGCCTCCCGTAGGAGT-3′) (Armitage et al., 2012; Harris et al., 2013; Yang et al., 2013). We used three different barcodes (each one for 0–1 mm, 2–4 mm, and 4–6 mm; ACGAGTGCGT, ACGCTCGACA, AGACGCACTC, respectively). Samples analyzed were: SP1-1, SP1-2, SP1-3; AU1-1, AU1-2, AU1-3; SP2-1, SP2-2, SP2-3; AU2-1, AU2-2, AU2-3; SP3-1, SP3-2, SP3-3; AU3-1, AU3-2, AU3-3 (SP and AU indicated the season analyzed; the first number, the year, and the second number, the layer). A PCR from each DNA was performed. The cycling conditions were 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 56°C for 40 s, 68°C for 40 s, and a final extension step at 68°C for 6 min. The resulting product was checked for size and purity on an agarose-SYBR Safe DNA gel that was subsequently stained (Invitrogen, San Diego, CA, United States). The amplicons were purified using a Pure Link kit (Invitrogen, San Diego, CA, United States) and quantified using Qubit and Bioanalyzer (Berlanga et al., 2016). A pool of amplicons was mixed in equimolar amounts (e.g., spring 1st year, amplicons obtained for 0–2, 2–4, and 4–6 mm), and then prepared for 454-pyrosequencing according to the manufacturer's instructions. Pyrosequencing coverage (depth sequencing) resulted in 99,216 total raw reads that after quality control processing resulted in 44,787 reads (see bioinformatic analyses section) for the 18 samples.

Shotgun metagenomic analysis was performed on samples belonging to the third year (SP3-1, SP3-2, SP3-3; AU3-1, AU3-2; AU3-3). We repeated the DNA extraction several times to reach the approximate concentration of 500 ng to 1 μg of DNA for each sample. Random shotgun metagenomics was performed in the Unity of Genomics of Scientific and Technological Centers, University of Barcelona (CCiTUB). Number of sequences ranged from 61370 to 140208. Major scaffold distribution lengths were 390–470 bp.

We combined 16S rRNA amplicons, PICRUSt and shotgun metagenomics using the best of each method to obtain the maximal information to try to describe precisely the taxonomical structure and functionality of the samples. The advantages of using 16S rRNA amplicons sequencing had normally better taxonomic resolution than shotgun metagenomics (Tessler et al., 2017), and the availability of bioinformatic tools for prediction of functions (PICRUSt) is particularly attractive to microbial ecologists as it allows them to study the genes (functions) of complex microbial communities with reasonable accuracy at a high taxonomic resolution (Mukherjee et al., 2017). Random shotgun sequencing of environmental DNA provides a direct and potentially less biased view of the functional attributes of microbial communities (Klatt et al., 2013). 16S rRNA gene regions recovered from the shotgun metagenomic data can span

the entire length of the genes; the PCR-based amplicon approach only targets the V1–V2 region. Therefore, the two approaches may not necessarily give identical results (Fierer et al., 2012).

Bioinformatics Analyses

For 16S rRNA amplicons, the raw data of each sample was preprocessed for demultiplex and quality control using a pipeline implemented in GPRO version 1.1 (Futami et al., 2011). Raw reads that contained < 150 nucleotides in size, ambiguities > 1, homopolimer > 8, as well as redundant sequences were removed from each metagenome dataset using screen.seqs and unique.seqs by Mothur1.31.2 (Schloss et al., 2009). Sequences were taxonomically classified using Silva database[1] (Quast et al., 2013). CD-HIT-EST from the CD-HIT 4.5.4 package (Fu et al., 2012) was used to define clusters of clones within each metagenome with a distance threshold of 0.03 (resulting in a cutoff at the species level). Alpha and beta diversity analyses of all samples were performed at 97% distance level of OTU. For diversity we rarified (normalized) samples to compare all the samples. Weighted UniFrac metrics was used to measure beta-diversity and to generate principal coordinates analysis plots, using the normalized OTU table. For the heatmap we used the OTU table at 0.10% genetic distance level (resulting in a cutoff at the family level) (Yarza et al., 2010), and make_otu_heatmap.py, and the script was modified by Stamp program. Hierarchical cluster analysis used for similarity measure was Pearson's correlation, and for the clustering algorithms, Ward's linkage.

Core microbiota was determined using compute_core_microbiome.py in qiime[2] (Caporaso et al., 2010). Core OTUs were defined as the OTUs that are present in at least 90% of the samples. From the set of OTUs that could be considered the core, we performed an ecological network of interactions. Ecological network was achieved by Molecular Ecological Network Approach (MENA)[3] (Deng et al., 2012). Ecological network worked with RMT (random matrix). To visualize the network it was used Cytoscape 3.5.1.

Metagenomes were predicted from the 16S rRNA data using PICRUSt (Langille et al., 2013) for samples corresponding to years 1, 2, and 3. This was prepared by the predict_metagenomes.py script against functional database of KEGG Orthology. Functional contributions of various taxa to different KOs were computed with the script metagenome_contributions.py (Mukherjee et al., 2017). For the third-year samples, gene annotation of the shotgun method was analyzed by the United States Department of Energy Joint Genome Institute[4] (Nordberg et al., 2014).

The DOE-JGI Metagenome Annotation Pipeline (MAP) supports the annotation of metagenomic sequences and it is organized in three stages: sequence data pre-processing, structural annotation, functional annotation and phylogenetic lineage prediction. Some of the processing methodology used by MAP was as follows: Unassembled 454 reads containing more than five occurrences of 'N's are removed. Sequences shorter than 150 bp after trimming are also removed. When two or more sequences are at least 95% identical, with their first 3 bp being identical as well, those sequences are considered to be replicates and only the longer copy is retained. For genomic assembler it is used the Velvet algorithm package. A good kmer size is just over half a read length, which prevents sequencing errors from forming bubbles. Ribosomal RNA genes (5S, 16S, 18S, 23S) are predicted using hmmsearch tool from the package HMMER 3.1b2. The pipeline runs against curated models, derived from full-length genes within IMG, while keeping the best scoring models. The identification of protein-coding genes is performed using a consensus of four different *ab initio* gene prediction tools: prokaryotic GeneMark.hmm (v.2.8), MetaGeneAnnotator (v. Aug 2008), Prodigal (v. 2.6.2) and FragGeneScan. Protein-coding genes with translations shorter than 32 amino acids are deleted. Assignment was made at 90% of the KO gene sequence that was covered by the alignment (Huntemann et al., 2016).

The numbers of the analysis projects in the JGI were Ga0197827, Ga0197828, Ga0197830, Ga0197833, Ga0197836, Ga0197838. For the 16S rRNA amplicons, sequence data were deposited on the NCBI database by the Bioproject PRJNA416849.

RESULTS

Phylogenetic Stratigraphy in the Camargue Microbial Mats

Camargue microbial mats composition of Bacteria, Archaea and Eukarya were based on data obtained by shotgun metagenomics. Microbial communities were dominated by Bacteria (92.4–94%), while Archaea and Eukarya represented 4–5% and 1–1.6%, respectively. The distribution of Archaea phyla in spring and autumn was similar but there were several differences across the three depths sampled (0–2, 2–4, and 4–6 mm). Thus, Archaeal relative abundances at those depths were 4.2, 4.6, and 5.1%, respectively. The major phyla were Euryarchaeota (80.6%), followed by Crenarchaeota (8%), *Candidatus* Micrarchaeota (3.9%), and Thaumarchaeota (ammonia-oxidizing archaea, 2.8%).

The eukaryotic diversity of the Camargue microbial mats was sparse, in contrast to the vast bacterial diversity. This was probably due to the broad metabolic capabilities of Bacteria, which enable them to occupy a broad range of chemical niches, whereas the metabolic versatility of eukaryotes is more limited, despite their ability to survive under high sulfide, fermentative, and anoxic conditions. Eukarya represented 1% of the total relative abundance of microorganisms from the mat, with the most representative eukaryotes those related to algae (Chlorophyta), plants (Streptophyta), fungi (Ascomycota), and Arthropoda (insects, mainly *Anopheles*). This result contrasts with the findings in the Guerrero Negro microbial mats, where the dominant eukaryotic organisms are bacterivorous nematodes (Feazel et al., 2008).

More than 30 phyla of Bacteria were recovered from amplicon sequencing of the 16S rRNA gene and shotgun

[1] http://www.arb-silva.de

[2] http://qiime.org/scripts/compute_core_microbiome.html

[3] http://ieg2.ou.edu/MENA

[4] http://www.jgi.doe.gov/

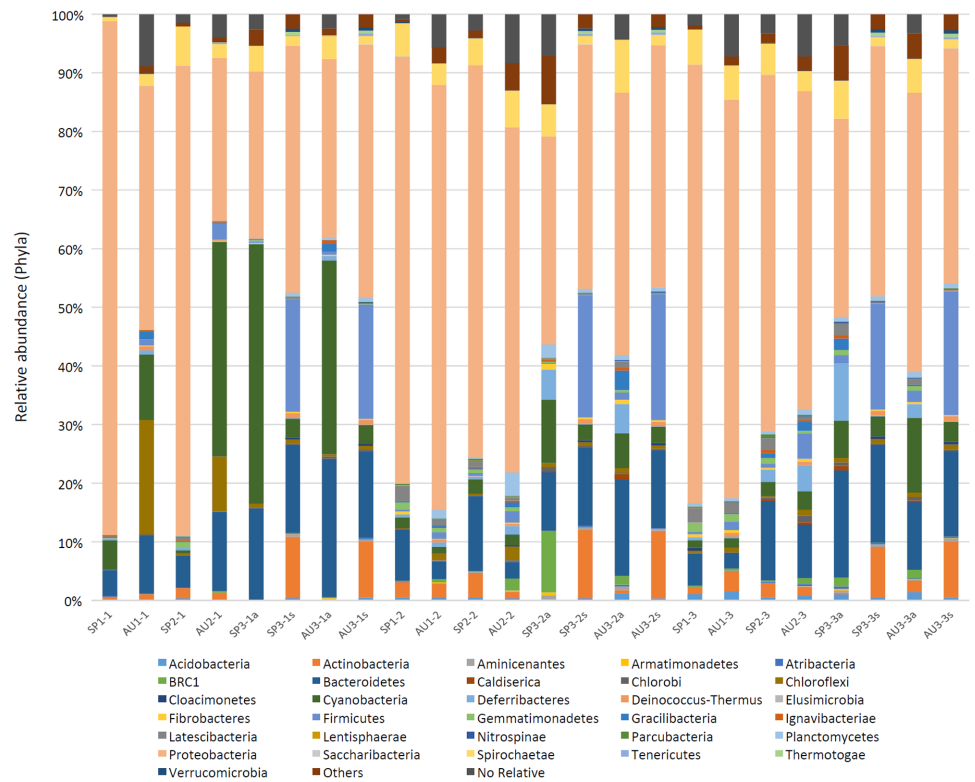


FIGURE 1 | Relative abundances of bacterial phyla from the Camargue microbial mats. Samples SP1-1, SP1-2, SP1-3; AU1-1, AU1-2, AU1-3; SP2-1, SP2-2, SP2-3; AU2-1, AU2-2, AU2-3; SP3-1, SP3-2, SP3-3; AU3-1, AU3-2; AU3-3. SP and AU indicate the season analyzed; the first number, year 1, 2, or 3; and the second number, the layer (1, 0–2 mm; 2, 2–4 mm; 3, 4–6 mm). Samples for the third year were accompanied by a and s, amplicons and shotgun, respectively.

metagenomics isolated from the Camargue microbial mats. Among the distinguished phyla, there were six that dominated: Proteobacteria (40.2–75.2%), Bacteroidetes (2.6–16.6%), Firmicutes (1.2–21.5%), Actinobacteria (1.0–10.4%), Cyanobacteria (1.2–36.6%), and Spirochaetes (1.4–6.6%) (**Figure 1**). The use of shotgun metagenome analysis in the 3rd-year samples did not yield additional phyla compared to the 16S rRNA amplicons analyzed using Silva database. However, it did detect a relatively high abundance of Actinobacteria, Bacteroidetes, and Firmicutes, and a lower relative abundance of Proteobacteria. The distributions of several phyla and their families depended on the layer analyzed and the season (**Figure 2** and **Supplementary Figures S1A,B**). Cyanobacteria were more abundant in the upper layer (0–2 mm) and in autumn. Alphaproteobacteria were the most abundant Proteobacteria, followed by Gammaproteobacteria and Deltaproteobacteria. Alphaproteobacteria, especially the family *Rhodobacteraceae*, were abundant in the upper layers (0–2 and 2–4 mm) and in the spring. Gammaproteobacteria were represented, in order descendent of relative abundance, by *Chromatiaceae*, *Ectothiorhodospiraceae*, and *Pseudoalteromonadaceae*. The *Chromatiaceace* family was present in all layers with a slightly increased abundance at 2–4 and 4–6 mm, but no difference between spring and autumn. *Ectothiorhodospiraceae* were more abundant in the upper layer (0–2 mm) than in the other, deeper layers. *Pseudoalteromonadaceae* were detected only in the autumn samples. Among the Deltaproteobacteria, *Desulfobacteraceae* was the most abundant family detected and their distribution in the different layers and in the two seasons was similar. *Chloroflexales* (phylum Chloroflexi) were more abundant in the autumn samples and in the upper layer. Several genera, such as *Thioalkalivibrio*, *Desulfotigum*, *Roseovarius*, etc., were detected through the years analyzed, but their distribution depended on the layer (**Figure 2**).

In the rarefaction curves for samples with respect to depth and season, at 0.03 similarity the samples did not reach an asymptote, suggesting their insufficient sequencing depths (**Supplementary Figure S2**). In our analyses we used the abundance estimator Chao1 and abundance-based coverage estimator (ACE), Shannon and Simpson diversity indexes, and Berger-Parker dominance index (**Table 1**). The highest diversity was found in the third layer (4–6 mm). The surface layer in all samples exhibited the lowest diversity, with a few strongly dominant OTUs, especially photosynthetic bacteria. A principal component analysis indicated that the population community structure of the upper layer (oxic/photic layer) differed from that of the transitional oxic-anoxic layer (2–4 mm) and the anoxic layer (4–6 mm), which were relatively close together (**Figure 3A**). These results were independent of the season and the sampling year. Samples for the oxic/photic layer (0–2 mm)

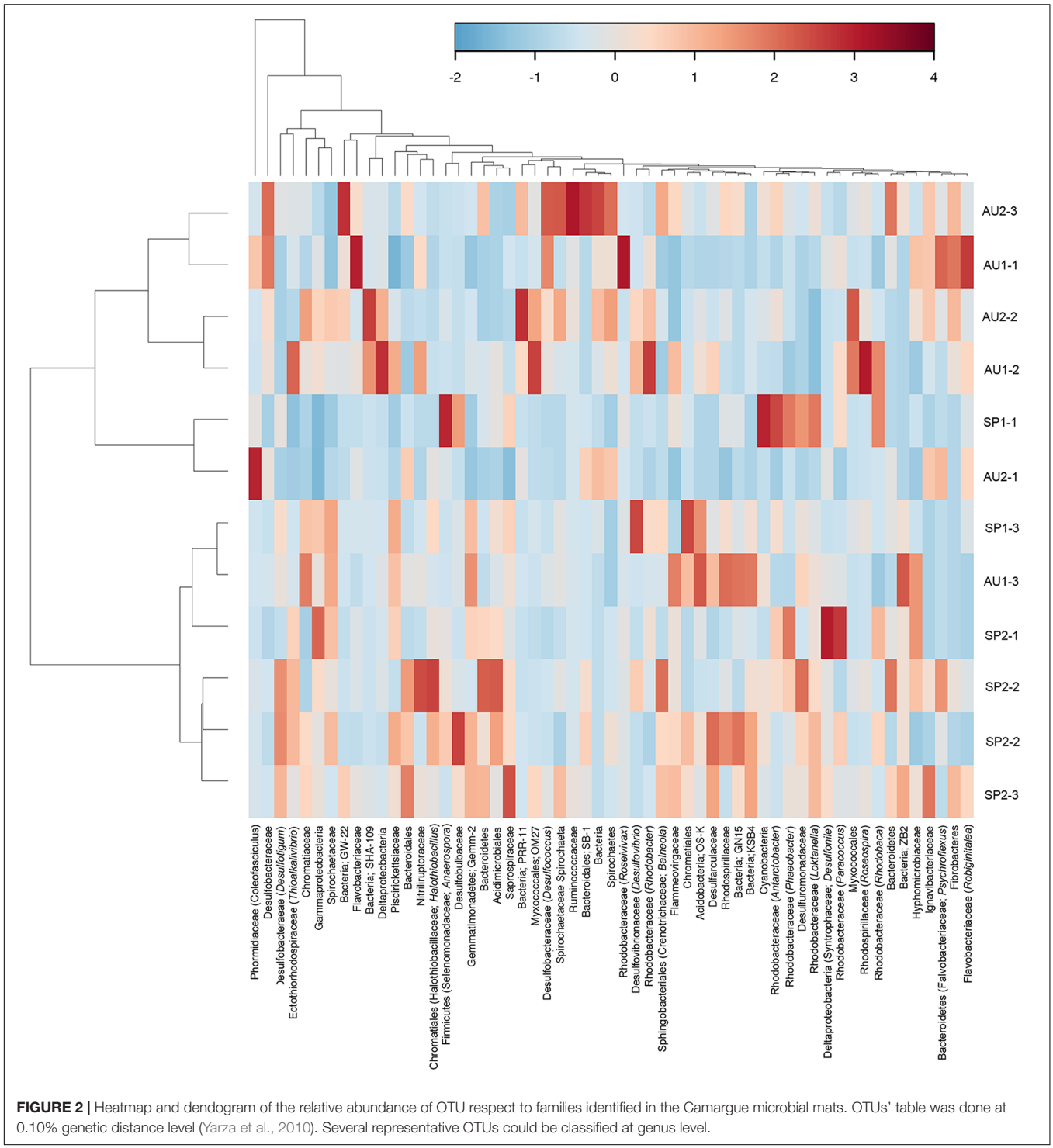


FIGURE 2 | Heatmap and dendogram of the relative abundance of OTU respect to families identified in the Camargue microbial mats. OTUs' table was done at 0.10% genetic distance level (Yarza et al., 2010). Several representative OTUs could be classified at genus level.

had the most distant community distribution, probably due to the high relative abundances of Cyanobacteria and Chloroflexi. The phylogenetic *P*-test in Unifrac indicated that the microbial communities were not significant different ($P > 0.05$). But pairwise significance tests using the t-Student based on taxa detected showed significant differences between oxic/photic layer and the other layers (transition oxic-anoxic layer and anoxic layer). Significant differences between two samples were based on a 95.0% confidence level. $P < 0.05$ was considered to indicate statistical significance. No significant differences were observed between samples from the same layer. In addition, we could observe a clear difference in distribution respect of the functional annotation KO metagenomes to layers (data from shotgun metagenome for the third year samples) (**Figure 3B**).

TABLE 1 | Diversity and richness indexes at 0.03 distance.

Sample[a]	Raw reads	Clean reads	No. OTUs	Chao 1	Ace	Bergerparker	Shannon	Simpson
SP1-1	7811	4114	205	634.04	856.45	0.127	4.51	0.0330
SP1-2	6656	3453	325	1206	1634.46	0.0588	5.33	0.0126
SP1-3	8285	3664	397	1387.5	1939.02	0.0392	5.71	0.0082
AU1-1	2825	1156	265	505.01	777.53	0.1203	4.38	0.0392
AU1-2	3053	1313	272	566.25	778.05	0.0422	4.89	0.0137
AU1-3	5361	1432	336	839.25	1148.86	0.0616	5.20	0.0121
SP2-1	3560	2042	291	896.01	1197.66	0.0929	5.02	0.0193
SP2-2	5393	2366	316	876.16	1113.66	0.0605	5.26	0.0119
SP2-3	8054	2866	454	1501.38	2265.98	0.0348	5.91	0.0058
AU2-1	3924	1859	282	711.37	1049.50	0.2478	4.13	0.0805
AU2-2	6022	2778	365	923	1190.00	0.0552	5.40	0.0118
AU2-3	4673	2405	463	1757	2591.69	0.0602	5.79	0.0082
SP3-1	3103	1151	423	765	1325.63	0.0417	5.28	0.0045
SP3-2	6728	3252	390	621	1678.86	0.0147	5.68	0.0009
SP3-3	6919	3130	401	878	1693.39	0.0078	5.86	0.0002
AU3-1	4989	2092	412	808	2188.89	0.0147	5.63	0.0011
AU3-2	5269	3053	433	774	1468.0	0.0104	5.81	0.0004
AU3-3	6591	2661	429	962	2188.9	0.014	5.89	0.0005

[a]*SP and AU indicate the season analyzed; the first number, year 1, 2, or 3; and the second number, the layer (1, 0–2 mm; 2, 2–4 mm; 3, 4–6 mm).*

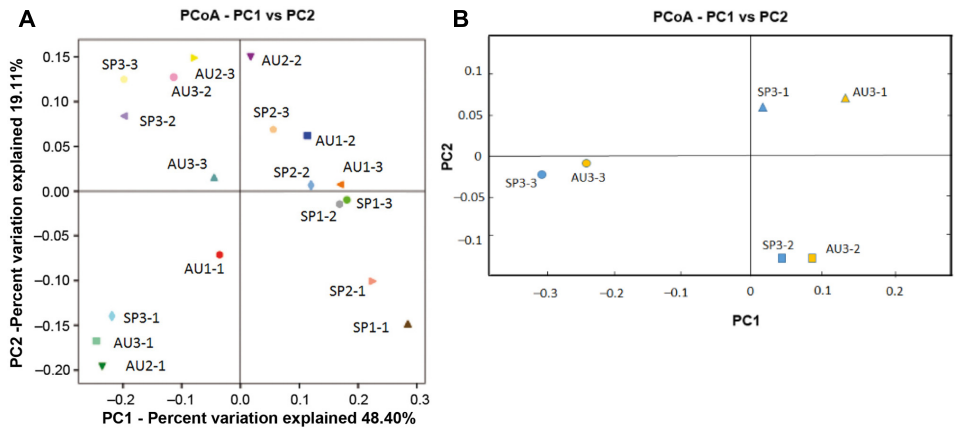


FIGURE 3 | Principal component analysis of the community distribution by year, season and layer from the Camargue microbial mats. **(A)** β-Diversity coupled with principal coordinates analysis was used to compare the bacterial composition in Camargue microbial mats by season and layer. Unifrac weighted was implemented in the QIIME program (Caporaso et al., 2010). Red squares represented the oxic/photic layer; Blue triangles, oxic-anoxic transition layer; Orange, anoxic layer. The phylogenetic P-test in Unifrac, indicated that the microbial communities were not significant different ($P > 0.05$). But pairwise significance tests using the t-Student showed significant differences between oxic/photic layer and the other layers (transition oxic-anoxic layer and anoxic layer). **(B)** Principal coordinates analysis of functional annotation of shotgun metagenomes processed in the JGI database [http://www.jgi.doe.gov/]. To compare the genomes (third year samples), we used the KO genes as a row with significant hits, and with a minimal function gene count for 5. The PCA analysis showed that the PC1 the percent of variation was explained by 18.7% and the PC2, 26.57%. The t-Student had similar results than for the taxonomical results in A.

The t-Student from taxonomical data showed the same results that those just mentioned.

Functional Stratigraphy in the Camargue Microbial Mats

To understand the metabolic potential of the Camargue microbial mats and identify their many different functional features, we used PICRUSt (based on 16S rRNA gene amplicon) and random shotgun metagenomics methodologies. The predicted proteins were classified as KEGG orthologs (KOs). The nearest sequence taxon index (NSTI) values is a measure of how closely related the OTUs in each sample are to the reference genomes in the database. In our case, the "nearest sequence taxon index" (NSTI) values per sample ranged for 0.072–0.172. The taxonomical classification could be accurate at family level, in few cases to genera level, but it was difficult to achieve the species level. This result could explain the values observed at NSTI. Respect to the shotgun metagenomics, KEGG pathways via KO (percentage) ranged from 14.85 to 17.17%; and KO genes ranged from 24.97 to 29.30% respect to the number of sequences

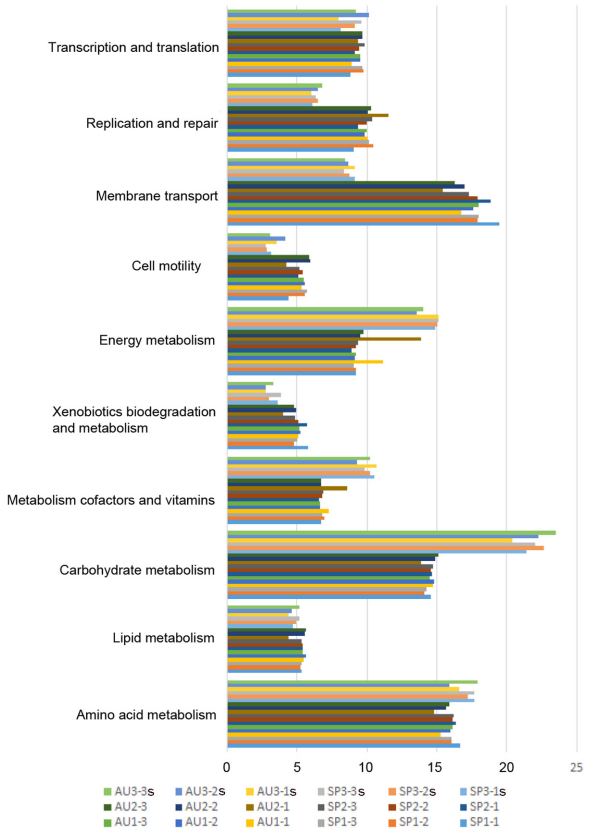


FIGURE 4 | Relative abundances of genes in KEGG classified by biological functional category from the Camargue microbial mats. Functional categories studied were: "metabolism," "cellular processes," "environmental information," and "genetic information processing." SP and AU indicate the season analyzed; the first number, year 1, 2, or 3; and the second number, the layer (1, 0–2 mm; 2, 2–4 mm; 3, 4–6 mm). For the third year, only data obtained from the shotgun (e.g. SP3-1s, etc.) has been represented.

(total sequences analyzed ranged from 61,370 to 140,208), and assignment was made at 90% of the KO gene sequence that was covered by the alignment (Huntemann et al., 2016).

The biological processes identified were essential for sustaining prokaryotic life in the environment. They include transcription and translation functions (8.7–9.8% relative abundance genes, based on the total number of genes detected in the sample) and replication and repair functions (9–10.2%). Other functional processes were related to cellular processes such as cell motility (4.3–5.9%). Genes related to membrane transport (17.2–19.5%) and to metabolic functions (56.3–58.4%), which included the metabolism of carbohydrates, lipids, amino acid, cofactors and vitamins, xenobiotic biodegradation, and energy metabolism. PICRUSt and shotgun metagenomic analyses revealed similar functional biological processes in all samples analyzed, except of carbohydrate metabolism and energy metabolism, which contained more genes detected by shotgun than by PICRUSt analyses (**Figure 4**).

Gene content analysis provides a basis for inferring the possible metabolic functions of dominant populations present in the community. Cell motility, represented by chemotaxis genes, such as *cherA, cheBR, motA, mcp, pixJ*, etc., peaked at the oxic-anoxic transition zone, but they were also important in the oxic/photic zone. These genes were associated to phototrophic organisms (Cyanobacteria, Alpha-Gammaproteobacteria, and Chloroflexi), but also to heterotrophic members, such as Bacteroidetes and Spirochaetes. Ferredoxins have a negative redox potential act as electron distributors in various metabolic pathways. The genes that codify different ferredoxins were detected in all layers, but they were especially abundant in the oxic-anoxic transition zone. Ferredoxins likely reflected diversification of redox reactions required for respiration (**Supplementary Table S1**). Osmotic regulation is required for microbial survival in hypersaline environments. Accumulation of osmoprotective molecules, in particular glycine betaine, is an adaptive mechanism to pawn the high salinity conditions. We searched for genes implicated on the glycine-betaine biosynthesis, such as *betA, betB, gbsA* (Wong et al., 2015). These genes were distributed through the layers (especially on the oxic/photic layer), and they were associated with different taxa, showing that the microbial mat community could be adapted to salinity conditions (**Supplementary Table S1**).

Genes associated to oxygenic photosynthesis and bacteriochlorphylls were detected in the upper layer (photic zone) in autumn and spring samples (Table S1). Regarding the photosynthetic reaction center in the anoxygenic photobreaksystem, *pufL* and *pufM* genes were detected and they belonged to Gammaproteobacteria (purple sulfur bacteria, *Chromatiaceae*). The possibility of alternative light energy usage by (bacterio)rhodopsin in different prokaryotic members of the mat cannot be confirmed because we could not detect related genes, even though retinal-based phototrophy could contribute as energy source in layers with low irradiance (Thiel et al., 2017).

In the studied metagenomes we identified the four known autotrophic carbon fixation pathways (the Calvin-Benson cycle, the reverse tricarboxylic acid cycle, the Wood–Ljundahl pathway, and the 3-hydroxypropionate bi-cycle) (Hügler et al., 2002, 2005; Ragsdale and Pierce, 2008; Berg, 2011; Thiel et al., 2017), which suggested the occurrence of a relatively diverse autotrophic community (**Supplementary Table S1**). Debris of predated bacteria by *Bdellovibrionaceae* and viruses may be another carbon source for other heterotrophs. *Bdellovibrionaceae* represented 1–9% of the relative abundance of Deltaproteobacteria. *Bdellovibrionaceae* are predatory bacteria upon a variety of Gram-negative bacteria. Viruses in the mat could be also involved in cell-lysis processes, based on CRISPR systems detected in the metagenomes.

Regarding nitrogen metabolism, we detected genes associated to nitrogen fixation. The oxic/photic zone contained the more diverse and abundant amount of nitrogen fixation genes (**Supplementary Table S1**). For the ammonium oxidation in nitrification, the main enzyme is the ammonia monooxygenase (*amoA*) that is present in both ammonia-oxidizing archaea and ammonium-oxidizing bacteria (Fan et al., 2015). However, *amoA* was not identified in the studied metagenomes. Nitritification provides the oxidant for anaerobic ammonium oxidation (anammox). We examined *hzoA/hzoB*

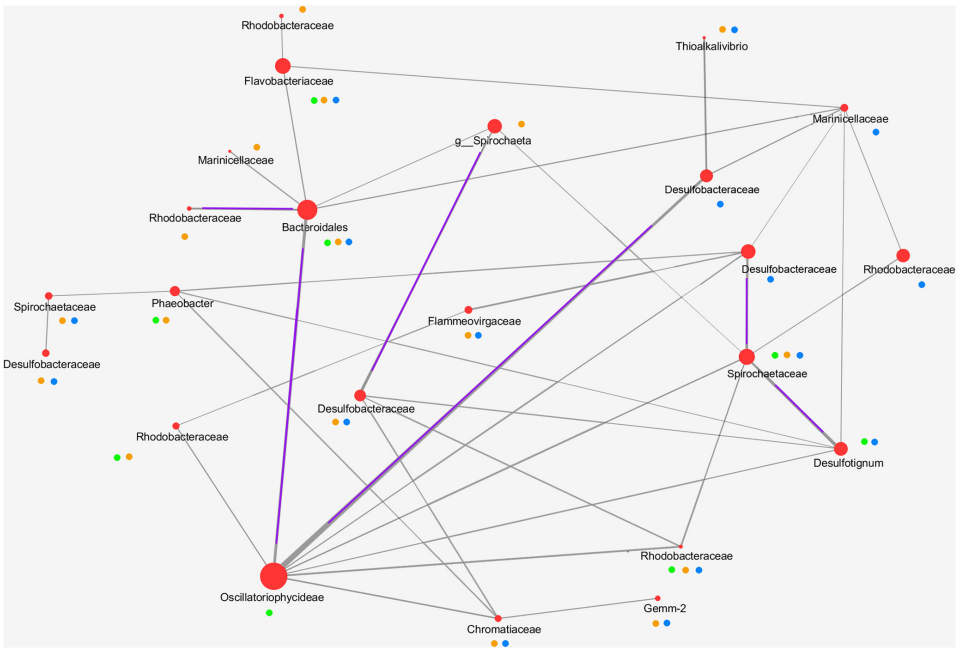


FIGURE 5 | Network showing potential interactions among microbial members of the Camargue microbial mats. Network was constructed by the in base of the "core" community using the Molecular Ecological Network Approach (MENA). Matrix similarity used the Pearson correlation coefficient. Nodes represented the representative OUT, and the size, their relative abundance. The thickness of the lines represented the strength of the correlation between different taxa. These positive interactions were partially marked in the figure with a purple line. Green, orange and blue circles indicated the layer (oxic/photic; oxic-anoxic transition and anoxic layers, respectively) where microorganisms were detected.

genes because their ubiquity and high expression in anammox bacteria (Planctomycetes) (Hirsch et al., 2011), but no records of those genes were found, although the phylum was detected in the Camargue microbial mats.

Sulfate reduction genes were present in the metagenome dataset and distributed similarly through the different layers (**Supplementary Table S1**). They affiliated to Deltaproteobacteria and Gammaproteobacteria. Sulfur oxidation activity was also found in the Camargue microbial mats, based upon the presence of the enzyme sulfide:quinone oxidoreductase (*sqr* gene).

Finally, to identify potential biotic interactions within the dominant, prokaryotic communities in the Camargue microbial mats, we constructed a network based on the core OTUs (**Figure 5**). The core OTU were determined by the shared OTUs at 90% in all samples. Several minority populations were not included as the "core community," and probably they could play important functions. Core community was performed by layer. We observed that there were no significant differences for one layer (e.g., 0–2 mm or 2–4 mm or 4–6 mm) among years and season. In the upper layer (oxic/photic zone), the core microbiota were represented Alphaproteobacteria (*Rhodobacterales*), Gammaproteobacteria (*Marinicellales* and *Chromatiales—Ectothiorhodospiraceae*) and Cyanobacteria, (*Coleofasciculus* [formerly, *Microcoleus*], *Oscillatoriales*). In the middle layer (transition oxic-anoxic zone), there were OTUs belonging to Alphaproteobacteria (*Rhodobacterales*), to Gammaproteobacteria (*Marinicellales, Chromatiales, Thiotrichales*), to Deltaproteobacteria (*Desulfobacterales*), to Bacteroidetes (*Flavobacteriales* and *Cytophagales*), to Spirochaetes, and to Cyanobacteria (*Oscillatoriales, Coleofasciculus*). In the bottom layer (anoxic zone), taxons were represented by Alphaproteobacteria (*Rhodobacterales*), OTUs to Gammaproteobacteria (*Marinicellales* and *Chromatiales*), Deltaproteobacteria (*Desulfobacterales*), Bacteroidetes (*Flavobacteriales. Bacteroidales* and *Cytophagales*), Spirochaetes, Planctomycetes, and Gemmatimonadetes. The network obtained showed that interactions among taxa could be done through different layers (**Figure 5**). The network was probably incomplete because there were not represented other populations, such as Firmicutes, Actinobacteria, Archaea, or other minor population with less than 0.1% of relative abundance (Planctomycetes, Nitrospinae, Saccharibacteria, etc.), which they could contribute and participate on metabolically interactions within the microbial mat. Cyanobacteria stablished the more diverse interactions with different population's members of the microbial mat. Positive correlations based the thickness of the lines (in the figure marked by a purple line) and were observed between Cyanobacteria and Deltaproteobacteria; Cyanobacteria and Bacteroidetes; Cyanobacteria and Rhodobacterales; and Spirochaetes and Deltaproteobacteria.

DISCUSSION

Microbial diversity within an ecosystem has most often been estimated based on the amplification of specific gene targets

(e.g., 16S rRNA) and random shotgun sequencing (Klatt et al., 2013; Warden et al., 2016; Cardoso et al., 2017). Our results shed light on the diversity of microbial communities, such as Bacteria (92.4–94% relative abundance), Archaea (4–5%) and Eukarya (1–1.6%), present in the Camargue microbial mats. Although an intrinsic bias of the method cannot be ruled out, as already noted by other authors (Amend et al., 2010; Zhou et al., 2011). Cardoso et al. (2017) found differences in taxonomic assignment based on whether the variable region of 16S rRNA V1–V3 vs. 16S rRNA V3–V4 sequences derived from the DNA template. They detected a higher abundance of Proteobacteria using the V1–V3 than the V3–V4 region, whereas the abundances of Bacteroidetes, Chloroflexi, and, particularly, some rare phyla were lower using the V1–V3 dataset. Nonetheless, the amplicon sequencing and shotgun metagenomics data obtained in this study confirmed the importance and numerical dominance of Proteobacteria in the Camargue microbial mats as well as in mats from elsewhere in the world (Ley et al., 2006; Ruvindy et al., 2016; Warden et al., 2016; Cardoso et al., 2017). Proteobacteria participate in the sulfur cycle, especially the purple sulfur bacteria belonging to the Gammaproteobacteria, purple non-sulfur bacteria belonging to the Alphaproteobacteria, and the sulfate-reducing bacteria belonging to the Deltaproteobacteria (Bolhuis et al., 2014; Ruvindy et al., 2016).

The distribution of Archaea phyla in spring and autumn was similar but there were several differences across the three depths. Euryarchaeota at the surface were dominated by the classes *Halobacteria* and *Methanomicrobia,* and in the deeper layers by *Methanobacteria*, *Halobacteria*, *Methanomicrobia*, and *Methanococci* (in order of their relative abundances). The dominance of Euryarchaeota was also described for other hypersaline microbial mats, except Guerrero Negro, which is dominated by Crenarchaeota (Robertson et al., 2009; Schneider et al., 2013; Fernandez et al., 2016; Wong et al., 2017).

Eukarya represented 1% of the total relative abundance of microorganisms present in the mat. The eukaryotic diversity of the Camargue mat was sparse, in contrast to the vast bacterial diversity. This was probably due to the broad metabolic capabilities of Bacteria, which enable them to occupy a broad range of chemical niches, whereas the metabolic versatility of eukaryotes is more limited. Also, some environmental factors such as salinity, oxygen and sulfide gradients could be limiting factors for the eukaryotic diversity. Halophiles are found in all three domains of life and they are components of brine communities. Within the Bacteria: *Cyanobacteria*, *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Spirochaetes*, and *Bacteroidetes*. Within the Archaea: *Halobacteria*, and for eukaryotes, Alveolates (ciliates and dinoflagellates), several Fungi (e.g., Wallemia, *Trimmatostroma, Hortaea*), chlorophytes, Euglenozoans, shrimp (e.g., *Artemia)* (Oren, 2008). Salinity (only) probably is not a limiting environmental factor for the development of eukaryotes, but their combination with daily changes in oxygen and sulfide may affect their survival in microbial mats.

Cyanobacteria were detected in the Camargue microbial mats in relative low numbers and they were dominated by the species *Coleofasciculus* (formerly *Microcoleus*) *chthonoplastes* (*Oscillatoriales*). This result may be a consequence of the methodology used, as the efficiency of cell lysis strongly varies among different microorganisms. Filamentous Cyanobacteria are heavily encapsulated by exopolysaccharides (EPS) and therefore they are difficult to lyse. Moreover, even when lysis is successful, nucleic acids may become trapped in the EPS and thus inaccessible for PCR and sequencing (Bolhuis et al., 2014). Ramos et al. (2017) reported that studies using Cyanobacteria-specific primers rendered high cyanobacterial diversity. However, the scarcity of cyanobacteria and their low diversity have been described in several mats (Ley et al., 2006; Fernandez et al., 2016).

In microbial mats, the import and export of microorganisms are low and the community composition is accordingly stable (Cardoso et al., 2017). The environmental conditions, including temperature and salinity, during the sampling period were not sufficiently different to significantly modify the microbial communities. Rather, vertical gradients of light and redox (oxic-anoxic) conditions were the likely determinants of mat community structure. In the presence of oxygen land high light intensity (oxic/photic zone, 0–2 mm), the prokaryotic communities in the surface layers were mainly composed of Cyanobacteria and anoxygenic phototrophs (Alphaproteobacteria, represented mainly by purple non-sulfur bacteria, *Rhodobacterales*, and *Rhodospirillaceae*). However, while *Rhodobacterales* species may prosper in the surface layer of the mat, most *Rhodospirillaceae* prefer anoxic conditions (Schneider et al., 2013). Archaea in the surface layer were represented by *Candidatus* Micrarchaeota, ammonia-oxidizing Thaumarchaeota, and Euryarchaeota (*Halobacteria*). *Halobacteria* uses bacteriorhodopsin to transform light energy into chemical energy by a process unrelated to chlorophyll-based photosynthesis. Chemotaxis and motility genes were assigned to phototrophs, such as Cyanobacteria, purple sulfur bacteria, and purple non-sulfur bacteria, consistent with the ability of these microorganisms to search for optimal environmental conditions, including light. The surface layer in all samples exhibited the lowest diversity, with a few strongly dominant OTUs, especially photosynthetic bacteria as observed by other authors (Armitage et al., 2012; Al-Najjar et al., 2014; Bolhuis et al., 2014). Upper layers can have extreme physicochemical conditions if the mat is desiccated, during the day it may have high light irradiance, temperature, and high salinity due to water evaporation. In our case, mats were flooded all seasons (cover by ca. 10–20 cm of water). Cyanobacteria were the major phylum at the top layer and they may be adapted to those conditions (Bolhuis et al., 2014; Al-Najjar et al., 2014; Pade and Hagemann, 2015).

The transition zone (2–4 mm) contained Alphaproteobacteria (purple non-sulfur bacteria), Gammaproteobacteria (*Chromatiaceae* —such as *Thiohalocapsa* and *Halochromatium*— were more abundant than *Ectothiorhodospiraceae* —such as *Thioalkalivibrio*—), *Candidatus* Chlorothrix, which was the most abundant genus of green non-sulfur bacteria of *Chloroflexales* (Chloroflexi), Deltaproteobacteria (sulfur reducing bacteria), and heterotrophic fermenting bacteria. In the anoxic zone (4–6 mm), Deltaproteobacteria and fermenters, especially Spirochaetes, comprised the major part of the bacterial community.

Microorganisms detected through layers could interact metabolically by the different genes detected that worked in carbon, nitrogen and sulfur nutrients cycles, being a self-sustaining system. Producers, such as photosynthetic microorganisms contributes to the nourishment of heterotrophs members of the community. Cyanobacteria (mainly *Coleofasciculus*) must provide a source of carbon to the heterotrophs, and a source of H_2 for sulfate-reducing bacteria (Deltaproteobacteria, such as *Desulfonema*, *Desulfotigum*, *Desulfococcus*, *Desulfonile*) (Lee et al., 2014). In the dark, cyanobacteria fermented their carbon reserves excreting low-molecular-weight organic acids and hydrogen (Hoffmann et al., 2015). Hydrogen can be utilized as electron donor by the anoxygenic photosynthetic bacteria (Nielsen et al., 2015). Nutritional interdependence among microbial populations is exemplified by an anaerobic community operating from hydrolytic to fermenting primary anaerobes, then to syntrophic bacteria and to homoacetocetic, methanogenic, or sulfidogenic secondary anaerobes. In diverse anoxic environments, spirochetes occupy an intermediate trophic level between the hydrolytic bacteria and these secondary anaerobes; this is because the main compounds produced by spirochete are acetate, H_2, and CO_2, which are normally consumed by sulfate-reducing bacteria and methanogens (Blazejak et al., 2005; Berlanga et al., 2008). In microbial mats, sulfate-reducing bacteria outcompete methanogens because of the high concentration of sulfate in the seawater.

Members of Bacteroidetes (such as *Psycroflexus, Robiginitalea*) were present in all samples. Bacteroidetes are able to grow under a wide range of physicochemical conditions (Farías et al., 2014; Wong et al., 2016) and to degrade polymeric compounds (Fernández-Gómez et al., 2013; Hania et al., 2017). Therefore, Bacteroidetes may play a key role in the degradation and cycling of mat carbon compounds. The family *Rhodothermaceae*, and especially the genus *Salinibacter*, has been detected in abundance in the upper oxic/photic zone in several microbial mats because the respective genera are halophilic and can use light as an additional energy source for growth (Sahl et al., 2008; Schneider et al., 2013). However, in the Camargue microbial mats their relative abundance was low.

Identifying microbes responsible for particular environmental functions is challenging. Microbial mats harbor different microbial symbiont populations with specialized functionalities. In this analysis, we described metabolic potentials and putative interactions among mat community members, leading to an initial overview of the metabolic potential of the entire mat community.

AUTHOR CONTRIBUTIONS

MB designed the work. MB and MP performed the experiments and analysis. All authors discussed and interpreted the results. MB and RG wrote the paper. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by grant CGL2009-08922 (Spanish Ministry of Economy and Competitiveness) to RG.

ACKNOWLEDGMENTS

The authors would like to thank Carlos Llorens, from the Unity of Genomics of Scientific and Technological Centers, University of Barcelona (CCiTUB), who participated in processing and quality control of the reads and clean reads obtained from 454-pyrosequencing, and the JGI staff members who contributed to obtaining the analysis of metagenome data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02619/full#supplementary-material

FIGURE S1 | Relative abundance of families respect on their phyla. **(A)** Relative abundance of Families detected in Bacteroidetes and relative abundance of subphyla from Proteobacteria. **(B)** Relative abundance of several families belonged to Alphaproteobacteria, Deltaproteobacteria and Gammaproteobacteria.

FIGURE S2 | Rarefaction curves from 16S rRNA amplicons from 18 samples. Rarefaction was done at 97% identity, and it was normalized by the number of sequences of the smaller dataset.

TABLE S1 | Key enzymes for functionality metagenomes.

REFERENCES

Al-Najjar, M. A. A., Ramette, A., Kühl, M., Hamza, W., Klatt, J. M., and Polerecky, L. (2014). Spatial patterns and links between microbial community composition and function in cyanobacterial mats. *Front. Microbiol.* 5:406. doi: 10.3389/fmicb.2014.00406

Amend, A. S., Seifert, K. A., and Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: does real abundance count? *Mol. Ecol.* 19, 5555–5565. doi: 10.1111/j.1365-294X.2010.04898.x

Armitage, D. W., Gallagher, K. L., Youngblut, N. D., Buckley, D. H., and Zinder, S. H. (2012). Millimeter-scale patterns of phylogenetic and trait diversity in a salt marsh microbial mat. *Front. Microbiol.* 3:293. doi: 10.3389/fmicb.2012.00293

Berg, I. A. (2011). Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. *Appl. Environ. Microbiol.* 77, 1925–1936. doi: 10.1128/AEM.02473-10

Berlanga, M., Aas, J. A., Paster, B. J., Boumenna, T., Dewhirst, F. E., and Guerrero, R. (2008). Phylogenetic diversity and temporal variation in the *Spirochaeta* populations from two Mediterranean microbial mats. *Int. Microbiol.* 11, 267–274.

Berlanga, M., Llorens, C., Comas, J., and Guerrero, R. (2016). Gut bacterial community of the xylophagous cockroaches *Cryptocercus punctulatus* and *Parasphaeria boleiriana*. *PLOS ONE* 11:e0152400. doi: 10.1371/journal.pone.0152400

Blazejak, A., Erséus, C., Amann, R., and Dubilier, N. (2005). Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm

(Oligochaeta) from the Peru margin. *Appl. Environ. Microbiol.* 71, 1553–1561. doi: 10.1128/AEM.71.3.1553-1561.2005

Bolhuis, H., Cretoiu, M., and Stal, L. J. (2014). Molecular ecology of microbial mats. *FEMS Microbiol. Ecol.* 90, 1–16. doi: 10.1111/1574-6941.12408

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

Cardoso, D. C., Sandionigi, A., Cretoiu, M. S., Casiraghi, M., Stal, L., and Bolhuis, H. (2017). Comparision of the active and resident community of a coastal microbial mat. *Sci. Rep.* 7:2969. doi: 10.1038/s41598-017-03095-z

Christian, N., Handorf, T., and Ebenhoh, O. (2007). Metabolic synergy: increasing biosynthetic capabilities by network cooperation. *Genome Inform. Ser.* 18, 320–329. doi: 10.1142/9781860949920_0031

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

Dillon, J. G., Miller, S., Bebout, B., Hullar, M., Pinel, N., and Stahl, D. A. (2009). Spatial and temporal variability in a stratified hypersaline microbial mat community. *FEMS Microbiol. Ecol.* 68, 46–58. doi: 10.1111/j.1574-6941.2009.00647.x

Dupraz, C., and Visscher, P. T. (2005). Microbial lithification in marine stromatolites and hypersaline mats. *Trends Microbiol.* 13, 429–438. doi: 10.1016/j.tim.2005.07.008

Fan, H., Bolhuis, H., and Stal, L. J. (2015). Nitrification and nitrifying bacteria in a coastal microbial mat. *Front. Microbiol.* 6:1367. doi: 10.3389/fmicb.2015.01367

Farías, M. E., Contreras, M., Rasuk, M. C., Kurth, D., Flores, M. R., Poiré, D. G., et al. (2014). Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: tebenquiche and La Brava, Salar de Atacama, Chile. *Extremophiles* 18, 311–329. doi: 10.1007/s00792-013-0617-6

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832

Feazel, L. M., Spear, J. R., Berger, A. B., Harris, J. K., Frank, D. N., Ley, R. E., et al. (2008). Eucaryotic diversity in a hypersaline microbial mat. *Appl. Environ. Microbiol.* 74, 329–332. doi: 10.1128/AEM.01448-07

Fernandez, A. B., Rasuk, M. C., Visscher, P. T., Contreras, M., Novoa, F., Poire, D. G., et al. (2016). Microbial diversity in sediment ecosystems (evaporites domes microbial mats and crust) of hypersaline laguna Tebenquiche, Salar de Atacama, Chile. *Front. Microbiol.* 7:1284. doi: 10.3389/fmicb.2016.01284

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S. G., González, J. M., et al. (2013). Ecology of marine bacteroidetes: a comparative genomics approach. *ISME J.* 7, 1026–1037. doi: 10.1038/ismej.2012.169

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et al. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Nat. Acad. Sci. U.S.A.* 109, 21390–21395. doi: 10.1073/pnas.1215210110

Fourçans, A., de Oteyza, T. G., Wieland, A., Solé, A., Diestra, E., van Bleijswijk, J., et al. (2004). Characterization of functional bacterial groups in a hypersaline microbial mat community (Salins-de-Giraud. Camargue. France). *FEMS Microbiol. Ecol.* 51, 55–70. doi: 10.1016/j.femsec.2004.07.012

Fourçans, A., Ranchou-Peyruse, A., Caumette, P., and Duran, R. (2008). Molecular analysis of the spatio-temporal distribution of sulfate-reducing bacteria (SRB) in Camargue (France). *Microb. Ecol.* 56, 90–100. doi: 10.1007/s00248-007-9327-x

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Futami, R., Muñoz-Pomer, L., Viu, J. M., Dominguez-Escriba, L., Covelli, L., Bernet, G. P., et al. (2011). GPRO: the professional tool for management, functional analysis and annotation of omic sequences and databases. *Biotechvana Bioinformatics: 2011-SOFT3.*

Großkopf, T., and Soyer, O. S. (2014). Synthetic microbial communities. *Curr. Opin. Microbiol.* 18, 72–77. doi: 10.1016/j.mib.2014.02.002

Guerrero, R., and Berlanga, M. (2006). Life's unity and flexibility: the ecological link. *Int. Microbiol.* 9, 225–235.

Guerrero, R., and Berlanga, M. (2013). An integrative ecogenetic study of minimal ecosystems: the microbial mats of Ebro Delta and the Camargue (Western Mediterranean). *Contrib. Sci.* 9, 117–139. doi: 10.2436/20.7010.01.172

Guerrero, R., and Berlanga, M. (2016). From the cell to the ecosystem: the physiological evolution of simbiosis. *Evol. Biol.* 43, 543–552. doi: 10.1007/s11692-015-9360-5

Hania, W. B., Joseph, M., Bunk, B., Spröer, C., Klenk, H.-P., Fardeau, M.-L., et al. (2017). Characterization of the first cultured representative of Bacteroidetes clade specialized on the scavenging of cyanobacteria. *Environ. Microbiol.* 19, 1134–1148. doi: 10.1111/1462-2920.13639

Harris, J. K., Caporaso, J. G., Walker, J. J., Spear, J. R., Gold, N. J., Robertson, C. E., et al. (2013). Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J.* 7, 50–60. doi: 10.1038/ismej.2012.79

Hirsch, M. D., Long, Z. T., and Song, B. (2011). Anammox bacterial diversity in various aquatic ecosystems based on the detection of hydrazine oxidase genes (*hzoA*/*hzoB*). *Microb. Ecol.* 61, 264–276. doi: 10.1007/s00248-010-9743-1

Hoffmann, D., Maldonado, J., Wojciechowski, M. F., and Garcia-Pichel, F. (2015). Hydrogen export from intertidal cyanobacterial mats: Sources, fluxes and the influence of community composition. *Environ. Microbiol.* 17, 3738–3753. doi: 10.1111/1462-2920.12769

Hügler, M., Menendez, C., Schägger, H., and Fuchs, G. (2002). Malonyl-Coenzyme A reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation. *J. Bacteriol.* 184, 2404–2410. doi: 10.1128/JB.184.9.2404-2410.2002

Hügler, M., Wirsen, C. O., Fuchs, G., Taylor, C. D., and Sievert, S. M. (2005). Evidence for autotrophic CO2 fixation via the reductive tricarboxylic acid cycle by members of the ε subdivision of *Proteobacteria*. *J. Bacteriol.* 187, 3020–3027. doi: 10.1128/JB.187.9.3020-3027.2005

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Tennessen, K., et al. (2016). The standard operation procedure of the DOE-JGI metagenome annotation pipeline (MAP v.4). *Stand. Gen. Sci.* 11:17. doi: 10.1186/s40793-016-0138-x

Klatt, C. G., Inskeep, W. P., Herrgard, M. J., Jay, Z. J., Rusch, D. B., Tringe, S. G., et al. (2013). Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Front. Microbiol.* 4:106. doi: 10.3389/fmicb.2013.00106

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 maker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Lee, J. Z., Burow, L. C., Woebken, D., Everroad, R. C., Kubo, M. D., Spormann, A. M., et al. (2014). Fermentation couples Chloroflexi and sulfate-reducing bacteria to Cyanobacteria in hypersaline microbial mats. *Front. Microbiol.* 5:61. doi: 10.3389/fmicb.2014.00061

Ley, R. E., Harris, J. K., Wilcox, J., Spear, J. R., Miller, S. R., Bebout, B. M., et al. (2006). Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl. Environ. Microbiol.* 72, 3685–3695. doi: 10.1128/AEM.72.5.3685-3695.2006

Long, R. A., Eveillard, D., Franco, S. L. M., Reeves, E., and Pinckney, J. L. (2012). Antagonistic interactions between heterotrophic bacteria as a potential regulator of community structure of hypersaline microbial mats. *FEMS Microbiol. Ecol.* 83, 74–81. doi: 10.1111/j.1574-6941.2012.01457.x

Mukherjee, A., Chettri, B., Langpoklakpam, J. S., Basak, P., Prasad, A., Mukherjee, A. K., et al. (2017). Bioinformatic approaches including predictive metagenomic profiling reveal characteristics of bacterial response to petroleum hydrocarbon contamination in diverse environments. *Sci. Rep.* 7:1108. doi: 10.1038/s41598-017-01126-3

Nielsen, M., Revsbech, N. P., and Kühl, M. (2015). Microsensor measurements of hydrogen gas dynamics in cyanobacterial microbial mats. *Front. Microbiol.* 6:726. doi: 10.3389/fmicb.2015.00726

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., et al. (2014). The genome portal of the department of energy joint genome institute: 2014 updates. *Nucleic Acids Res.* 42, D26–D31. doi: 10.1093/nar/gkt1069

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems* 4:2. doi: 10.1186/1746-1448-4-2

Pade, N., and Hagemann, M. (2015). Salt acclimatation of Cyanobacteria and their application in biotechnology. *Life* 5, 25–49. doi: 10.3390/life5010025

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

Ragsdale, S. W., and Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathways of CO2 fixation. *Biochim. Biophys. Acta* 1784, 1873–1898. doi: 10.1016/j.bbapap.2008.08.012

Ramos, V. M. C., Castelo-Branco, R., Leão, P. N., Martins, J., Carvalhal-Gomes, S., Sobrinho da Silva, F., et al. (2017). Cyanobacterial diversity in microbial mats from the hypersaline lagoon system of Araruama, Brazil: an in-depth polyphasic study. *Front. Microbiol.* 8:1233. doi: 10.3389/fmicb.2017.01233

Robertson, C. E., Spear, J. R., Harris, J. K., and Pace, N. R. (2009). Diversity and stratification of Archaea in a hypersaline microbial mat. *Appl. Environ. Microbiol.* 75, 1801–1810. doi: 10.1128/AEM.01811-08

Ruvindy, R., White, R. A. III., Neilan, B. A., and Burns, B. P. (2016). Unravelling core microbial metabolisms in the hypersaline microbial mats of shark bay using high-throughput metagenomics. *ISME J.* 10, 183–196. doi: 10.1038/ismej.2015.87

Sachs, J. L., and Hollowell, A. C. (2012). The origins of cooperative bacterial communities. *mBio* 3:e00099-12. doi: 10.1128/mBio.00099-12

Saghaï, A., Gutiérrez-Preciado, A., Deschamps, P., Moreira, D., Bertolino, P., Ragon, M., et al. (2017). Unveiling microbial interactions in stratified mat communities from a warm saline shallow pond. *Environ. Microbiol.* 19, 2405–2421. doi: 10.1111/1462-2920.13754

Sahl, J. W., Pace, N. R., and Spear, J. R. (2008). Comparative molecular analysis of endoevaporitic microbial communities. *Appl. Environ. Microbiol.* 74, 6444–6446. doi: 10.1128/AEM.00879-08

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, D., Arp, G., Reimer, A., Reitmer, J., and Daniel, R. (2013). Phylogenetic analysis of a microbialite-forming microbial mat from hypersaline lake of the Kiritimati Atoll, Central Pacific. *PLOS ONE* 8:e66662. doi: 10.1371/journal.pone.0066662

Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., et al. (2017). Lage-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci. Rep.* 7:6589. doi: 10.1038/s41598-017-06665-3

Thiel, V., Hügler, M., Ward, D. M., and Byant, D. A. (2017). The dark side of the mushroom spring microbial mat: Life in the shadow of chlorophototrophs. II. Metabolic functions of abundant community members predicted from metagenomic analyses. *Front. Microbiol.* 8:943. doi: 10.3389/fmicb.2017.00943

Villanueva, L., Navarrete, A., Urmeneta, J., Geyer, R., Guerrero, R., and White, D. C. (2007). Monitoring diel variations of physiological status and bacterial diversity in an estuarine microbial mat: an integrated biomarker analysis. *Microb. Ecol.* 54, 523–531. doi: 10.1007/s00248-007-9224-3

Ward, D. M., Bateson, M. M., Ferris, M. J., Kühl, M., Wieland, A., Koeppel, A., et al. (2006). Cyanobacteria 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. B* 361, 1997–2008. doi: 10.1098/rstb.2006.1919

Warden, J. G., Casaburi, G., Omelon, C. R., Bennett, P. C., Breecker, D. O., and Foster, J. S. (2016). Characterization of microbial mat microbiomes in the modern thrombolite ecosystem of lake Clifton, Western Australia using shotgun metagenomics. *Front. Microbiol.* 7:1064. doi: 10.3389/fmicb.2016.01064

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669–1681. doi: 10.1038/ismej.2015.235

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* 10, 2557–2568. doi: 10.1038/ismej.2016.45

Wieland, A., and Kühl, M. (2006). Regulation of photosynthesis and oxygen consumption in a hypersaline cyanobacterial mat (Camargue, France) by irradiance, temperature and salinity. *FEMS Microb. Ecol.* 55, 195–210. doi: 10.1111/j.1574-6941.2005.00031.x

Wierzchos, J., Berlanga, M., Ascaso, C., and Guerrero, R. (2006). Micromorphological characterization and lithification of microbial mats from the Ebro Delta (Spain). *Int. Microbiol.* 9, 289–295.

Wong, H. L., Ahmed-Cox, A., and Burns, P. B. (2016). Molecular ecology of hypersaline microbial mats: current insights and new directions. *Microorganisms* 4:6. doi: 10.3390/microorganisms4010006

Wong, H. L., Smith, D.-L., Visscher, P. T., and Burns, B. P. (2015). Niche differentiation of bacterial communities at a millimeter scale in Shark Bay microbial mats. *Sci. Rep.* 5:15607. doi: 10.1038/srep15607

Wong, H. L., Visscher, P. T., White, R. A. III, Smith, D.-L., Patterson, M. M., and Burns, B. P. (2017). Dynamics of archaea at fine spatial scales in Shark Bay mat microbiomes. *Sci. Rep.* 7:46160. doi: 10.1038/srep46160

Yang, S., Wen, X., Jin, H., and Wu, Q. (2013). Pyrosequencing investigation into the bacterial community in permafrost soils along the Chin-Russia crude oil pipeline (CRCOP). *PLOS ONE* 7:e52730. doi: 10.1371/journal.pone.0052730

Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., Glöckner, F. O., et al. (2010). Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. *Syst. Appl. Microbiol.* 33, 291–299. doi: 10.1016/j.syapm.2010.08.001

Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.-H., Tu, Q., et al. (2011). Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J.* 5, 1303–1313. doi: 10.1038/ismej.2011.11

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 Berlanga, Palau and Guerrero. 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: 11 August 2017
doi: 10.3389/fmicb.2017.01546

Metagenomic Analysis of Cecal Microbiome Identified Microbiota and Functional Capacities Associated with Feed Efficiency in Landrace Finishing Pigs

Zhen Tan[1], Ting Yang[1], Yuan Wang[1], Kai Xing[1], Fengxia Zhang[1], Xitong Zhao[1], Hong Ao[2], Shaokang Chen[3], Jianfeng Liu[1] and Chuduan Wang[1]**

[1] National Engineering Laboratory for Animal Breeding, MOA Key Laboratory of Animal Genetics and Breeding, Department of Animal Genetics and Breeding, China Agricultural University, Beijing, China, [2] The State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [3] Beijing General Station of Animal Husbandry, Beijing, China

OPEN ACCESS

Edited by:
Sabela Balboa Méndez, Universidade de Santiago de Compostela, Spain

Reviewed by:
Congying Chen, Jiangxi Agricultural University, China
Tatsuya Unno, Jeju National University, South Korea

***Correspondence:**
Jianfeng Liu
liujf@cau.edu.cn
Chuduan Wang
cdwang@cau.edu.cn

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

Received: 15 May 2017
Accepted: 31 July 2017
Published: 11 August 2017

Citation:
Tan Z, Yang T, Wang Y, Xing K, Zhang F, Zhao X, Ao H, Chen S, Liu J and Wang C (2017) Metagenomic Analysis of Cecal Microbiome Identified Microbiota and Functional Capacities Associated with Feed Efficiency in Landrace Finishing Pigs. Front. Microbiol. 8:1546. doi: 10.3389/fmicb.2017.01546

Feed efficiency (FE) appears to vary even within closely related pigs, and may be partly affected by the diversity in the composition and function of gut microbes. To investigate the components and functional differences of gut microbiota of low and high FE pigs, high throughput sequencing and *de novo* metagenomics were performed on pig cecal contents. Pigs were selected in pairs with low and high feed conversion ratio. The microorganisms of individuals with different FE were clustered according to diversity. The genus *Prevotella* was the most enriched in both groups, and the abundance of species *Prevotella* sp. *CAG:604* was significantly increased in low efficiency individuals compared to that in animals showing high efficiency. In contrast, other differential species, including lactic acid bacteria, were all enriched in the group with good feeding characteristics. Functional analysis based on the Kyoto Encyclopedia of Genes and Genomes databases demonstrated that differential genes for the metabolism of carbohydrates were most abundant in both groups, but pathways of pyruvate-related metabolism were more intense in pigs with higher FE. All these data indicated that the microbial environment was closely related to the growth traits of pigs, and regulating microbial composition could aid developing strategies to improve FE for pigs.

Keywords: metagenomics, cecal microbiome, feed efficiency, feed conversion ratio, pigs

INTRODUCTION

Genetics and the environment together shape the performance of domestic animals, and diet plays an extremely important role in production (Aggrey et al., 2010). A major way to reduce costs in the pig industry is to improve feed efficiency (FE), because feed accounts for more than 60% of the production cost (Jing et al., 2015). To enhance pig production, FE should be improved, and this can be measured using residual feed intake (RFI) or feed conversion ratio (FCR). The FCR is the amount of feed consumed per unit of body weight gain during a specified period, and is calculated as the feed intake divided by the weight gained. Thus, an animal with a high FCR is less efficient at converting feed into body mass than one with a low FCR. In previous studies, the heritability values for FCR ranged from 0.13 to 0.31 (Do et al., 2013).

Owing to the rapid development of metagenomic studies, the gut microbiota was identified as markedly influencing animal health and performance. Correlations have been determined between intestinal microbes and host. The scientific community's understanding of intestinal microflora has improved. The gut microbiota improves the energy harvesting capacity (Turnbaugh et al., 2006). Changes in the gut physiology and gut microbial composition were reported to improve FE, together with genetic changes (Lumpkins et al., 2010). The microflora can have a negative impact on the host in various ways, including using excessive amounts of energy, and causing host energy diversion to the immune system by inducing inflammatory responses (Turnbaugh et al., 2006). Variation in the diversity of gut microbes has been associated with differences between different breeds of pigs (Yang et al., 2014), and various growth stages (Kim et al., 2015) or intestinal segments (Kim and Isaacson, 2015), as a result of different genetic and environmental factors. Some nutrients such as resistant starch (RS), which cannot be digested completely in the small intestine, need to be fermented by cecal and colonic microbes to produce short-chain fatty acids (SCFAs) (Knudsen et al., 2012). Feed conversion efficiency is closely related to the genetic diversity of the gut microbiota (Singh et al., 2012, 2014). Therefore, even for animals reared under the same environmental conditions, there is a clear link between the gut microbiota and pig productivity.

The gut microbiota was demonstrated to participate in the regulation of energy harvesting efficiency of the host, which was significantly associated with body weight gain (Looft et al., 2012, 2014; Kim et al., 2015; Kim and Isaacson, 2015). The microorganisms harbored in the gastrointestinal tract of animals have coevolved with their hosts (Ley et al., 2006). The study of community structure and functional capacity of gut microbiota was helpful for better understanding of the relationships between microbial function and host physiology and metabolism. Revealing the taxonomic composition and functional capacity of gut microbiota and their interaction with the host should facilitate understanding the roles they play in the host, and may improve pig production by increasing the component of FE associated with microorganisms. The study on the fecal microbiome in different fatness found that the cecal microbiome has the stronger ability to degrade xylan, pectin and cellulose (Yang et al., 2016) and taxon and functional capacity of fecal microbiota in low and high FCR broilers (Singh et al., 2014), there also has been not many researches on the functional capacity of gut microbial species associated with FE in animals, including pigs.

In this study, high throughput sequencing of metagenomes were undertaken to investigate differences between the microbial communities found in the cecal contents of female finishing Landrace pigs with high and low FE. We determined whether the different structural and functional characteristics of the bacterial populations between the two groups (high and low FE) were correlated with pig production performance. The putative microorganisms identified were associated with nutrient digestion and growth traits. This study may lead to an improved understanding of digestion in the large intestine, while providing novel insights into growth traits.

MATERIALS AND METHODS

Animals and Tissues

In this trial, feed intake and body weight of 120 female Landrace pigs of 120–165 days of age were recorded using a Velos (Nedap Co., Ltd., Groenlo, Netherland) automated individual feeding system, which recorded each pig's intake of feed and weight by recognizing the electronic ear mark during the feed eating per time. The FCR (FCR, feed intake divided by the weight gained) was then obtained. The pigs were provided the same commodity feed and clean water *ad libitum* throughout the experiment, and housed in an environmentally controlled room (10 pigs in each pen), provided by the Tianjin Ninghe Primary Pig Breeding Farm (Ninghe, China). All the experimental pigs were kept healthy and were provided a full range of feed lacking antibiotics and medicines. Pedigree information of all animals was available. The FCR in the high FE group (20 animals) was significantly different from that of the low FE group (20 animals) (**Supplementary Figure S1**). Two full-sib pairs and two half-sib pairs were selected, with each pair having opposite FCR phenotypes (Supplementary Table S1). We defined the low (Lce) group as individuals with low FE and high FCR values, and pigs with high FE and low FCR values were in the high (Hce) group.

The eight selected pigs were euthanized at 166 days of age, and digesta samples were collected from the cecum lumen of each pig within 20 min of euthanization. All samples were collected in sterile tubes and stored in liquid nitrogen as soon as possible until required for further analysis. All the methods were carried out in accordance with the approved guidelines (GB/T 17236–2008) from the Quality Supervision, Inspection, and Quarantine of the People's Republic of China. All experimental protocols were approved by the Animal Welfare Committee of China Agricultural University (permit number: DK996).

DNA Preparation and Sequencing

DNA was extracted and purified from cecal contents using QIAamp DNA Stool Mini Kit (Qiagen Ltd., Germany) following the manufacturer' s instructions. The integrity of DNA was checked using 1% agarose gel electrophoresis, and the concentration of DNA was measured using a UV-Vis spectrophotometer (NanoDrop 2000c, United States).

Metagenomic DNA libraries were constructed with an insert size of 350 base pairs (bp) for each sample following the manufacturer's instructions (Illumina) (Qin et al., 2010). The libraries for metagenomic analysis were sequenced on an Illumina HiSeq 2500 platform by an Illumina HiSeq - PE150 bp strategy.

All the sequencing data were deposited in the National Center for Biotechnology Information Short Read Archive under Accession no. SRP108960.

Data Analysis

Quality Control

Illumina raw reads were treated to remove reads with low qualities, trim the read sequences, remove adaptor, and remove the host sequences. Specifically, reads with adaptor or more than

3 N bases were removed. The 3′ end of the read was trimmed and bases with low quality (Phred score < 20) were removed, as were reads with original length shorter than 50 bp. Host (pig) genomic DNA sequences were removed by SOAPaligner (v 2.21) (Li et al., 2008).

Taxonomy Annotation

The remaining clean reads were assembled by the SOAPdenovo software (v 1.05) (Li et al., 2008), mapping to microbial genomes in NCBI. The parameter is −m 4 −r 2 −m 100 −x 1000 (Qin et al., 2014). Then, the aligned reads were classified at Kingdom, Phylum, Class, Order, Family, Genus, and Species levels, and abundance was determined. The taxonomy profile was constructed at different levels. Principal component analysis (PCA) was used to determine whether there was any similarity between the composition of samples from the same condition. Box plots were generated with the coordinate values of each group, to judge if the first principal component and the second principal component had a significant effect on the sample distribution. Based on the non-parametric Kruskal–Wallis (KW) sum-rank test, LDA Effect Size (LEfSe) analysis was performed at the species level to determine the community or species that had a significant effect on the division of the samples (Segata et al., 2011).

Construction of Gene Set and Differential Gene Analysis

Using SOAPdenovo software (version 1.05), which is based on the De-Bruijn graph (Li et al., 2010), the filtered data were assembled by different sizes of Kmer (51, 55, 59, and 63). Scaffolds were broken into new scaftigs at their gaps. Meanwhile, we estimated the number of scaftigs ≥ 500bp, and selected assembly results with maximum N50 values (Qin et al., 2014).

MetaGeneMark (v 2.10) was employed for prediction of bacterial open reading frames (ORFs) (Noguchi et al., 2006). Gene sequence lengths less than 100 bp were filtered out, and the remaining sequences were translated into the corresponding amino acid sequences.

All predicted genes were clustered (identity > 95%, coverage > 90%) using CD-HIT[1] (Li and Godzik, 2006). Selecting the longest gene sequence in each cluster and removing the redundant genes allowed construction of a non-redundant gene set.

Clean reads were compared to the non-redundant gene set by SOAPaligner (Li et al., 2008) matching the parameters described above. The abundance of each gene in the sample can be calculated based on the number of reads mapped to each gene. The sum of the abundance of each gene in one sample is 1, and the abundance is expressed as the relative abundance. According to the abundance of each gene, the differential genes were screened via the Wilcox rank-sum test, p-value was calculated (Qin et al., 2014).

Functional Annotation Analysis

The set of genes was aligned with the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene database using BLAST (BLAST Version[2] 2.2.28+) to obtain the KO annotation information from the KEGG database (Kanehisa et al., 2004). The differential genes between the groups by rank sum test were aligned with the KEGG database. Thus, the metabolic pathway information of each differential gene was obtained.

Gene sequences were used by BLAST to compare with the Carbohydrate-Active enzymes database (CAZy) to get the information on species source, the functional classification of EC, the gene and protein sequence, and the protein structure of CAZymes (Lombard et al., 2014), and with the Antibiotic Resistance Genes Database (ARDB) to obtain the type, quantity, and function information of antibiotic resistance genes (Liu and Pop, 2009).

RESULTS

Sequencing of the Cecal Microbiota

The DNA of cecal digesta was extracted, fragmented, and sequenced by Illumina HiSeq PE-150, which generated a total of 62 Gbp of clean reads for eight samples, with an average sample size in the high group of 7.4 Gbp, and 8.1 Gbp in the low group. However, only 10 million reads were mapped to microbial genomes (Supplementary Table S2), because about three quarters of the reads originated from eukaryotic DNA, which belongs to feed residue. The microbial DNA could not be separated from eukaryotic DNA of the feed residue easily in experimental conditions, leading to a high proportion of contaminating reads. In addition, there were possibly many reads from microbes that exist in the environment but do not have genomic sequences present in the NCBI database (Hong et al., 2016).

Taxonomic Composition of Cecal Microbiota

The taxonomic composition was calculated using metagenomics methods at the phylum and genus levels (**Supplementary Figures S2**, **S3**). A study by Looft et al. (2014) showed that, at the phylum level, Bacteroidetes was the dominant group in the cecum (Looft et al., 2014), and this result was close to the results from metagenomics analysis in this study. Approximately half of the bacteria were unclassified or accounted for extremely small percentages of all groups at the genus level.

To evaluate the similarity of individuals in a group, and to elucidate whether the microbial composition had a connection with FE, PCA was performed based on the abundance profiling of microbes at the genus and species levels. **Figure 1A** shows that there were clear significant differences between the Hce and Lce groups in the first dimension ($P = 0.029$).

Different Species of Cecal Microbiota between High and Low FE Groups

The PCA analysis revealed that differences in the organismal structure of cecal microbiota were present between the Hce and

[1]http://www.bioinformatics.org/cd-hit/

[2]http://blast.ncbi.nlm.nih.gov/Blast.cgi

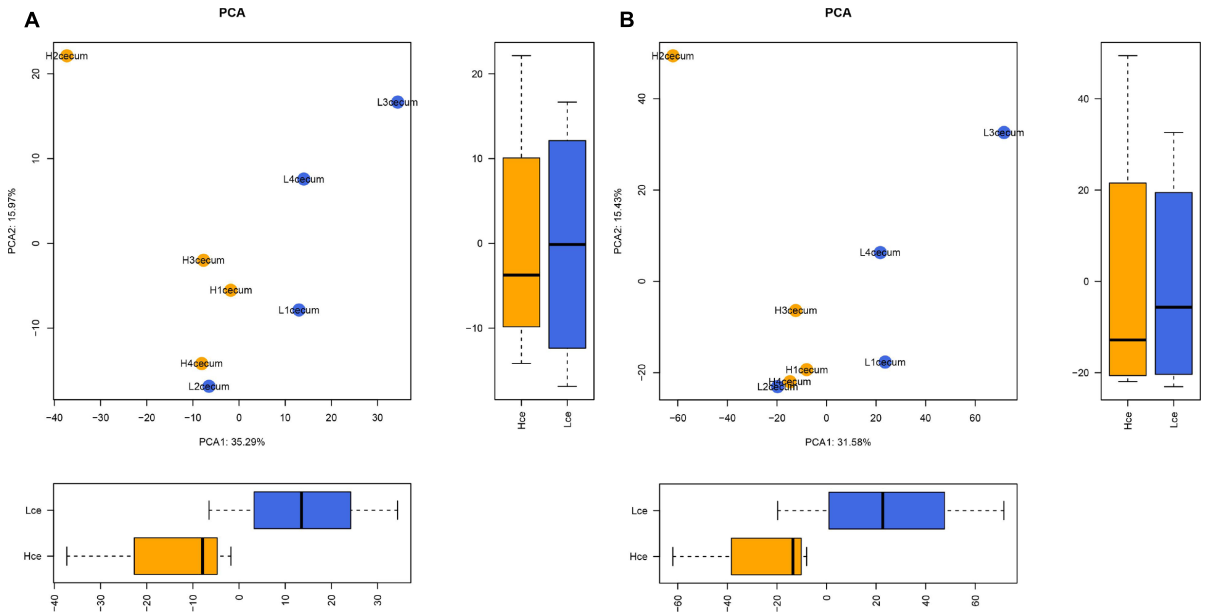


FIGURE 1 | Microbial composition at the genus and species levels differs between cecal digest samples of high and low feed efficiency (FE). **(A)** Principal component analysis (PCA) of metagenomics based genera. **(B)** PCA of metagenomics based species. The boxes were tested by Wilcoxon test in the first dimension and second dimension between groups.

Lce groups (**Figure 1**). Differences in metagenomic taxonomic composition at the species level were identified. Over 300 species exhibited significant differences by the Wilcoxon test, and are shown in Supplementary Table S3. Interestingly, a comparable enrichment analysis of species between the two groups by linear discriminant analysis (LDA) plot revealed that some species were unique biomarkers to cecal microbes of the high or low FE group (**Figure 2**). The LDA Effect Size (LEfSe) based on non-parametric Kruskal–Wallis (KW) sum-rank test results showed only one species, *Prevotella* sp. *CAG:604*, had a significant effect, and could be considered as a potential biomarker for the Lce group. The remainder belonged to the Hce group, and *Oscillibacter* sp_*ER4* had the highest LDA score. *Oscillibacter* is known as a probiotic and producer of anti-inflammatory metabolites (Li et al., 2016). These biomarkers associated with nutrient metabolism and disease prevention showed significantly different abundance, consistent with the better overall health of the Hce group compared with that of the Lce group.

By mapping to the microbial genome, metagenomics species (MGS) were identified in both groups, with 314 species showing significant abundance differences (by Wilcoxon rank-sum tests) (Supplementary Table S3). From taxonomic characterization of the MGS, the top 20 different species showing enrichment are plotted in **Figure 3**. Nineteen of the top twenty species that differed were more enriched in the Hce group. *Prevotella* sp. *CAG:604* was the only single species concentrated more in the low group compared to that in the high group. *Prevotella* sp. *CAG:604* contains some genes that encode proteins involved in nutrient and energy metabolism, such as BN731_01873, ychF, gpmI, queF, speA, fmt, etc. There were three MGS of *Lactobacillus* (*Lactobacillus ruminis*, *L. amylophilus*, and *Lactobacillus* sp. *N54.MSG-719*) in the nineteen species enriched in the Hce group, with *Lactobacillus* spp. often considered as probiotics.

A heatmap plot (**Figure 4**) shows when the major different species were clustered based on species similarity and relative abundance, *Prevotella* sp. *CAG:604* was clustered to a unique branch compared to other species. These results indicated that the species *Prevotella* sp. *CAG:604* might be a potential biomarker for distinguishing between the cecum microbiota of the high and low FE groups.

Comparison of Functionality of the Cecal Microbiome between High and Low FE Groups

The functional classification results could help to clarify the metabolic differences between groups. The different aspects between high and low FCR phenotypes may possibly indicate the microbes that affect nutrient metabolism.

The different predicted genes are listed in Supplementary Table S4, with more than forty thousand (48221) assembled sequences showing significant differences by Wilcox rank-sum test with a P-value < 0.05 (Qin et al., 2014). To elucidate the distribution of samples based on different genes, PCA was performed for the eight samples (**Supplementary Figure S4**). The composition of predicted genes displayed visible differences between the groups.

The different genes were aligned to the CAZy, and categorized into seven CAZy types (**Supplementary Figure S5**) (Lombard et al., 2014). Glycoside hydrolases were enriched most in both high and low groups, and next were the glycosyl transferases.

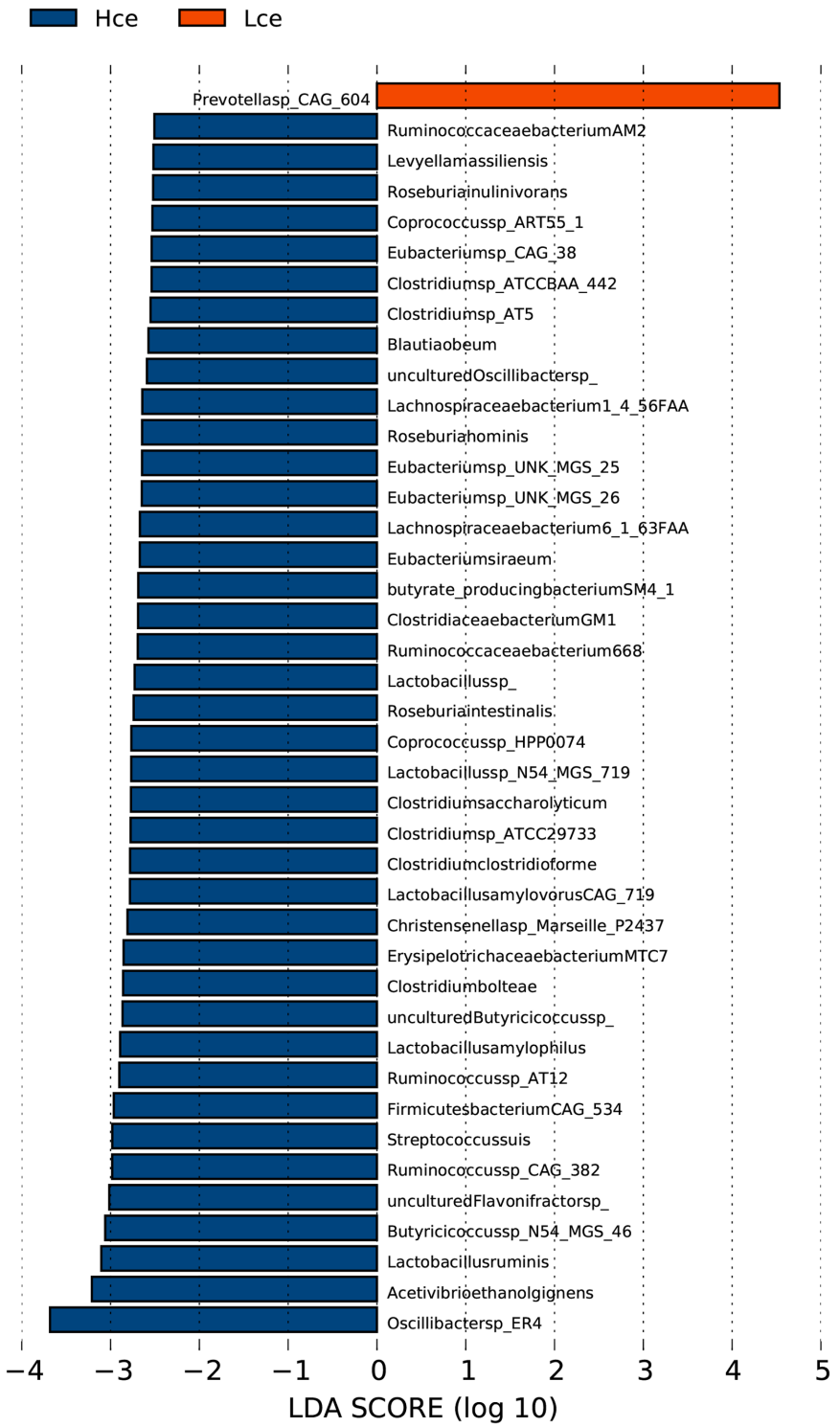


FIGURE 2 | LDA Effect Size (LEfSe) analysis of cecal microbiota between high feed conversion ratio (FCR) and low groups at the species level. LDA (linear discriminant analysis) plot indicates biomarkers found by ranking accordingly to their effect size (2.0) of the species.

Many coexisting microbes compete for nutrients in the cecal fluid, and some microbes may produce antibiotics or toxins to inhibit the growth of others. Analyzing genes by the ARDB can allow annotation of the abundance and the strength of antibiotic resistance genes (Hong et al., 2016). The heatmap of **Supplementary Figure S6** shows that from the abundant AR

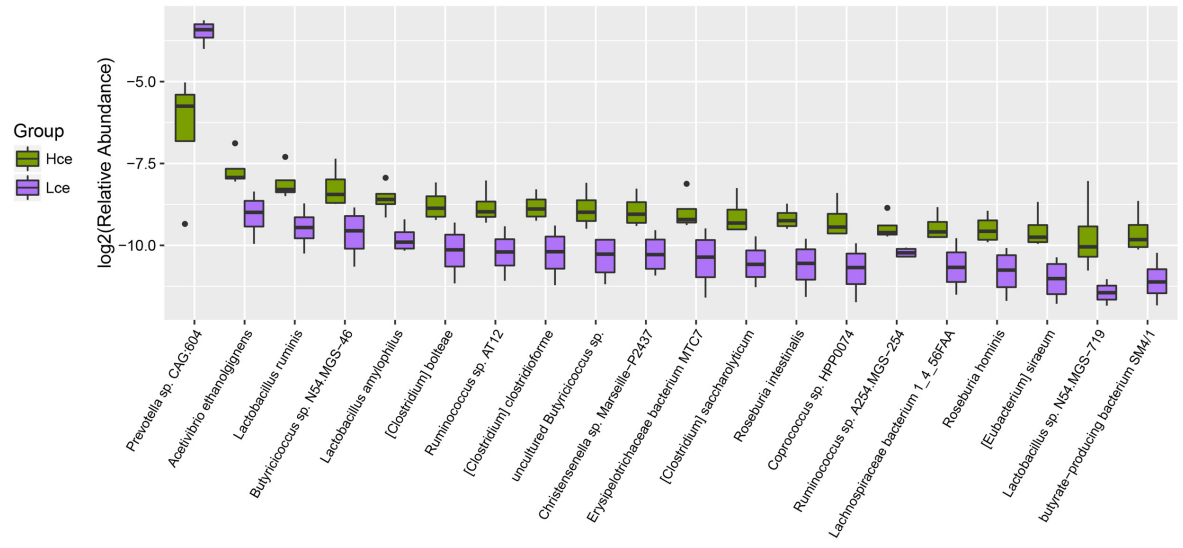


FIGURE 3 | Boxplot for top 20 differentially enriched species between the Hce and Lce groups. Hce, cecal microbiota of high FE group. Lce, cecal microbiota of low FE group.

types after *z*-score processing, the type "MacAB" was the most aligned in both groups. However, the AR types clustered by samples were separated into two major areas. This suggested that the microbe-contributed AR genes were different between the two groups, and regulating the microbial composition might affect the growth of undesirable microbes in response to certain antibiotics.

Differential genes were dominated by the carbohydrate metabolism category in the clustering base subsystem of KEGG pathways, as expected, in **Figure 5**, where the differential genes between the groups by rank sum test were aligned with the KEGG database. The metabolic pathway information of each differential gene was obtained. Amino acid metabolism, energy metabolism, nucleotide metabolism, metabolism of cofactors and vitamins, and transcription were represented in relatively high abundance in both groups.

The differential pathways were then investigated (Supplementary Table S5), and 11 pathways were significantly different with $p < 0.05$. Apart from nitrogen metabolism and other glycan degradation, nine of the eleven pathways showed higher expression in the Hce group: ATP-binding cassette (ABC) transporters, bacterial chemotaxis, D-Arginine and D-ornithine metabolism, flagellar assembly, lysine degradation, phenylalanine metabolism, sulfur relay system, synthesis and degradation of ketone bodies, and two-component system. Bacterial chemotaxis and flagellar assembly are closely related, affected microbiological performance and development. ABC transporters play roles in the import of essential nutrients, the export of toxic molecules, and also mediate the transport of many other physiological substrates (Davidson et al., 2008). The differential pathways categorized as synthesis and degradation of ketone bodies, sulfur relay system, lysine degradation, other glycan degradation, D-Arginine and D-ornithine metabolism, and phenylalanine metabolism were classified to nutrient metabolism. The higher abundance of these pathways indicted that the positive activities of some nutrient metabolism pathways in Landrace pigs may be involved in faster growth.

In addition, some differential genes in the Hce group that were related to ABC transporters, lysine degradation, and phenylalanine metabolism could be mapped to *Desulfovibrio piger*, which is considered to be an intestinal sulfate-reducing bacterium (Kushkevych, 2015) expressing pyruvate-ferredoxin oxidoreductase. *Lactobacillus* was matched to the two-component system in the Hce group, but *Prevotella* matched in the Lce group. *Prevotella* also matched nitrogen metabolism and other glycan degradation, two pathways that were enriched in the Lce group.

As shown in **Figure 6**, the differential genes were mapped to KEGG, the significantly abundant pathways differing between groups were obtained, and the majority of the pathways were interrelated and connected with degradation of nutrients. All the pathway information of **Figure 6** is annotated at the KEGG website. The EC 2.6.1.21 (D-amino-acid transaminase) enriched in Hce is responsible for degradation of amino acids. EC 2.3.1.9 is defined as acetyl-CoA *C*-acetyltransferase, which is involved in many pathways of degradation and metabolism.

DISCUSSION

To reduce false-positive results caused by genetic background noise and the number of replicates (Xing et al., 2016), two full-sib pairs and two half-sib pairs of pigs with opposing FCR phenotypes were used. Previous studies on broilers involving feed conversion efficiency trials have revealed that fecal bacteria are linked to body weight gain (Singh et al., 2014). The compositions and functional annotations of cecal microbiota

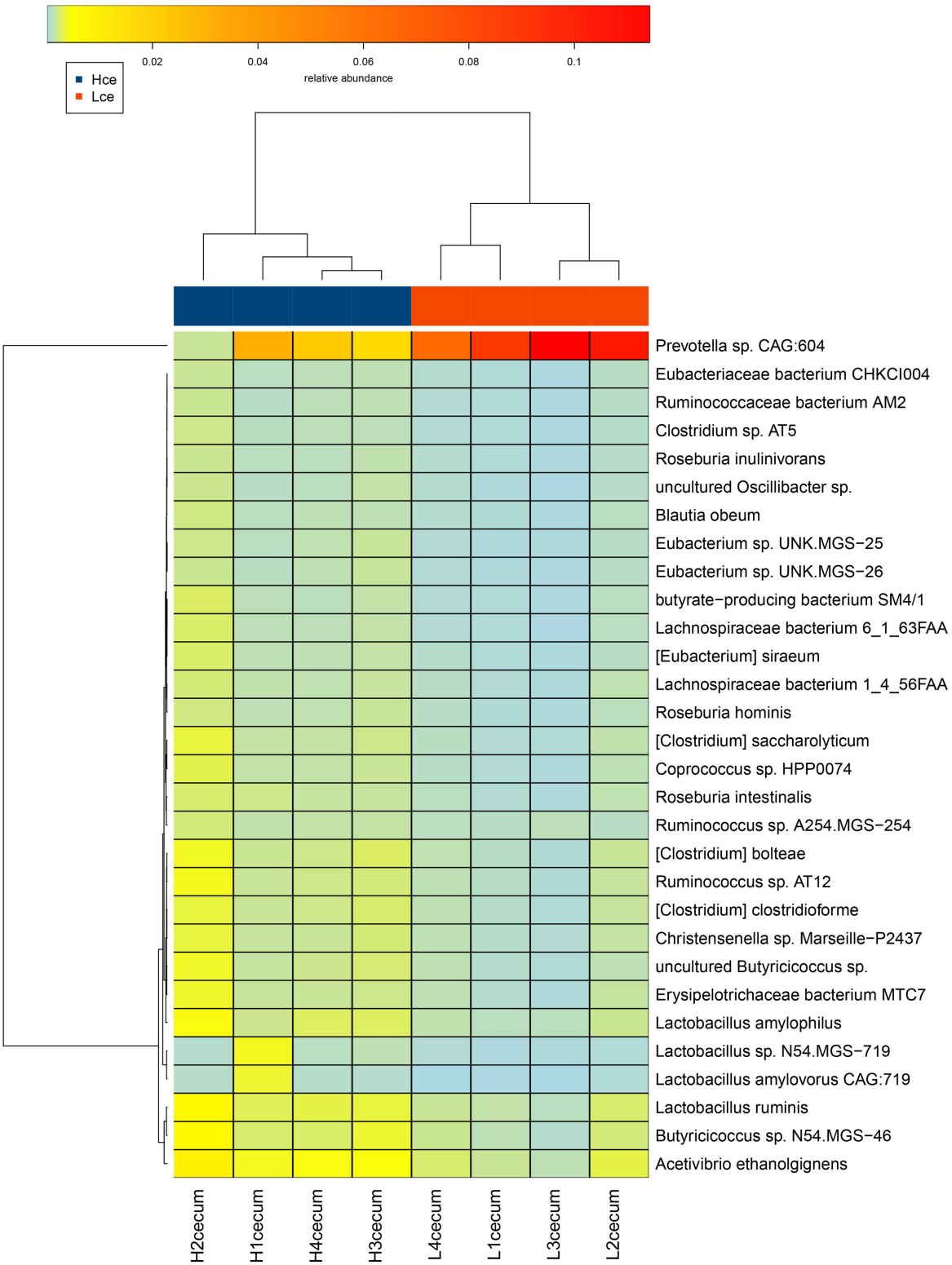


FIGURE 4 | Heatmap cluster analysis of species based on differentially abundant cecal microbiota between the high and low FE groups. The relative levels of abundance are depicted visually from green to red; red represents the highest abundance, whereas green represents the lowest level of abundance. Vertical clustering indicates the similarity in expression between all species in different samples. Horizontal clustering indicates the similarity in abundance of the species of each sample.

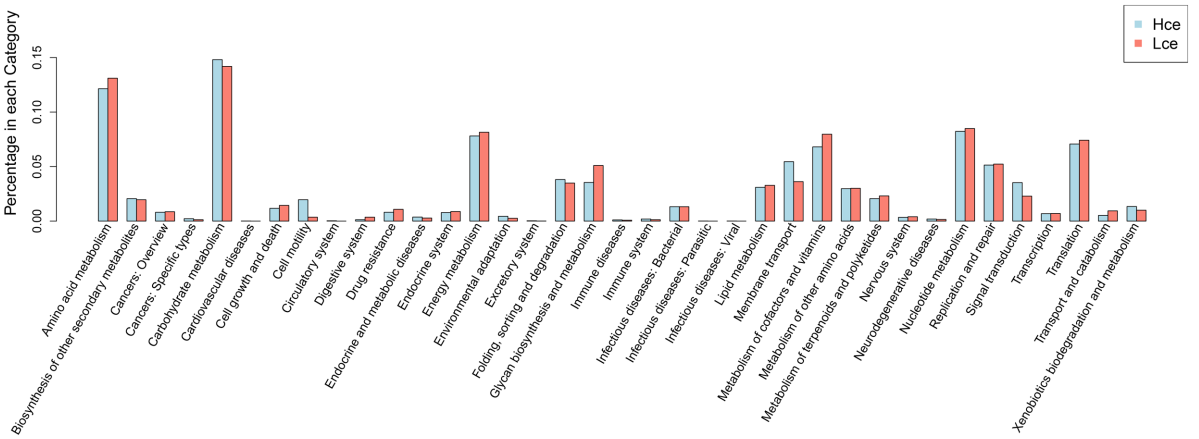


FIGURE 5 | Distribution of category of differential genes by KEGG. KEGG, Kyoto Encyclopedia of Genes and Genomes; Hce, cecal predicted genes of high FE group. Lce, cecal predicted genes of low FE group.

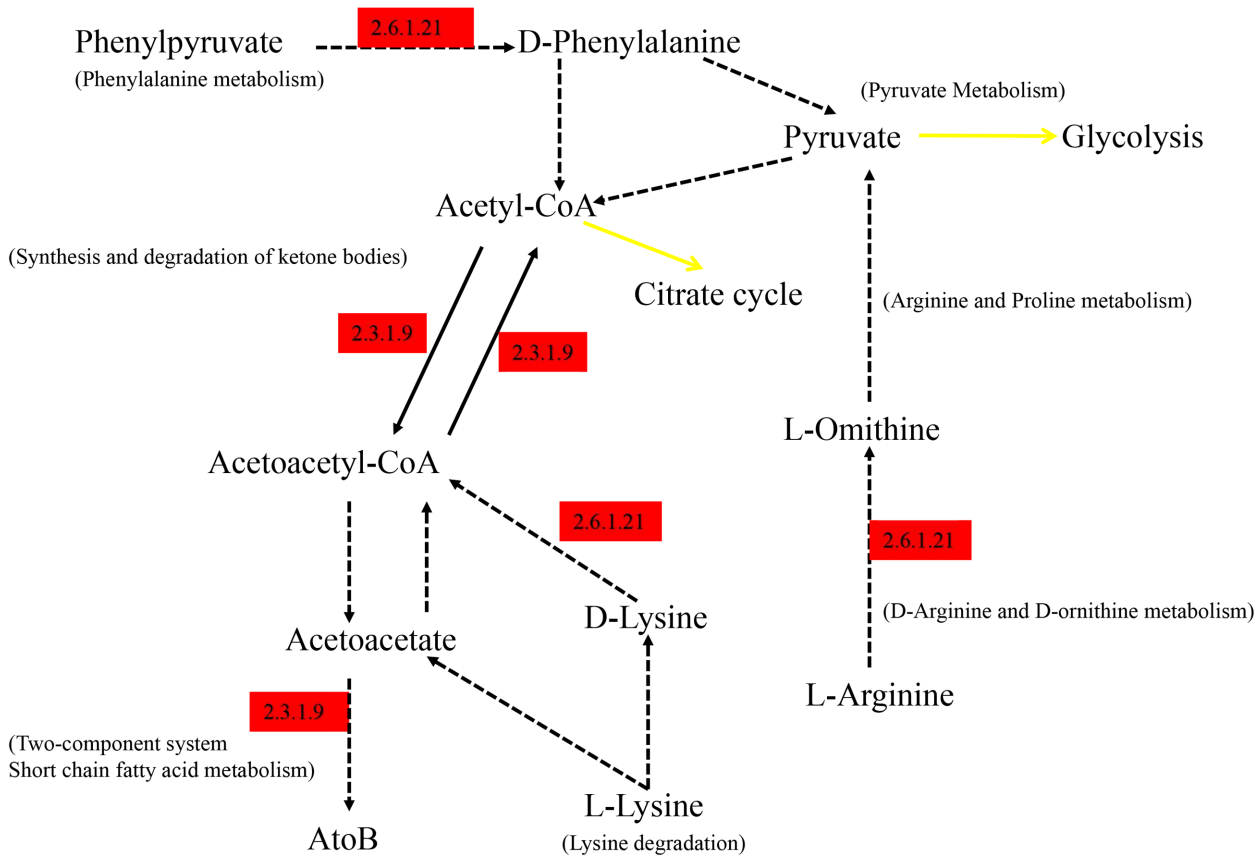


FIGURE 6 | Summarized KEGG pathways of genes significantly abundant in Hce group comparing to Lce. Hce, cecal predicted genes of high FE group. Lce, cecal predicted genes of low FE group. Genes significantly abundant in the Hce are annotated in red; the black solid lines correspond to only one reaction on KEGG pathway; the yellow solid lines indicate a link to another map; the dotted lines indicate more than one reaction.

were characterized for pigs with divergent FE phenotypes during the finishing period here. We searched for the presence of unique genera and/or genes potentially associated with digestion and absorption of nutrients.

The number of mapped reads was a little smaller, but excluding host DNA (average about 2%) and eukaryotic DNA of the feed residue (appropriate 75%), the proportion of mapped reads will be up to 20%, besides that the number of contigs was

normal by the reference (Xiao et al., 2016). While the assembled contig length was 155.7 M ± 40.5 M, and total gene number was 999,791 ± 255,664 in Xiao et al found. The assembled contig length was 143.7 M, and total gene number was 650,604 in average in my study.

Firmicutes and Bacteroidetes were the most abundant phyla in cecal microbiota of piglets in both groups, accounting for more than 80% of the bacterial community, consistent with previous studies (Pedersen et al., 2013; Kim and Isaacson, 2015). The predominant three genera in the cecum of both groups were *Prevotella*, *Bacteroides*, and *Lactobacillus*. The Lce group had a higher proportion of *Prevotella* than the Hce group (**Supplementary Figure S3**). A higher abundance of *Prevotella* is probably related to the presence of fructo-oligosaccharides and starch in the lower intestine (Metzler-Zebeli et al., 2013). *Prevotella* was also considered to be a polysaccharide-degrading bacterial genus, associated with the ability to degrade mucin and plant-based carbohydrates (Lamendella et al., 2011; Patel et al., 2014).

Previous reports showed that *Lactobacillus* was involved in the metabolism of bile acids, which is related to obesity and metabolic syndrome, and the metabolism of phenolic, benzoyl, and phenyl derivatives, which is associated with weight loss (Nicholson et al., 2012). The different abundance of *Lactobacillus* might cause the diversity of energy metabolites in the high and low FE groups. *Lactobacillus*, a member of the lactic acid bacteria (LAB), is commonly used as a probiotic. Some species of *Lactobacillus* help to shape the composition of the gut microbiota by producing antimicrobial bacteriocins (Kim and Isaacson, 2015). Bacteroides fermented various sugar derivatives from plant material and have some benefits for the host, and members of Bacteroides affect the lean or obese phenotype in humans (Ridaura et al., 2013). Several species of *Lactobacillus* are LAB that are used as probiotics, which convert carbohydrates to lactic acid in homofermentation or heterofermentation, or to acetic acid in heterofermentation. The acidified environment produced by these two acids can inhibit the growth of other microbes (Sami et al., 1997).

The other top different MGS were also correlated with conversion and metabolism: *Acetivibrio ethanolgignens* could not degrade cellulose, but can produce ethanol; *Butyricicoccus* sp. *N54.MGS-46* may be a butyrate-producing microbe (Takada et al., 2016); and *Roseburia*, Lachnospiraceae and *butyrate-producing bacterium SM4/1* were also regarded as butyrate-producing bacteria (Duncan et al., 2002; Meehan and Beiko, 2014). Other studies showed that many species belonging to the family Erysipelotrichaceae were enriched in diet-induced obese animals (Turnbaugh et al., 2008). The more vigorous bacteria produced volatile fatty acids and their derivatives in the Hce group compared to that in the Lce group, suggesting that the digestion of the Hce group is more effective than that that of the Lce group and consistent with high FE.

From heatmaps and PCA plots, intra-group samples could be separated from inter-group ones. The *Prevotella* sp. *CAG:604* enriched in the Lce group was clustered away from other species enriched in the Hce group. The differential genes between the groups were analyzed with the KEGG gene database, and metabolic pathway information of differential genes was obtained. Carbohydrate metabolism was the metabolic pathway most enriched in both groups of differential genes. The non-digestible carbohydrates in the small intestine, including cellulose, xylan, and resistant starch, are fermented in the large intestine by microbiota to yield energy for microbial growth, and the end products are volatile fatty acids and their derivatives (Tremaroli and Backhed, 2012).

Carbohydrate metabolism was observed in both groups but involved significantly different genes. CAZymes in assembled different genes were screened for potential complex carbohydrate degradation (**Supplementary Figure S5**). Glycoside hydrolases (GHs) were enriched in both groups, indicating that the majority were related to carbohydrate hydrolysis. However, they were not the same genes in the two groups. Moreover, this result illustrated the pathways of carbohydrate metabolism might be variant between the Hce and Lce groups. From **Supplementary Figure S5**, more genes were concentrated in the Lce group, inconsistent with results from other studies that higher CAZy enzymes were associated with better digestive capacity of pigs (Ramayo-Caldas et al., 2016). When the results were clustered by differences in genes, rather than pathways, they differed.

The predicted genes were specifically annotated using the ARDB. Highly abundant annotations listed in a heatmap (**Supplementary Figure S6**) showed that the individuals of two groups could be grouped into two categories based on antibiotic resistance genes, with the concentrations in different categories. AR gene analysis suggested that diversity existed between individuals with various growth phenotypes. Suppressing the growth of undesirable microbes by regulating microbial composition might be helpful to promote healthy growth of animals (Hong et al., 2016). Health programs could possibly be developed to improve the growth efficiency of pigs by targeting different levels of the specific resistance gene in individuals with different FE.

The KEGG results were consistent with previous reports. The Hce group harbored a relatively high abundance of genes associated with pyruvate- and butyrate-related metabolism (**Figure 6**), part of carbohydrate metabolism, and one of the main sources of SCFAs that were later absorbed by the host (Krishnan et al., 2015). By mapping the differential genes in different pathways to the bacterial genome, *D. piger* was related to ABC transporters, lysine degradation, and phenylalanine metabolism, *Lactobacillus* matched to the two-component system in the Hce group, and *Prevotella* was related to nitrogen metabolism, other glycan degradation and the two-component system in the Lce group. These results were consistent with higher abundance of *Lactobacillus* in the Hce group and *Prevotella* in the Lce group.

Moreover, these differential pathways could be installed in series, and the EC 2.3.1.9 and EC 2.6.1.21 participating in these metabolic pathways were highly expressed in the Hce group. EC 2.3.1.9 (Acetoacetyl-CoA thiolase) catalyzes the cleavage of acetoacetyl-CoA into acetyl-CoA and its reverse reaction (Fox et al., 2014), and EC 2.6.1.21 (D-Amino acid

aminotransferase) catalyzes the inter-conversion between various D-amino acids and alpha-keto acids (Fuchikami et al., 1998). They are responsible for fatty acid degradation and potentially produce medium chain fatty acids for energy supplementation, helping growth of the host.

The pathways of bacterial chemotaxis and flagellar assembly were also increased in the Hce group, suggesting a better growth environment for microorganisms in the Hce group than in the Lce group. These metabolic differences, and especially resulting compounds, possibly influence both microbial composition and growth of the host.

CONCLUSION

In summary, there were differences in the cecal microbiota of individuals with different FE. Microorganisms that differed in abundance were mainly related to carbohydrate metabolism and may affect the growth of the host. The cecum of individuals with high FE contained more potential probiotics such as some species of *Lactobacillus*. Functional analysis revealed that the differentially expressed genes affect the host's energy absorption mainly through the pathway of pyruvate-related metabolism. Taken together, these results indicated that the microbial environment was closely related to the growth traits of pigs, and regulation of microorganism composition could be applied to the pig production industry.

AUTHOR CONTRIBUTIONS

JL, CW, and ZT planned the project and designed the experiments. ZT conducted the experiments and carried out the data analysis with help from TY, KX, and YW. ZT, YW, KX, FZ, XZ, HA, and SC contributed reagents preparation and samples collection. ZT wrote the manuscript, which was critically reviewed by JL and CW.

FUNDING

This research was financially supported by Beijing Innovation Consortium of Agriculture Research System (BAIC02-2016).

ACKNOWLEDGMENTS

We are grateful to the reviewers of this manuscript for their constructive suggestions. The authors are also indebted to Qingbo Wang and Weiliang Zhou from the Tianjin Ninghe primary pig breeding farm for helping with pig experiments, and to the molecular quantitative genetics team at the China Agricultural University for their expertise.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01546/full#supplementary-material

FIGURE S1 | Feed Conversion Ratio (FCR) calculated in high and low groups. Significant difference was tested of 20 individuals at the ends of the high and low FCR, respectively, by one way variance analysis.

FIGURE S2 | Taxonomic composition calculated by different methods at the phylum level based on metagenomics.

FIGURE S3 | Taxonomic composition calculated by different methods at the genus level based on metagenomics.

FIGURE S4 | principal component analysis (PCA) for significant different predicted genes between the Hce and Lce groups. Hce, cecal predicted genes of high FE group. Lce, cecal predicted genes of low FE group.

FIGURE S5 | Number of different abundant genes clustered by CAZy in Hce and Lce groups. CAZy, Carbohydrate-Active Enzymes Database; Hce, cecal predicted genes of high FE group; Lce, cecal predicted genes of low FE group.

FIGURE S6 | Heatmap diagram showing the abundance of antibiotic resistant genes differs between the cecal microbiota of high and low FE groups. FE, feed efficiency. Homogenization control of row by *z*-score.

REFERENCES

Aggrey, S. E., Karnuah, A. B., Sebastian, B., and Anthony, N. B. (2010). Genetic properties of feed efficiency parameters in meat-type chickens. *Genet. Sel. Evol.* 42:25. doi: 10.1186/1297-9686-42-25

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

Do, D. N., Strathe, A. B., Jensen, J., Mark, T., and Kadarmideen, H. N. (2013). Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *J. Anim. Sci.* 91, 4069–4079. doi: 10.2527/jas.2012-6197

Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S., and Flint, H. J. (2002). *Roseburia intestinalis* sp nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* 52, 1615–1620. doi: 10.1099/ijs.0.02143-0

Fox, A. R., Soto, G., Mozzicafreddo, M., Garcia, A. N., Cuccioloni, M., Angeletti, M., et al. (2014). Understanding the function of bacterial and eukaryotic thiolases II by integrating evolutionary and functional approaches. *Gene* 533, 5–10. doi: 10.1016/j.gene.2013.09.096

Fuchikami, Y., Yoshimura, T., Gutierrez, A., Soda, K., and Esaki, N. (1998). Construction and properties of a fragmentary D-amino acid aminotransferase. *J. Biochem.* 124, 905–910.

Hong, X., Chen, J., Liu, L., Wu, H., Tan, H., Xie, G., et al. (2016). Metagenomic sequencing reveals the relationship between microbiota composition and quality of Chinese Rice Wine. *Sci. Rep.* 6:26621. doi: 10.1038/srep26621

Jing, L., Hou, Y., Wu, H., Miao, Y., Li, X., Cao, J., et al. (2015). Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential Residual Feed Intake in pigs. *Sci. Rep.* 5:11953. doi: 10.1038/srep11953

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi: 10.1093/nar/gkh063

Kim, H. B., and Isaacson, R. E. (2015). The pig gut microbial diversity: understanding the pig gut microbial ecology through the next generation high throughput sequencing. *Vet. Microbiol.* 177, 242–251. doi: 10.1016/j.vetmic.2015.03.014

Kim, J., Nguyen, S. G., Guevarra, R. B., Lee, I., and Unno, T. (2015). Analysis of swine fecal microbiota at various growth stages. *Arch. Microbiol.* 197, 753–759. doi: 10.1007/s00203-015-1108-1

Knudsen, K. E. B., Hedemann, M. S., and Laerke, H. N. (2012). The role of carbohydrates in intestinal health of pigs. *Anim. Feed Sci. Technol.* 173, 41–53. doi: 10.1016/j.anifeedsci.2011.12.020

Krishnan, S., Alden, N., and Lee, K. (2015). Pathways and functions of gut microbiota metabolism impacting host physiology. *Curr. Opin. Biotechnol.* 36, 137–145. doi: 10.1016/j.copbio.2015.08.015

Kushkevych, I. V. (2015). Kinetic properties of pyruvate ferredoxin oxidoreductase of intestinal sulfate-reducing bacteria *Desulfovibrio piger* Vib-7 and *Desulfomicrobium* sp Rod-9. *Pol. J. Microbiol.* 64, 107–114.

Lamendella, R., Domingo, J. W., Ghosh, S., Martinson, J., and Oerther, D. B. (2011). Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol.* 11:103. doi: 10.1186/1471-2180-11-103

Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848. doi: 10.1016/j.cell.2006.02.017

Li, J., Sung, C. Y. J., Lee, N., Ni, Y., Pihlajamaki, J., Panagiotou, G., et al. (2016). Probiotics modulated gut microbiota suppresses hepatocellular carcinoma growth in mice. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1306–E1315. doi: 10.1073/pnas.1518189113

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713–714. doi: 10.1093/bioinformatics/btn025

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., et al. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20, 265–272. doi: 10.1101/gr.097261.109

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

Liu, B., and Pop, M. (2009). ARDB-antibiotic resistance genes database. *Nucleic Acids Res.* 37, D443–D447. doi: 10.1093/nar/gkn656

Lombard, V., Golaconda Ramulu, H., 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

Looft, T., Allen, H. K., Cantarel, B. L., Levine, U. Y., Bayles, D. O., Alt, D. P., et al. (2014). Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. *ISME J.* 8, 1566–1576. doi: 10.1038/ismej.2014.12

Looft, T., Johnson, T. A., Allen, H. K., Bayles, D. O., Alt, D. P., Stedtfeld, R. D., et al. (2012). In-feed antibiotic effects on the swine intestinal microbiome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1691–1696. doi: 10.1073/pnas.1120238109

Lumpkins, B. S., Batal, A. B., and Lee, M. D. (2010). Evaluation of the bacterial community and intestinal development of different genetic lines of chickens. *Poult. Sci.* 89, 1614–1621. doi: 10.3382/ps.2010-00747

Meehan, C. J., and Beiko, R. G. (2014). A phylogenomic view of ecological specialization in the lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol. Evol.* 6, 703–713. doi: 10.1093/gbe/evu050

Metzler-Zebeli, B. U., Schmitz-Esser, S., Klevenhusen, F., Podstatzky-Lichtenstein, L., Wagner, M., and Zebeli, Q. (2013). Grain-rich diets differently alter ruminal and colonic abundance of microbial populations and lipopolysaccharide in goats. *Anaerobe* 20, 65–73. doi: 10.1016/j.anaerobe.2013.02.005

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

Noguchi, H., Park, J., and Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* 34, 5623–5630. doi: 10.1093/nar/gkl723

Patel, D. D., Patel, A. K., Parmar, N. R., Shah, T. M., Patel, J. B., Pandya, P. R., et al. (2014). Microbial and Carbohydrate Active Enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet. *Gene* 545, 88–94. doi: 10.1016/j.gene.2014.05.003

Pedersen, R., Ingerslev, H., Sturek, M., Alloosh, M., Cirera, S., Christoffersen, B. O., et al. (2013). Characterisation of gut microbiota in Ossabaw and Gottingen minipigs as models of obesity and metabolic syndrome. *PLoS ONE* 8:e56612. doi: 10.1371/journal.pone.0056612

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

Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513, 59–64. doi: 10.1038/nature13568

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

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, 1241214–1241214. doi: 10.1126/science.1241214

Sami, M., Yamashita, H., Hirono, T., Kadokura, H., Kitamoto, K., Yoda, K., et al. (1997). Hop-resistant *Lactobacillus brevis* contains a novel plasmid harboring a multidrug resistance-like gene. *J. Ferment. Bioeng.* 84, 1–6. doi: 10.1016/S0922-338X(97)82778-X

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

Singh, K. M., Shah, T., Deshpande, S., Jakhesara, S. J., Koringa, P. G., Rank, D. N., et al. (2012). High through put 16S rRNA gene-based pyrosequencing analysis of the fecal microbiota of high FCR and low FCR broiler growers. *Mol. Biol. Rep.* 39, 10595–10602. doi: 10.1007/s11033-012-1947-7

Singh, K. M., Shah, T. M., Reddy, B., Deshpande, S., Rank, D. N., and Joshi, C. G. (2014). Taxonomic and gene-centric metagenomics of the fecal microbiome of low and high feed conversion ratio (FCR) broilers. *J. Appl. Genet.* 55, 145–154. doi: 10.1007/s13353-013-0179-4

Takada, T., Watanabe, K., Makino, H., and Kushiro, A. (2016). Reclassification of *Eubacterium desmolans* as *Butyricicoccus desmolans* comb. nov., and description of *Butyricicoccus faecihominis* sp nov., a butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* 66, 4125–4131. doi: 10.1099/ijsem.0.001323

Tremaroli, V., and Backhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature* 489, 242–249. doi: 10.1038/nature11552

Turnbaugh, P. J., Baeckhed, F., Fulton, L., and Gordon, J. I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* 3, 213–223. doi: 10.1016/j.chom.2008.02.015

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

Xiao, L., Estelle, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., et al. (2016). A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* doi: 10.1038/nmicrobiol.2016.161 [Epub ahead of print].

Xing, K., Zhu, F., Zhai, L., Chen, S., Tan, Z., Sun, Y., et al. (2016). Identification of genes for controlling swine adipose deposition by integrating transcriptome, whole-genome resequencing, and quantitative trait loci data. *Sci. Rep.* 6:23219. doi: 10.1038/srep23219

Yang, H., Huang, X., Fang, S., Xin, W., Huang, L., and Chen, C. (2016). Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness. *Sci. Rep.* 6:27427. doi: 10.1038/srep27427

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.* 27, 898–906. doi: 10.5713/ajas.2013.13621

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 Tan, Yang, Wang, Xing, Zhang, Zhao, Ao, Chen, Liu 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) 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 November 2018
doi: 10.3389/fmicb.2018.02890

Echinococcus granulosus Infection Results in an Increase in *Eisenbergiella* and *Parabacteroides* Genera in the Gut of Mice

Jianling Bao[1,2], Huajun Zheng[3,4], Yuezhu Wang[3], Xueting Zheng[1], Li He[1], Wenjing Qi[1], Tian Wang[1], Baoping Guo[1], Gang Guo[1], Zhaoxia Zhang[1], Wenbao Zhang[1,2], Jun Li[1]* and Donald P. McManus[5]*

[1] State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asian, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China, [2] College of Public Health, Xinjiang Medical University, Urumqi, China, [3] Key Laboratory of Reproduction Regulation of NPFPC, SIPPR, IRD, Fudan University, Shanghai, China, [4] Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China, [5] Molecular Parasitology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

OPEN ACCESS

Edited by:
Jesus L. Romalde,
Universidade de Santiago de Compostela, Spain

Reviewed by:
Catherine Eichwald,
University of Zurich, Switzerland
Majid Fasihi Harandi,
Kerman University of Medical Sciences, Iran

***Correspondence:**
Wenbao Zhang
wenbaozhang2013@163.com
Jun Li
1742712944@163.com;
1742712944@qq.com

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

Received: 16 December 2017
Accepted: 12 November 2018
Published: 29 November 2018

Citation:
Bao J, Zheng H, Wang Y, Zheng X, He L, Qi W, Wang T, Guo B, Guo G, Zhang Z, Zhang W, Li J and McManus DP (2018) Echinococcus granulosus Infection Results in an Increase in Eisenbergiella and Parabacteroides Genera in the Gut of Mice. Front. Microbiol. 9:2890. doi: 10.3389/fmicb.2018.02890

Cystic echinococcosis (CE) is a chronic infectious disease caused by *Echinococcus granulosus*. To confirm whether the infection impacts on the gut microbiota, we established a mouse model of *E. granulosus* infection in this study whereby BALB/c mice were infected with micro-cysts of *E. granulosus*. After 4 months of infection, fecal samples were collected for high-throughput sequencing of the hypervariable regions of the 16S rRNA gene. Sequence analysis revealed a total of 13,353 operational taxonomic units (OTUs) with only 40.6% of the OTUs having genera reference information and 101 of the OTUs were significantly increased in infected mice. Bioinformatics analysis showed that the common core microbiota were not significantly changed at family level. However, two genera (*Eisenbergiella* and *Parabacteroides*) were enriched in the infected mice ($P_{AMOVA} < 0.05$) at genus level. Functional analysis indicated that seven pathways were altered in the *E. granulosus* Infection Group compared with the Uninfected Group. Spearman correlation analysis showed strong correlations of IgG, IgG1 and IgG2a with nine major genera. *E. granulosus* cyst infection may change the gut microbiota which may be associated with metabolic pathways.

Keywords: ***Echinococcus granulosus*, cystic echinococcosis, mice, microbiome, immunoglobins**

INTRODUCTION

Cystic echinococcosis (CE) is a cosmopolitan zoonosis caused by the cystic stage of the dog tapeworm *Echinococcus granulosus* (McManus et al., 2012). The disease causes serious health problems and economic losses, especially in Central Asia (including western China), northern Africa and South America (McManus et al., 2012). *E. granulosus* requires two hosts [an intermediate host including sheep, goats, cattle or wild herbivores and a definitive host such as dogs (or wolves and other carnivores)] to complete its life-cycle. Humans also become infected as an incidental host by ingesting eggs released from *E. granulosus* in carnivore feces. After hatching in the stomach and

intestine, the eggs (oncospheres) penetrate the gut mucosa, enter the blood circulation and finally develop in internal organs mainly the liver (70%) and lungs (20%) of humans. Serological studies showed that in endemic areas, 4–26% of the population were seropositive against *E. granulosus* antigens (Gavidia et al., 2008), indicating that a large number of individuals had been infected. CE infection strongly impacts on host immune responses (Zhang et al., 2003, 2012) with high levels of antibodies, particularly IgG1, IgG4 and IgE (Zhang et al., 2003) and predominant Th2 cytokines including IL-4, IL-5, IL-6, and IL-10 (Rigano et al., 1997; Casaravilla et al., 2014), indicating that the host immune response against CE differs from a bacterial or viral infection (Zeng et al., 2017).

The gut microbiota play an important role in human health (Chen et al., 2017) impacting on metabolism, immunity, development and the behavior of the host (Thaiss et al., 2016). In addition, microbiota components are impacted by medical conditions such as cancer (Jensen et al., 2015; O'Keefe, 2016; Sonnenburg and Backhed, 2016). Similar changes occur in experimental models as well (Gkouskou et al., 2014; Yu et al., 2018). Studies showed that helminth infection in the gut induced typical Th2 immune responses which may control the mcriobiota in the gut of mice (Ramanan et al., 2016; Guernier et al., 2017; Wegener Parfrey et al., 2017). However, it is not known whether *E. granulosus* infection impacts on the gut microbiota of humans or mice. Mice have been used for *E. granulosus* larval infection including primary (Zhang et al., 2001) and secondary infection (Gottstein, 2001; Mourglia-Ettlin et al., 2016). Mice models play an important role in studies of developmental biology and host specificity in echinococcosis (Nakaya et al., 2006). Recently, mouse models were successfully used for drug screening and development (Elissondo et al., 2007; Wang et al., 2017). To increase the success of secondary infection, we developed a method using micro-cysts cultured *in vitro* to infect mice (Zhang et al., 2005), and obtained more than 70% of cyst recovery from 50 PSC-generated cysts.

In this study, BALB/C mice were infected with micro-cysts of *E. granulosus* and their fecal samples were collected for sequencing the variable regions of 16S rRNA genes of gut commensal bacteria to determine their composition and diversity. We show that *E. granulosus* impacted on the gut microbiota of the mice with microbiota changes likely being associated with the altered host immune status in infected individuals.

MATERIALS AND METHODS

Ethics Statement

The protocols for using mice in the study were approved by the Ethics Committee of The First Affiliated Hospital of Xinjiang Medical University (FAH-XMU, Approval No. IACUC-20120625003). The "Guidelines for the Care of Laboratory Animals" by the Ministry of Science and Technology of the People's Republic of China (2006) were rigidly followed in the use of these animals.

Preparation of Cultured *E. granulosus* Hydatid Cysts

Fresh *E. granulosus* sensu stricto protoscoleces (PSC) were aspirated from hydatid cysts from sheep livers collected from a slaughterhouse in Urumqi, Xinjiang Uyghur Autonomous Region, China. The PSC were digested with 1% (w/v) pepsin and cultured to obtain echinococcal cysts using published procedures (Zhang et al., 2005; Wang et al., 2014).

Animal Infection and Sample Collection

Pathogen-free female BALB/c mice, aged 6 weeks, were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd. All animals were housed at the animal facility of the FAH-XMU. The BALB/c mice were randomly divided into two groups: *E. granulosus* infected group (Infected Group) and uninfected group (Uninfected Group). The mice in the *E. granulosus* Infection Group were intraperitoneally (i.p) transplanted with 35 small (diameter, 200–300 μm) hydatid cysts suspended in 0.4 mL RPMI 1640 medium through a 1.0 mL syringe as described in our previous study (Zhang et al., 2005; Wang et al., 2016). Every mouse in the Uninfected Group was i.p. injected with 1.0 mL RPMI 1640 medium. After 4 months, the stool of each mouse was collected daily and about 1 g stool was collected over 1 week. The stool samples were stored at −80°C until use.

Antibody Isotype and Subtype Assays

Mice were sacrificed after the last fecal sample collection, then serum samples were obtained from peripheral blood and stored at −20°C for use. The sera were analyzed by ELISA for different immunoglobulins to hydatid cyst fluid (HCF) including IgG, IgG1, IgG2a, IgG2b and IgG3 and IgM. In brief, each well of MaxiSorb immune-plates (Nunc International) was coated with 100 μL of sheep HCF antigens at a concentration of 2 mg/mL in carbonate/bicarbonate buffer (pH 9.6) (Kittelberger et al., 2002) and incubated overnight at 4°C. After three washes with PBS, the wells were each blocked with 200 μL of 5% (v/v) skim milk in PBS for 1 h at 37°C. Mouse serum was diluted 1–200. For each well, 100 μL of the diluted serum was added and incubated at 37°C for 1 h. After three washes with PBS containing 0.05% Tween 20 (PBST), each well was added 100 μL of diluted anti-mouse monoclonal IgG, IgG1, IgG2a, IgG2b and IgG3 and IgM conjugate (BETHYL). The reaction was developed by adding 100 μL of 2,2-azino-di-[ethyl-benzothiazoline sulfonate] substrate solution (Sigma). After incubation for 30 min in the dark at room temperature, optical density values were read at 405 nm by an ELISA reader (Thermo, Waltham, United States).

Echinococcal Cyst Number and Size

The abdominal cavity of each mouse was opened after sacrifice, and hydatid cysts in the cavity were removed and the numbers of hydatid cysts were counted and their sizes were measured.

PCR Amplification and 16S rRNA Gene Sequencing

Genomic DNA was extracted from stool samples using KAPA HiFi HotStart ReadyMix Kits (Kapa Biosciences, Woburn, MA,

TABLE 1 | The number and average size of cysts in the *E. granulosus* infection group.

	Eg1	Eg2	Eg3	Eg4	Eg5	Eg11	Eg12	Eg13	Eg14	Eg15	Eg31	Eg32	Eg33	Eg35
Number of cysts	12	15	17	37	21	16	17	14	15	11	12	8	23	13
Average size of cysts(cm)	0.58	0.54	0.58	0.72	0.74	0.55	0.67	0.42	0.49	0.71	0.55	0.75	1.01	0.77

Eg, mice infected with E. granulosus.

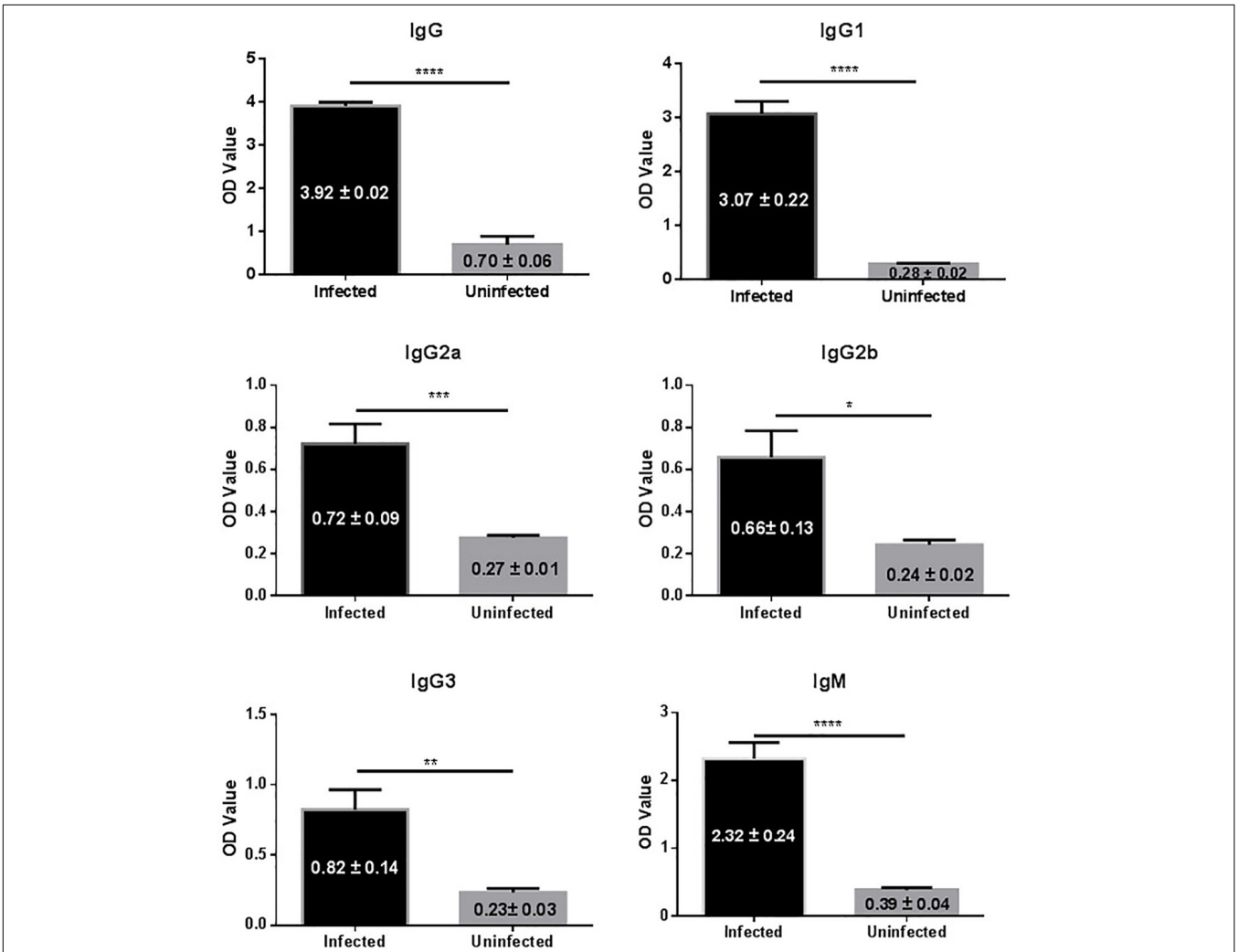


FIGURE 1 | IgG antibody isotypes and subtypes in sera of mice infected with *E. granulosus*. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.005$, $^{****}p < 0.001$.

United States). The V3–V4 hypervariable region of the 16S rRNA gene was amplified by PCR and sequenced; the length of the V3–V4 hypervariable region was approximately 469 bp. Amplicon pools were prepared for sequencing with AMPure XT beads (Beckman Coulter Genomics, Danvers, MA, United States) and quantification with the Library Quantification Kit for Illumina (Kapa Biosciences, Woburn, MA, United States), respectively. The libraries were sequenced on 300PE MiSeq runs.

Bioinformatics and Statistical Analysis

Mothur (version 1.39.5) was used to assemble the paired FASTQ files (Schloss et al., 2011). The selected quality DNA sequences were confirmed using the following criteria: (1) no contaminant sequences, (2) containing no ambiguous bases, (3) the size length ≥350 bp, (4) containing no chimeric sequences, and (5) primers were trimmed. The average length of selected DNA sequences was 414 bp (350–446 bp). The selected DNA sequences were then grouped into operational taxonomic units (OTUs) by comparing with SILVA reference databases (V128) (Quast et al., 2013) at 97% similarity. The minimum reads number of samples (24,097) was used for data normalization. Community richness, evenness and diversity analysis (Shannon, Simpson, Shannoneven, Simpsonenven, ACE, Chao and Good's coverage) were analyzed using the Mothur *T*-test (with 95% confidence

TABLE 2 | The diversity evaluation of the microbiomes of mice infected with cystic echinococcosis and uninfected mice.

				Richness		Evenness		Diversity	
Group	**Sample**	**OTUs**	**Coverage (%)**	**Chao**	**ACE**	**Simpsoneven**	**Shannoneven**	**Shannon**	**Simpson**
Infected	14	9,118	98.50	18,212.718	26,873.315	0.01444	0.646977	5.899138	0.007595
Uninfected	11	8,423	98.25	16,846.368	25,099.202	0.01675	0.660714	5.972012	0.007084

Infected, mice infected with E. granulosus; Uninfected, uninfected mice; OTUs, operational taxonomic units; Chao, Chao index, ACE, ACE index.

intervals, p-value <0.05). Taxonomy was assigned using the online software Ribosomal Database Project (RDP) classifier (80% threshold) (Wang et al., 2007) based on the RDP (Cole et al., 2009). LEfSe (Segata et al., 2011) was also performed to detect abundance taxa (p-value <0.05) between the two groups and estimate linear discriminant analysis effect size (LDA score >2.0). Then Mothur was performed to check the LEfSe results using "metastats" command. Differences among the two groups were also assessed using Analysis of Molecular Variance (AMOVA) in Mothur. Microbiome functions were analyzed using PICRUSt (Langille et al., 2013) based on the KEGG pathways by normalizing the 16S rRNA copy numbers. The input file (biom file) of PICRUSt was calculated using the Mothur software command "classify.otu" and "make.biom", and then the input file was uploaded to the online PICRUSt for function analysis. Differences were determined using STAMP (Parks et al., 2014).

Correlation of Antibody Isotypes and Bacteria

Statistical analysis program-R Package was performed to calculate the coefficient relationship between bacterial genera present and immunoglobulin isotypes using the non-parametric Spearman rank correlation algorithm. A coefficient of >0.68 or <-0.68 was considered to represent strong correlation (Taylor, 1990).

RESULTS

Infection and Blood Serum Isotypes

In this study 14 mice were transplanted with 35 micro-cysts of *E. granulosus*. All the mice were successfully infected with an average number of 16 (SD $\pm$ 7.0) cysts and an average size of 6.4 mm (0.1–21 mm) in diameter (**Table 1**). Serological antibody tests showed that these infected mice had a predominantly IgG1 antibody response against HCF antigens, followed by IgG2b, IgG2a and IgG3 (**Figure 1**), indicating *E. granulosus* infection induced a predominant Th2 response.

Bacterial Populations in the Stool Samples

Stool samples from the 25 mice were collected for gastrointestinal microbiota analysis, including 14 samples from the mice infected with cysts of *E. granulosus* (Infected Group) and 11 control samples from mice without infection (Uninfected Group,). A total of 1,383,569 16S rRNA genes were identified by high-throughput DNA sequencing analysis after filtering through quality control filters. The gene numbers ranged from 24,097 (from one in Uninfected Group) to 86,478 genes. To normalize the data to avoid statistical bias, 24,097 genes from the mice with the lowest gene number were used as a baseline for normalization of all the sequences. OTU (97% similarity) analysis was used to estimate richness, evenness and diversity of the bacterial communities. A total of 13,353 OTUs were obtained including 9,118 OTUs from mice infected with *E. granulosus*, and 8,423 OTUs from the uninfected mice (**Table 2**). No significant difference was evident between the two groups of mice in term of OTU numbers ($p > 0.05$). The Good's coverage was over 93.5% (93.5∼97.8%) for each sample, and over 98% for the two groups, respectively (**Table 2**), meaning that the sequencing depth was sufficient to undertake microbiota analysis with two groups.

Core Microbiome in the Gut of Mice

Ribosomal Database Project analysis showed that 99.7% of the 16S rRNA genes were aligned into nine phyla with the common bacteria *Firmicutes*, *Bacteroidetes* and *Proteobacteria* being dominant in both infected and uninfected groups. RDP analysis clustered 93.5% of the genes (OTUs) into 58 families and 13 families were identified as the major taxa and core microbiomes co-existing in the two groups. The genes in those families accounted for 91.61 and 94.27% of the microbiome community in the infected group and uninfected group, respectively (**Table 3**). Among the 13 families, *Lachnospiraceae* was mostly predominant in both groups, accounting for 41.42 and 43.92% of the total microbiome, respectively. *Ruminococcaceae* and

TABLE 3 | The major families of microbiota in *E. granulosus* infected mice and uninfected mice.

Family	**Uninfected (%)**	**Infected (%)**	***p*-value**
Lachnospiraceae	41.42	43.92	$p > 0.05$
Porphyromonadaceae	14.21	15.75	$p > 0.05$
Ruminococcaceae	11.52	10.87	$p > 0.05$
Rikenellaceae	6.61	5.33	$p > 0.05$
Bacteroidaceae	5.80	5.93	$p > 0.05$
Helicobacteraceae	3.31	3.94	$p > 0.05$
Prevotellaceae	2.32	1.94	$p > 0.05$
Desulfovibrionaceae	1.94	2.30	$p > 0.05$
Coriobacteriaceae	0.14	0.20	$p > 0.05$
Deferribacteraceae	1.82	2.73	$p > 0.05$
Lactobacillaceae	1.80	0.90	$p > 0.05$
Enterobacteriaceae	0.39	0.17	$p > 0.05$
Erysipelotrichaceae	0.33	0.29	$p > 0.05$

Uninfected, uninfected mice; Infected, mice infected with E. granulosus.

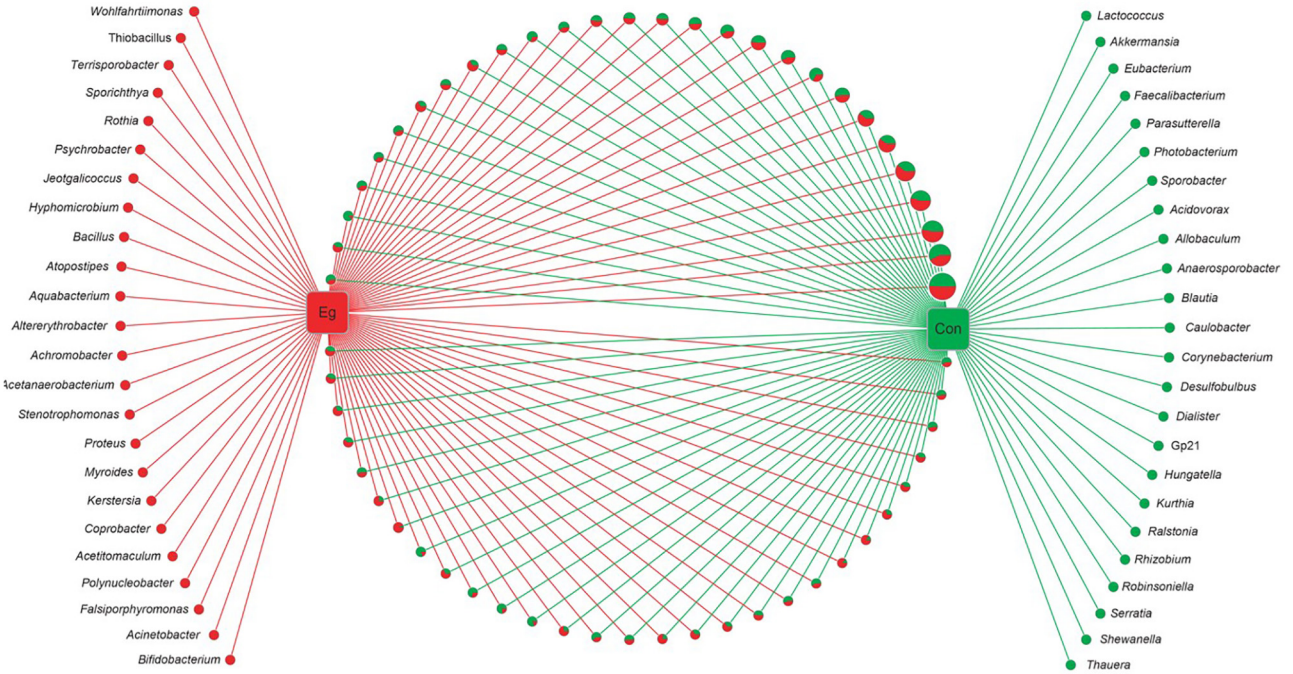


FIGURE 2 | Networks of all bacterial genera revealed in *E. granulosus* uninfected and infected animal groups. Green circle: the unique genera of the Uninfected Group; red circle: the unique genera of Infected Group. In pie charts, green and red represents the genus proportion of the uninfected and infected animal groups; all the pie charts present all the genera common in both groups. Circle sizes represent the read numbers.

Porphyromonadaceae were also dominant (>10% of the entire microbiome in both groups).

Among the 58 families, 40.4% of genes (OUTs) have genus reference information and were aligned into 105 classified genera (**Figure 2**). There were 57 genera co-existing in both groups. However, there were 24 genera present in the Infection Group and another 24 genera in the Uninfected Group. The proportion of all the group unique genera was less than 0.01%, and no significant differences were found between the two groups.

Among the 105 classified genera, 33 were core genera (with each genus comprising >0.1% of total the microbiome), including *Bacteroides*, *Odoribacter*, *Clostridium XlVa*, *Helicobacter*, *Alistipes*, *Barnesiella* and *Mucispirillum* (**Table 4**). Among the predominant genera, there were 27 ubiquitous (core) genera which were consistently found in all samples and comprised more than 38% of the total microbiome.

At the OTU level, there were significant differences between the two groups; 101 OTUs were significantly increased and 49 OTU were significantly decreased in the infected mouse group ($p < 0.05$) (**Table 5**). Of note, most (59.6%) OTUs were unclassified into genera as no classification information is available for these OTUs.

Bacterial Composition in Different Groups

LEfSe analysis showed the composition of the bacterial populations in the guts of the infected and uninfected mice was similar, whereas richness, evenness and diversity were only slightly changed (**Table 2** and **Figure 3**). In contrast, AMOVA analysis showed significant difference between the two groups for the microbiota ($P_{AMOVA} = 0.029$). Species richness (OTU, ACE and Chao index) was higher in the *E. granulosus* Infected Group, and the evenness (Shannoneven and Simpsoneven) was lower in this group compared with the uninfected mice. As richness and evenness combined, there was no significant difference in the diversity between the two groups ($p > 0.5$). Consequently, the results of these analyses indicated that *E. granulosus* infection did not alter much of the composition of the core bacteria present in the mouse gut significantly, although some rare bacteria in very low abundance were increased.

Among the 13 core families, LEfSe analysis showed no significant difference between the *E. granulosus* Infected Group and Uninfected Group in terms of microbiome. Among the major abundant genera, three showed significant differences between the groups (**Table 4**, LDA > 2, $p < 0.05$). The infected mice significantly increases two genera, including Eisenbergiella (1.9 times) and Parabacteroides (17.5 times) compared with the uninfected mice ($p < 0.05$).

Predicted Functional Potential Changes in the Microbiomes of the *E. granulosus* Infection and Uninfected Groups

We used PICRUSt to predict and compare the microbial functional potential changes between the two groups. A total of 47 Kos were found to be significantly increased in the Infected

TABLE 4 | The abundance of genera (>0.1%) in the gut microbiomes of *E. granulosus* infected mice and uninfected mice determined by LEfSe analysis.

Order	Family	Genus	Feature	Uninfected (%)	Infected (%)	*p*-value
Coriobacteriales	Coriobacteriaceae	Enterorhabdus	Ubiquitous	0.06	0.08	
Bacteroidales	Bacteroidaceae	Bacteroides	Ubiquitous	5.80	5.93	
Bacteroidales	Porphyromonadaceae	Barnesiella	Ubiquitous	2.06	2.99	
Bacteroidales	Porphyromonadaceae	Odoribacter	Ubiquitous	3.86	4.33	
Bacteroidales	Porphyromonadaceae	Parabacteroides	Ubiquitous	0.29*	0.56*	0.023976
Bacteroidales	Prevotellaceae	Alloprevotella	Ubiquitous	1.55	1.13	
Bacteroidales	Prevotellaceae	Prevotella	Ubiquitous	0.73	0.80	
Bacteroidales	Rikenellaceae	Rikenella	Ubiquitous	1.81	1.67	
Deferribacterales	Deferribacteraceae	Mucispirillum	Ubiquitous	1.81	2.72	
Lactobacillales	Lactobacillaceae	Lactobacillus	Ubiquitous	1.80	0.90	
Clostridiales	Clostridiaceae 1	Clostridium sensu stricto		0.12	0.17	
Clostridiales	Lachnospiraceae	Acetatifactor	Ubiquitous	1.10	1.01	
Clostridiales	Lachnospiraceae	Clostridium XlVa	Ubiquitous	2.70	4.19	
Clostridiales	Lachnospiraceae	Clostridium XlVb	Ubiquitous	0.40	0.28	
Clostridiales	Lachnospiraceae	Dorea	Ubiquitous	0.10	0.07	
Clostridiales	Lachnospiraceae	Eisenbergiella		0.02*	0.35*	0.001998
Clostridiales	Lachnospiraceae	Lachnospiraceae incertae_sedis	Ubiquitous	0.56	0.61	
Clostridiales	Lachnospiraceae	Roseburia	Ubiquitous	0.06	0.11	
Clostridiales	Ruminococcaceae	Anaerotruncus	Ubiquitous	0.31	0.31	
Clostridiales	Ruminococcaceae	Butyricicoccus	Ubiquitous	0.35	0.35	
Clostridiales	Ruminococcaceae	Clostridium IV	Ubiquitous	0.04	0.05	
Clostridiales	Ruminococcaceae	Flavonifractor	Ubiquitous	1.62	1.78	
Clostridiales	Ruminococcaceae	Intestinimonas	Ubiquitous	0.15*	0.11*	0.035964
Clostridiales	Ruminococcaceae	Oscillibacter	Ubiquitous	1.69	1.29	
Clostridiales	Ruminococcaceae	Pseudoflavonifractor	Ubiquitous	0.68	0.73	
Clostridiales	Ruminococcaceae	Ruminococcus		0.13	0.02	
Erysipelotrichales	Erysipelotrichaceae	Erysipelotrichaceae incertae_sedis	Ubiquitous	0.21	0.14	
Erysipelotrichales	Erysipelotrichaceae	Turicibacter		0.03	0.10	
Desulfovibrionales	Desulfovibrionaceae	Bilophila		0.03	0.03	
Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio	Ubiquitous	0.34	0.67	
Campylobacterales	Helicobacteraceae	Helicobacter	Ubiquitous	3.31	3.94	
Enterobacteriales	Enterobacteriaceae	Escherichia/Shigella	Ubiquitous	0.39	0.17	
Anaeroplasmatales	Anaeroplasmataceae	Anaeroplasma		0.16	0.10	

*Uninfected, uninfected mice; Infected, mice infected with E. granulosus. *significantly different between uninfected mice and infected mice; ubiquitous, the genus was identified as present in all samples.*

Group ($p < 0.05$) (**Table 6**). At KEGG level 3, 7 pathways were identified as being significant difference ($p < 0.05$) between the Infected Group and Uninfected Group, and the changed pathways belonged to the metabolism category, including "Biotin metabolism," "Biosynthesis and Biodegradation of secondary metabolites," "Ether lipid metabolism," "Steroid biosynthesis," "Aminobenzoate degradation," "Tryptophan metabolism" and "Limonene and pinene degradation."

Correlations Between Bacterial Composition and Immunoglobulin Isotypes

Spearman correlation analysis showed strong correlations of IgG, IgG1 and IgG2a with nine major genera (**Table 7** and **Figure 4**). The numbers of *Enterorhabdus*, *Barnesiella* and *Clostridium* XlVa were positively correlated with IgG1, IgG2a and IgG2b levels, respectively. IgA was positively associated with increased numbers of genera *Clostridium* IV, *Lachnospiraceae Incertae sedis* and *Mucispirillum*. In addition, IgG, IgG1, IgG2b and IgG3 were associated with decreased numbers of genera *Escherichia/Shigella*, *Ruminococcus*, *Ruminococcus/Intestinimonas* and *Ruminococcus*, respectively (**Table 7**).

DISCUSSION

The ecological balance of the microbiota in the gut is crucial for maintaining healthy condition (Cani et al., 2008). Disruption of the balance of the gut microbiota is associated with a range of diseases, including colorectal cancer, autoimmune diseases, metabolic diseases, among others (Sokol et al., 2008; Jiang et al., 2015). In this study, we showed that *E. granulosus* infection increased two genera of gut microbiota: *Eisenbergiella* and genus

TABLE 5 | The abundance of OTU ($p < 0.01$) in the infected mice and uninfected mice determined by LEfSe analysis.

#OTU	Taxonomy	Uninfected: mean rel.freq.(%)	Infected: mean rel.freq.(%)	*p*-values
OTU00452	Clostridiales;Lachnospiraceae;Clostridium XlVa	0.0139587	0.04387	0.000351
OTU00048	Desulfovibrionales;Desulfovibrionaceae;unclassified_Desulfovibrionaceae	0.0067907	0.934319	0.000472
OTU00409	Clostridiales;Lachnospiraceae;Acetatifactor	0.0033954	0.040906	0.001051
OTU00832	Desulfovibrionales;Desulfovibrionaceae;unclassified_Desulfovibrionaceae	0	0.013043	0.001199
OTU01146	Clostridiales;Lachnospiraceae;unclassified_Lachnospiraceae	0.0075453	0.000296	0.001367
OTU01119	Clostridiales;Lachnospiraceae;unclassified_Lachnospiraceae	0.0003773	0.005336	0.001834
OTU00050	Bacteroidales;Porphyromonadaceae;Parabacteroides	0.1011065	0.450856	0.002202
OTU00186	Clostridia;Clostridiales;unclassified_Clostridiales	0.1844817	0.046538	0.00282
OTU00122	Clostridiales;Lachnospiraceae;Eisenbergiella	0.0211267	0.335252	0.002849
OTU00410	Clostridiales;Ruminococcaceae;Oscillibacter	0.0045272	0.034681	0.003348
OTU00058	Campylobacterales;Helicobacteraceae;Helicobacter	0.0580985	0.596103	0.003788
OTU00720	Clostridia;Clostridiales;unclassified_Clostridiales	0.0015091	0.01245	0.00456
OTU01984	Desulfovibrionales;Desulfovibrionaceae;unclassified_Desulfovibrionaceae	0	0.002668	0.006591
OTU01307	Clostridiales;Lachnospiraceae;unclassified_Lachnospiraceae	0.0003773	0.004743	0.007635
OTU01051	Bacteria;Firmicutes;unclassified_Firmicutes	0.0003773	0.007411	0.007741
OTU02591	Desulfovibrionales;Desulfovibrionaceae;unclassified_Desulfovibrionaceae	0	0.001779	0.008089
OTU01588	Bacteroidales;Porphyromonadaceae;Odoribacter	0.0033954	0.000296	0.008438
OTU00715	Desulfovibrionales;Desulfovibrionaceae;unclassified_Desulfovibrionaceae	0.0120724	0.003261	0.008709
OTU00460	Clostridiales;Lachnospiraceae;unclassified_Lachnospiraceae	0.0049044	0.049206	0.009747

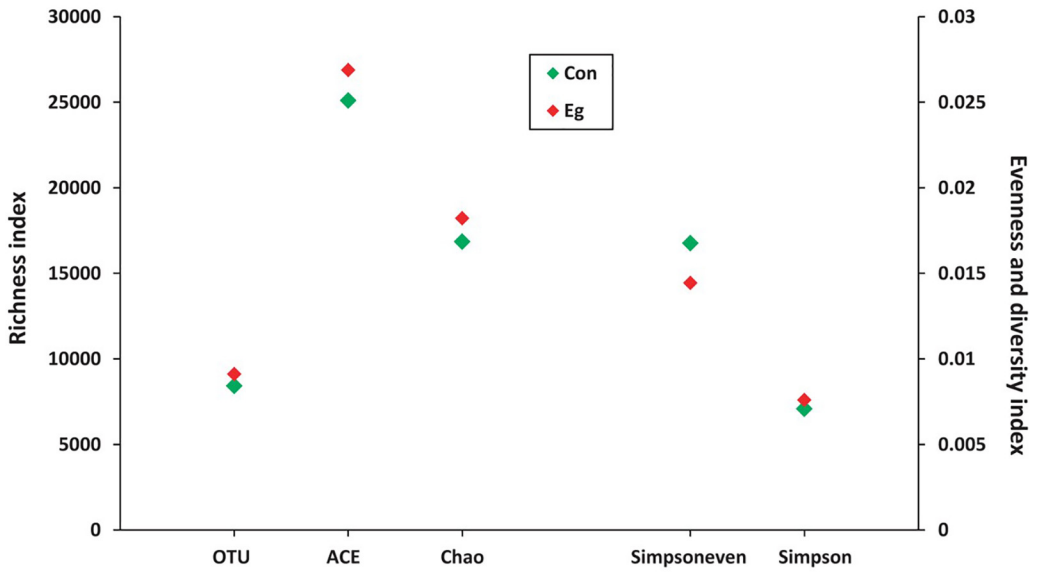


FIGURE 3 | The microbial diversity variability in mice infected (Infected Group) or uninfected (Uninfected Group) with *E. granulosus*. LEfSe analysis showed the composition of the bacterial populations in the guts of the infected and uninfected mice was similar, whereas richness, evenness and diversity were no changed.

Parabacteroides, with most genera remaining unchange. The two genera are in the family *Lachnospiraceae*. Their increase may impact on human health (Plieskatt et al., 2013), and may be associated with diabetes in mice (Kameyama and Itoh, 2014).

At the OTU level, there were 150 OTUs significantly changed in the infected mice. However, among the OTUs, only 49 OTUs have taxonomic information at genus level with 101 without predicted taxonomic classification information, which limited our further analysis (**Table 5**). Additionally, the LEfSe analysis also showed genera *Eisenbergiella* and *Parabacteroides* increased in the infected mice, suggesting that these two genera of bacteria might be biomarkers for *E. granulosus* infection. In our study, the genus *Eisenbergiella* in the family *Lachnospiraceae* was up-regulated significantly in the infected mice, however, there is very limited biological function information on this genus. Combined with antibody analysis in this study, this genus of bacteria may be associated with a Th2 response.

Another increased genus in Infected Group is *Parabacteroides* whereas species belonging to the genus *Parabacteroides* are saccharolytic (Rajilic-Stojanovic and de Vos, 2014), being

TABLE 6 | Functional predictions using the PICRUSt base on 16S rRNA gene copy numbers.

KO accession	Annotation	Infected	Uninfected	*p*-value
K00655	1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]	17,607	15,826	0.01887
K02312	2,3-dihydroxybenzoate-AMP ligase [EC:2.7.7.58]	4,214	3,074	2.43E−02
K01826	5-carboxymethyl-2-hydroxymuconate isomerase [EC:5.3.3.10]	2	0	0.015531
K01905	Acetyl-CoA synthetase (ADP-forming) [EC:6.2.1.13]	4,216	3,063	0.02436
K12554	Alanine adding enzyme [EC:2.3.2.-]	1	0	0.020538
K05341	Amylosucrase [EC:2.4.1.4]	4,223	3,068	0.023681
K03325	Arsenite transporter, ACR3 family	5,022	3,773	1.16E−02
K01338	ATP-dependent Lon protease [EC:3.4.21.53]	15,636	13,406	0.023016
K01305	Beta-aspartyl-dipeptidase (metallo-type) [EC:3.4.19.-]	4,245	3,082	0.023369
K01012	Biotin synthetase [EC:2.8.1.6]	15,209	12,941	0.013677
K03519	Carbon-monoxide dehydrogenase medium subunit [EC:1.2.99.2]	4,219	3,067	0.024608
K01647	Citrate synthase [EC:2.3.3.1]	14,507	12,412	0.023345
K01739	Cystathionine gamma-synthase [EC:2.5.1.48]	4,250	3,105	0.023501
K00901	Diacylglycerol kinase [EC:2.7.1.107]	5,333	4,032	0.012241
K11754	Dihydrofolate synthase/folylpolyglutamate synthase [EC:6.3.2.12 6.3.2.17]	26,899	24,234	0.011904
K07458	DNA mismatch endonuclease, patch repair protein [EC: 3.1.-.-]	4,883	3,692	0.01095
K06212	Formate transporter	4,268	3,116	0.021367
K13892	Glutathione transport system ATP-binding protein	4,226	3,072	0.024579
K04653	Hydrogenase expression/formation protein HypC	7,620	6,473	0.017468
K04654	Hydrogenase expression/formation protein HypD	7,624	6,475	0.016744
K04655	Hydrogenase expression/formation protein HypE	7,696	6,567	0.019062
K04656	Hydrogenase maturation protein HypF	7,630	6,479	0.016758
K04652	Hydrogenase nickel incorporation protein HypB	7,888	6,740	0.021211
K09384	Hypothetical protein	4,225	3,076	2.37E−02
K09703	Hypothetical protein	4,217	3,064	0.024347
K07301	Inner membrane protein	8,454	7,324	0.000592
K03779	L(+)-tartrate dehydratase alpha subunit [EC:4.2.1.32]	4,264	3,106	0.024966
K00879	L-fuculokinase [EC:2.7.1.51]	4,246	3,099	0.023788
K08369	MFS transporter, putative metabolite:H+ symporter	4,280	3,112	0.021188
K02018	Molybdate transport system permease protein	10,849	8,353	2.09E−02
K03637	Molybdenum cofactor biosynthesis protein C	5,564	4,311	0.024072
K07474	Phage terminase small subunit	4,318	3,190	0.022446
K02759	PTS system, cellobiose-specific IIA component [EC:2.7.1.69]	4,336	3,207	0.022898
K02760	PTS system, cellobiose-specific IIB component [EC:2.7.1.69]	4,448	3,372	0.019282
K02777	PTS system, glucose-specific IIA component [EC:2.7.1.69]	4,677	3,516	0.016117
K10026	Queuosine biosynthesis protein QueE	4,760	3,668	2.47E−02
K12996	Rhamnosyltransferase [EC:2.4.1.-]	4,219	3,080	0.024267
K03086	RNA polymerase primary sigma factor	17,918	16,008	3.41E−03
K05297	Rubredoxin-NAD+ reductase [EC:1.18.1.1]	4,352	3,216	0.020307
K03438	S-adenosyl-methyltransferase [EC:2.1.1.-]; 16S rRNA (cytosine1402-N4)-methyltransferase [EC:2.1.1.199]	16,924	15,121	0.014162
K07313	Serine/threonine protein phosphatase 1 [EC:3.1.3.16]	5,067	3,888	0.010963
K02945	Small subunit ribosomal protein S1	16,311	14,363	1.68E−03
K05814	sn-glycerol 3-phosphate transport system permease protein	4,248	3,118	0.023484
K11928	Sodium/proline symporter	4,419	3,257	0.022096
K01695	Tryptophan synthase alpha chain [EC:4.2.1.20]	9,531	8,500	0.021878
K07665	Two-component system, OmpR family, copper resistance phosphate regulon response regulator CusR	7,719	6,661	0.012516
K04784	Yersiniabactin non-ribosomal peptide synthetase	4,212	3,062	0.024426
Average		2,156	2,116	

The numbers of Infected and Uninfected were the normalization of copy numbers. Uninfected, uninfected mice; Infected, mice infected with E. granulosus.

producers of short chain fatty acids (SCFAs) including acetate, propionate and butyrate as bacterial fermentation products (Cummings et al., 1987; Correa-Oliveira et al., 2016; Lloyd-Price et al., 2016). SCFAs act as links between the microbiota and the host immune system (Correa-Oliveira et al., 2016). The liver is the major systemic organ for SCFA metabolism and consumption

TABLE 7 | Genera of bacteria correlating highly with serum antibody isotypes in *E. granulosus* infected mice and uninfected mice.

Genus	Uninfected (%)	Infected (%)	Factor	Uninfected	Infected	Spearman Coef	*p*-value
Barnesiella	2.06	2.99	IgG2a	0.2758	0.8682	0.9	0.037386
Ruminococcus	0.13	0.02	IgG1	0.2801	3.4034	−0.9	0.037386
Ruminococcus	0.13	0.02	IgG2b	0.2434	0.8148	−0.9	0.037386
Ruminococcus	0.13	0.02	IgG3	0.2329	1.0469	−1	0
Ruminococcus	0.13	0.02	IgM	0.3879	2.7058	−0.9	0.037386
Clostridium IV	0.04	0.05	IgA	0.1563	0.3758	0.9747	0.004818
Enterorhabdus	0.06	0.08	IgG1	0.2801	3.4034	0.9	0.037386
Intestinimonas	0.15	0.11	IgG2b	0.2434	0.8148	−0.9	0.037386
Escherichia/Shigella	0.39	0.17	IgG	0.7014	3.9271	−1	0
Lachnospiracea Incertaesedis	0.56	0.61	IgA	0.1564	0.3758	0.9747	0.004818
Mucispirillum	1.81	2.72	IgA	0.1564	0.3758	0.9747	0.004818
Clostridium XlVa	2.70	4.19	IgG2b	0.2434	0.8148	0.9	0.037386

Uninfected, uninfected mice; Infected, mice infected with E. granulosus.

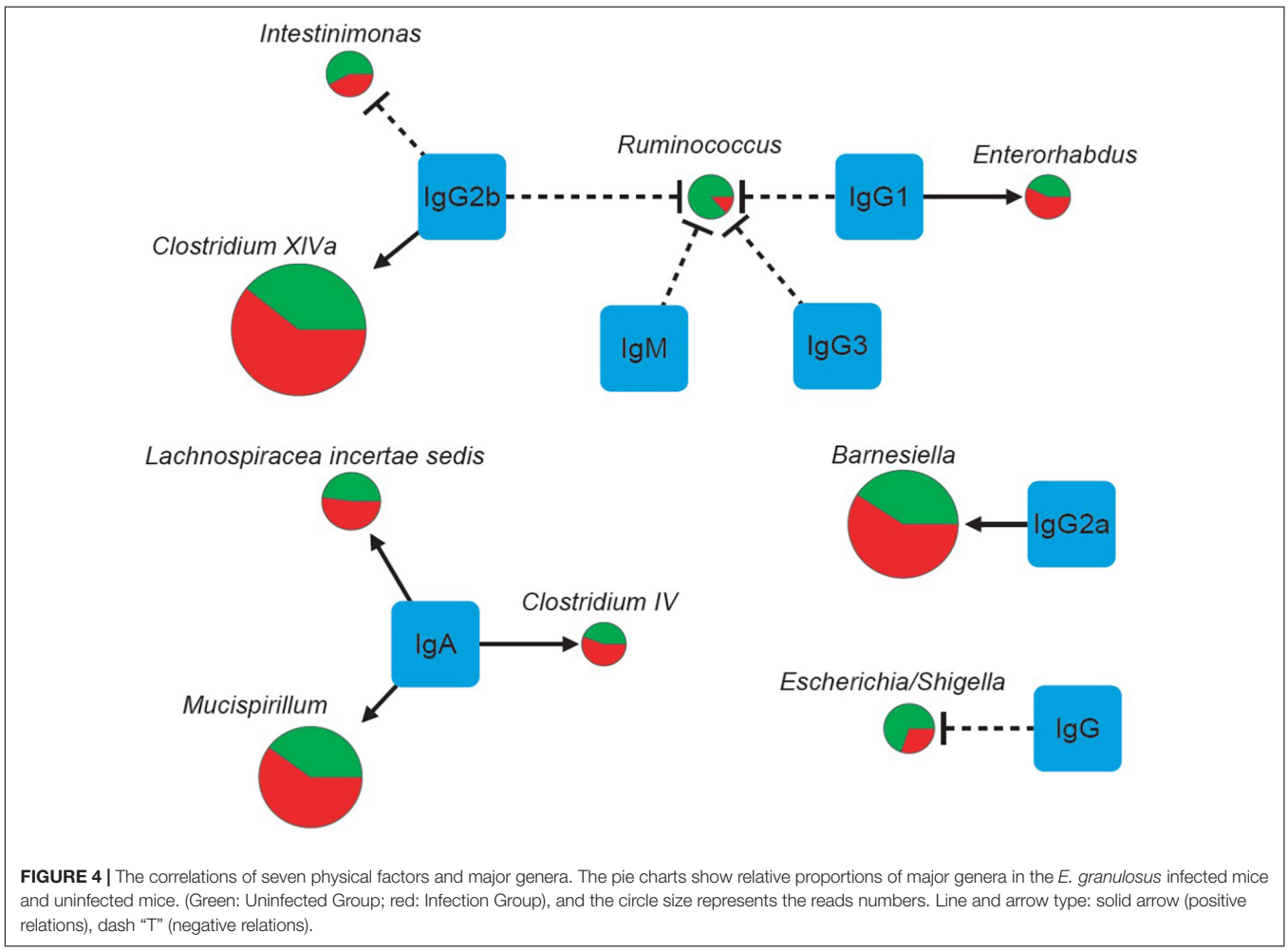


FIGURE 4 | The correlations of seven physical factors and major genera. The pie charts show relative proportions of major genera in the *E. granulosus* infected mice and uninfected mice. (Green: Uninfected Group; red: Infection Group), and the circle size represents the reads numbers. Line and arrow type: solid arrow (positive relations), dash "T" (negative relations).

(Kim et al., 2014), SCFAs released by the gut and equaled by hepatic uptake (Bloemen et al., 2009). *Parabacteroides* has evolved to contain a gene encoding a major capsid protein (Rosenwald et al., 2014) one of the phage orthologous groups (Kristensen et al., 2013). One report demonstrated that *Parabacteroides* was prevalent in diabetic (Wu et al., 2010). The increasing of genus *Parabacteroides* in hydatid infection may be associated with hepatic alteration.

Functional predictions showed seven pathways of the gut microbiota in the *E. granulosus* infection group were altered compared with the uninfected group. These pathways included biotin, lipid metabolism, and tryptophan metabolism.

The synergistic effect of bacteria leads to the difference of gut flora metabolic pathways due to some or all intestinal bacteria involving in metabolism. Biotin metabolism in the intestine is regulated through transcriptional and post-transcriptional mechanisms. Its balance plays a key role in regulating the absorption and the function of biotin in tissues (Zoetendal et al., 2012). Based on pathway impact analysis, we found that tryptophan metabolism was decreased in the *E. granulosus* infection group. In mice infected with *schistosomes*, tryptophan or compounds from tryptophan metabolism were up-regulated and increased in urine which indicate possible problems in tryptophan metabolism in these infected animals (Njagi et al., 1992; Wang et al., 2004).

We showed that the bacterial composition of nine major genera had strong correlations with the levels of IgG, IgG1 and IgG2a antibodies against HCF antigens (**Figure 4**). The numbers of genera *Enterorhabdus*, and *Clostridium* XlVa were positively correlated with IgG1 and IgG2b levels, indicating that these bacteria can be tolerated with those Th2 associated antibodies or Th2 responses may benefit those genera of bacteria. Meanwhile, IgG2a, a Th1 associated antibody, was associated with increased number of genus *Barnesiella*, indicating Th1 has a role for increasing genus *Barnesiella*. Our data also showed that Th2 associated antibodies IgG1 and IgG2b and IgG3 decreased numbers of genera *Escherichia/Shigella*, and *Ruminococcus*, *Ruminococcus/Intestinimonas* (**Table 7** and **Figure 4**). Interestingly, *Intestinimonas* decreased significantly in *E. granulosus* infection group as a differential genus by LEfSe analysis, perhaps it is associated with some kinds of change, then we concluded that it is highly related to IgG2b by Spearman coefficient correlation analysis and is consistent with the immune background of *E. granulosus* infection. So IgG2b may play impartment role in inhibition of *Intestinimonas*. *Clostridium* has been found to be associated with a number of diseases. It showed that *Clostridium* may participate in antibiotic-associated diarrhea (Buffie et al., 2015) and damages the human intestine *in vitro* (Fernandez Miyakawa et al., 2005). *Barnesiella* is present in the healthy intestinal tract and is influenced by antibiotics, and intestinal colonization with *Barnesiella* confers resistance to intestinal domination and bloodstream infection with vancomycin-resistant *Enterococcus* (Ubeda et al., 2013).

In summary, we explored gut microbiota in mice infected with *E. granulosus*, and found that chronic *E. granulosus* infection increased 101 OTUs including two genera of gut microbiota in mice. Functional prediction showed seven pathways of gut microbiota were altered, and bacterial composition of major genera had positive correlations with IgG1 and IgG2b in *E. granulosus* infected mice.

Whereas more than 85% of the genomic sequences between mouse and *Homo* are conserved, overall gene expression and its regulation are considerably different between the two species (Hugenholtz and de Vos, 2018). Human and mouse seems to be similar at phylum level, Bacteroidetes and Firmicutes are the two major bacterial phyla of the intestinal tract (Rawls et al., 2006). However, we do not know whether *E. granulosus* infection will affect human intestinal flora in the same way as the mouse and further studies are now required to understand the further possible mechanisms associated with altered colonization resistance after helminth infection and to determine changes in the gut microbiota of patients with CE.

ACCESSION NUMBERS

The sequence data have been submitted to the GeneBank Sequence Read Archive (Accession Number PRJNA396089).

AUTHOR CONTRIBUTIONS

WZ, JL, and DM contributed to conception and design of the study. JB organized the database. LH, WQ, and TW finished the animal experiments. HZ, ZZ, and YW performed the statistical analysis. JB and YW wrote the first draft of the manuscript. GG, XZ, and BG wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

The study was financially supported by The National Natural Science Foundation of China (NSFC 81460308 and NSFC U1303203).

REFERENCES

Bloemen, J. G., Venema, K., Van De Poll, M. C., Olde Damink, S. W., Buurman, W. A., and Dejong, C. H. (2009). Short chain fatty acids exchange across the gut and liver in humans measured at surgery. *Clin. Nutr.* 28, 657–661. doi: 10.1016/j.clnu.2009.05.011

Buffie, C. G., Bucci, V., Stein, R. R., Mckenney, P. T., Ling, L., Gobourne, A., et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 517, 205–208. doi: 10.1038/nature13828

Cani, P. D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A. M., Delzenne, N. M., et al. (2008). Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes Metab. Res. Rev.* 57, 1470–1481. doi: 10.2337/db07-1403

Casaravilla, C., Pittini, A., Ruckerl, D., Seoane, P. I., Jenkins, S. J., Macdonald, A. S., et al. (2014). Unconventional maturation of dendritic cells induced by particles from the laminated layer of larval *Echinococcus granulosus*. *Infect. Immun.* 82, 3164–3176. doi: 10.1128/IAI.01959-14

Chen, Y. C., Greenbaum, J., Shen, H., and Deng, H. W. (2017). Association between gut microbiota and bone health: potential mechanisms and prospective. *J. Clin. Endocrinol. Metab.* 102, 3635–3646. doi: 10.1210/jc.2017-00513

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37, D141–D145. doi: 10.1093/nar/gkn879

Correa-Oliveira, R., Fachi, J. L., Vieira, A., Sato, F. T., and Vinolo, M. A. (2016). Regulation of immune cell function by short-chain fatty acids. *Clin. Transl. Immunol.* 5:e73. doi: 10.1038/cti.2016.17

Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P., and Macfarlane, G. T. (1987). Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* 28, 1221–1227. doi: 10.1136/gut.28.10.1221

Elissondo, M., Ceballos, L., Dopchiz, M., Andresiuk, V., Alvarez, L., Bruni, S. S., et al. (2007). In vitro and in vivo effects of flubendazole on *Echinococcus granulosus* metacestodes. *Parasitol. Res.* 100, 1003–1009. doi: 10.1007/s00436-006-0381-y

Fernandez Miyakawa, M. E., Pistone Creydt, V., Uzal, F. A., Mcclane, B. A., and Ibarra, C. (2005). Clostridium perfringens enterotoxin damages the human intestine in vitro. *Infect. Immun.* 73, 8407–8410. doi: 10.1128/IAI.73.12.8407-8410.2005

Gavidia, C. M., Gonzalez, A. E., Zhang, W., Mcmanus, D. P., Lopera, L., Ninaquispe, B., et al. (2008). Diagnosis of cystic echinococcosis, central Peruvian Highlands. *Emerg. Infect. Dis.* 14, 260–266. doi: 10.3201/eid1402.061101

Gkouskou, K. K., Deligianni, C., Tsatsanis, C., and Eliopoulos, A. G. (2014). The gut microbiota in mouse models of inflammatory bowel disease. *Front. Cell. Infect. Microbiol.* 4:28. doi: 10.3389/fcimb.2014.00028

Gottstein, B. (2001). Animal models in parasitology. *Contrib. Microbiol.* 9, 31–44. doi: 10.1159/000060387

Guernier, V., Brennan, B., Yakob, L., Milinovich, G., Clements, A. C., and Soares Magalhaes, R. J. (2017). Gut microbiota disturbance during helminth infection: can it affect cognition and behaviour of children? *BMC Infect. Dis.* 17:58. doi: 10.1186/s12879-016-2146-2

Hugenholtz, F., and de Vos, W. M. (2018). Mouse models for human intestinal microbiota research: a critical evaluation. *Cell. Mol. Life Sci.* 75, 149–160. doi: 10.1007/s00018-017-2693-8

Jensen, A. B., Ajslev, T. A., Brunak, S., and Sorensen, T. I. (2015). Long-term risk of cardiovascular and cerebrovascular disease after removal of the colonic microbiota by colectomy: a cohort study based on the Danish National Patient Register from 1996 to 2014. *BMJ Open* 5:e008702. doi: 10.1136/bmjopen-2015-008702

Jiang, H., Ling, Z., Zhang, Y., Mao, H., Ma, Z., Yin, Y., et al. (2015). Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav. Immun.* 48, 186–194. doi: 10.1016/j.bbi.2015.03.016

Kameyama, K., and Itoh, K. (2014). Intestinal colonization by a Lachnospiraceae bacterium contributes to the development of diabetes in obese mice. *Microbes Environ.* 29, 427–430. doi: 10.1264/jsme2.ME14054

Kim, C. H., Park, J., and Kim, M. (2014). Gut microbiota-derived short-chain fatty acids, T cells, and inflammation. *Immune Netw.* 14, 277–288. doi: 10.4110/in.2014.14.6.277

Kittelberger, R., Reichel, M. P., Jenner, J., Heath, D. D., Lightowlers, M. W., Moro, P., et al. (2002). Evaluation of three enzyme-linked immunosorbent assays (ELISAs) for the detection of serum antibodies in sheep infected with *Echinococcus granulosus*. *Vet. Parasitol.* 110, 57–76. doi: 10.1016/S0304-4017(02)00308-4

Kristensen, D. M., Waller, A. S., Yamada, T., Bork, P., Mushegian, A. R., and Koonin, E. V. (2013). Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J. Bacteriol.* 195, 941–950. doi: 10.1128/JB.01801-12

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

Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome. *Genome Med.* 8:51. doi: 10.1186/s13073-016-0307-y

McManus, D. P., Gray, D. J., Zhang, W., and Yang, Y. (2012). Diagnosis, treatment, and management of echinococcosis. *BMJ* 344:e3866. doi: 10.1136/bmj.e3866

Mourglia-Ettlin, G., Cucher, M., Arbildi, P., Rosenzvit, M., and Dematteis, S. (2016). Natural and induced antibodies contribute to differential susceptibility to secondary cystic echinococcosis of Balb/c and C57Bl/6 mice. *Immunobiology* 221, 103–115. doi: 10.1016/j.imbio.2015.07.016

Nakaya, K., Mamuti, W., Xiao, N., Sato, M. O., Wandra, T., Nakao, M., et al. (2006). Usefulness of severe combined immunodeficiency (scid) and inbred mice for studies of cysticercosis and echinococcosis. *Parasitol. Int.* 55(Suppl.), S91–S97. doi: 10.1016/j.parint.2005.11.014

Njagi, E. N., Bender, D. A., and Okelo, G. B. (1992). Tryptophan metabolism and vitamin B6 nutritional status in patients with schistosomiasis mansoni and in infected mice. *Parasitology* 104(Pt 3), 433–441. doi: 10.1017/S0031182000063691

O'Keefe, S. J. (2016). Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* 13, 691–706. doi: 10.1038/nrgastro.2016.165

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

Plieskatt, J. L., Deenonpoe, R., Mulvenna, J. P., Krause, L., Sripa, B., Bethony, J. M., et al. (2013). Infection with the carcinogenic liver fluke *Opisthorchis viverrini* modifies intestinal and biliary microbiome. *FASEB J.* 27, 4572–4584. doi: 10.1096/fj.13-232751

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

Rajilic-Stojanovic, M., and de Vos, W. M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* 38, 996–1047. doi: 10.1111/1574-6976.12075

Ramanan, D., Bowcutt, R., Lee, S. C., Tang, M. S., Kurtz, Z. D., Ding, Y., et al. (2016). Helminth infection promotes colonization resistance via type 2 immunity. *Science* 352, 608–612. doi: 10.1126/science.aaf3229

Rawls, J. F., Mahowald, M. A., Ley, R. E., and Gordon, J. I. (2006). Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* 127, 423–433. doi: 10.1016/j.cell.2006.08.043

Rigano, R., Profumo, E., and Siracusano, A. (1997). New perspectives in the immunology of *Echinococcus granulosus* infection. *Parassitologia* 39, 275–277.

Rosenwald, A. G., Murray, B., Toth, T., Madupu, R., Kyrillos, A., and Arora, G. (2014). Evidence for horizontal gene transfer between *Chlamydophila pneumoniae* and Chlamydia phage. *Bacteriophage* 4:e965076. doi: 10.4161/21597073.2014.965076

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

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J. J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16731–16736. doi: 10.1073/pnas.0804812105

Sonnenburg, J. L., and Backhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. *Nature* 535, 56–64. doi: 10.1038/nature18846

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Sonogr.* 6, 35–39. doi: 10.1177/875647939000600106

Thaiss, C. A., Zmora, N., Levy, M., and Elinav, E. (2016). The microbiome and innate immunity. *Nature* 535, 65–74. doi: 10.1038/nature18847

Ubeda, C., Bucci, V., Caballero, S., Djukovic, A., Toussaint, N. C., Equinda, M., et al. (2013). Intestinal microbiota containing Barnesiella species cures vancomycin-resistant *Enterococcus faecium* colonization. *Infect. Immun.* 81, 965–973. doi: 10.1128/IAI.01197-12

Wang, H., Li, J., Guo, B., Zhao, L., Zhang, Z., Mcmanus, D. P., et al. (2016). In vitro culture of *Echinococcus multilocularis* producing protoscoleces and mouse infection with the cultured vesicles. *Parasit. Vectors* 9:411. doi: 10.1186/s13071-016-1687-y

Wang, H., Li, J., Pu, H., Hasan, B., Ma, J., Jones, M. K., et al. (2014). *Echinococcus granulosus* infection reduces airway inflammation of mice likely through enhancing IL-10 and down-regulation of IL-5 and IL-17A. *Parasit. Vectors* 7:522. doi: 10.1186/s13071-014-0522-6

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, W., Li, J., Yao, J., Wang, T., Li, S., Zheng, X., et al. (2017). In vitro and in vivo efficacies of novel carbazole aminoalcohols in the treatment of cystic echinococcosis. *J. Antimicrob. Chemother.* 72, 3122–3130. doi: 10.1093/jac/dkx250

Wang, Y., Holmes, E., Nicholson, J. K., Cloarec, O., Chollet, J., Tanner, M., et al. (2004). Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12676–12681. doi: 10.1073/pnas.0404878101

Wegener Parfrey, L., Jirku, M., Sima, R., Jalovecka, M., Sak, B., Grigore, K., et al. (2017). A benign helminth alters the host immune system and the gut microbiota in a rat model system. *PLoS One* 12:e0182205. doi: 10.1371/journal.pone.0182205

Wu, X., Ma, C., Han, L., Nawaz, M., Gao, F., Zhang, X., et al. (2010). Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Curr. Microbiol.* 61, 69–78. doi: 10.1007/s00284-010-9582-9

Yu, L. C., Wei, S. C., and Ni, Y. H. (2018). Impact of microbiota in colorectal carcinogenesis: lessons from experimental models. *Intest. Res.* 16, 346–357. doi: 10.5217/ir.2018.16.3.346

Zeng, M. Y., Inohara, N., and Nunez, G. (2017). Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* 10, 18–26. doi: 10.1038/mi.2016.75

Zhang, W., Li, J., and Mcmanus, D. P. (2003). Concepts in immunology and diagnosis of hydatid disease. *Clin. Microbiol. Rev.* 16, 18–36. doi: 10.1128/CMR.16.1.18-36.2003

Zhang, W., Wen, H., Li, J., Lin, R., and Mcmanus, D. P. (2012). Immunology and immunodiagnosis of cystic echinococcosis: an update. *Clin. Dev. Immunol.* 2012:101895. doi: 10.1155/2012/101895

Zhang, W., You, H., Zhang, Z., Turson, G., Hasyet, A., and Mcmanus, D. P. (2001). Further studies on an intermediate host murine model showing that a primary *Echinococcus granulosus* infection is protective against subsequent oncospheral challenge. *Parasitol. Int.* 50, 279–283. doi: 10.1016/S1383-5769(01)00086-1

Zhang, W. B., Jones, M. K., Li, J., and Mcmanus, D. P. (2005). *Echinococcus granulosus*: pre-culture of protoscoleces in vitro significantly increases development and viability of secondary hydatid cysts in mice. *Exp. Parasitol.* 110, 88–90. doi: 10.1016/j.exppara.2005.02.003

Zoetendal, E. G., Raes, J., Van Den Bogert, B., Arumugam, M., Booijink, C. C., Troost, F. J., et al. (2012). The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* 6, 1415–1426. doi: 10.1038/ismej.2011.212

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 Bao, Zheng, Wang, Zheng, He, Qi, Wang, Guo, Guo, Zhang, Zhang, Li and McManus. 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: 19 December 2017
doi: 10.3389/fmicb.2017.02501

Genome-Scale Data Call for a Taxonomic Rearrangement of *Geodermatophilaceae*

Maria del Carmen Montero-Calasanz[1,2], Jan P. Meier-Kolthoff[2], Dao-Feng Zhang[3], Adnan Yaramis[1,4], Manfred Rohde[5], Tanja Woyke[6], Nikos C. Kyrpides[6], Peter Schumann[2], Wen-Jun Li[3]* and Markus Göker[2]**

[1] School of Biology, Newcastle University, Newcastle upon Tyne, United Kingdom, [2] Leibniz Institute, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, [3] State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, [4] Department of Biotechnology, Middle East Technical University, Ankara, Turkey, [5] Central Facility for Microscopy, Helmholtz Centre for Infection Research, Braunschweig, Germany, [6] Department of Energy, Joint Genome Institute, Walnut Creek, CA, United States

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Haitham Sghaier, Centre National des Sciences et Technologies Nucléaires, Tunisia
Tomoo Sawabe, Hokkaido University, Japan

***Correspondence:**
Maria del Carmen Montero-Calasanz
maria.montero-calasanz@ncl.ac.uk
Wen-Jun Li
liwenjun3@mail.sysu.edu.cn
Markus Göker
markus.goeker@dsmz.de

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

Received: *08 August 2017*
Accepted: *01 December 2017*
Published: *19 December 2017*

Citation:
Montero-Calasanz MC, Meier-Kolthoff JP, Zhang D-F, Yaramis A, Rohde M, Woyke T, Kyrpides NC, Schumann P, Li W-J and Göker M (2017) Genome-Scale Data Call for a Taxonomic Rearrangement of Geodermatophilaceae. Front. Microbiol. 8:2501. doi: 10.3389/fmicb.2017.02501

Geodermatophilaceae (order *Geodermatophilales*, class *Actinobacteria*) form a comparatively isolated family within the phylum *Actinobacteria* and harbor many strains adapted to extreme ecological niches and tolerant against reactive oxygen species. Clarifying the evolutionary history of *Geodermatophilaceae* was so far mainly hampered by the insufficient resolution of the main phylogenetic marker in use, the 16S rRNA gene. In conjunction with the taxonomic characterisation of a motile and aerobic strain, designated YIM M13156^T and phylogenetically located within the family, we here carried out a phylogenetic analysis of the genome sequences now available for the type strains of *Geodermatophilaceae* and re-analyzed the previously assembled phenotypic data. The results indicated that the largest genus, *Geodermatophilus*, is not monophyletic, hence the arrangement of the genera of *Geodermatophilaceae* must be reconsidered. Taxonomic markers such as polar lipids and fatty-acids profile, cellular features and temperature ranges are indeed heterogeneous within *Geodermatophilus*. In contrast to previous studies, we also address which of these features can be interpreted as apomorphies of which taxon, according to the principles of phylogenetic systematics. We thus propose a novel genus, *Klenkia*, with the type species *Klenkia marina* sp. nov. and harboring four species formerly assigned to *Geodermatophilus, G. brasiliensis, G. soli, G. taihuensis,* and *G. terrae*. Emended descriptions of all species of *Geodermatophilaceae* are provided for which type-strain genome sequences are publicly available. Our study again demonstrates that the principles of phylogenetic systematics can and should guide the interpretation of both genomic and phenotypic data.

Keywords: *Klenkia, Geodermatophilus*, *Modestobacter, Blastococcus*, GBDP, GGDC, phylogenetic systematics, polyphasic taxonomy

INTRODUCTION

The order *Geodermatophilales* (Sen et al., 2014) comprises the sole family *Geodermatophilaceae*, which was initially proposed by Normand et al. (1996), although no type genus was designated at that time, confirmed later by Stackebrandt et al. (1997), formally described by Normand (2006) and later emended by Zhi et al. (2009). The family accommodates the genera *Blastococcus*

(Ahrens and Moll, 1970; Skerman et al., 1980; Hezbri et al., 2016b), *Modestobacter* (Mevs et al., 2000) and the type genus *Geodermatophilus* (Luedemann, 1968; Skerman et al., 1980).

Geodermatophilus was historically poorly studied due to difficulties in culturing novel isolates (Urzì et al., 2004). Overcoming those technical difficulties, the number of validly named species within the genus dramatically increased from a single species, *G. obscurus*, in 2011 to twenty-one species at the time of writing (Parte, 2014). The number of *Geodermatophilus* isolates is expected to continue to raise in coming years, as indicated by metagenomics studies carried out in arid and hyper-arid habitats (Neilson et al., 2012; Giongo et al., 2013). Species belonging to the genus are indeed mainly isolated from arid soils and characterized by tolerance against oxidative stress (Gtari et al., 2012; Montero-Calasanz et al., 2013a, 2014b, 2015; Hezbri et al., 2015a, 2016a) although some isolates from rhizospheric soils and lake sediments have also been classified within the genus.

In the post-genomic era, the integration of genomic information in microbial systematics (Klenk and Göker, 2010) in addition to physiological and chemotaxonomic parameters as taxonomic criteria is strongly suggested for classifying prokaryotes (Ramasamy et al., 2014). This particularly holds in groups such as *Geodermatophilaceae*, which are only incompletely resolved in phylogenies inferred from the most commonly applied marker gene, the 16S rRNA gene. Nevertheless, except for the genome sequences generated within our project only the genome (Ivanova et al., 2010) and proteome sequence of *G. obscurus* (Sghaier et al., 2016) were publicly available.

Based on phylogenies inferred from genome-scale data and on a re-interpretation of the available phenotypic evidence according to the principles of phylogenetic systematics (Hennig, 1965; Wiley and Lieberman, 2011), this study introduces the new genus *Klenkia* into *Geodermatophilaceae,* whose type species is *Klenkia marina* sp. nov. Accordingly, we also propose the reclassification of *G. brasiliensis* as *Klenkia brasiliensis* comb. nov., *G. soli* as *K. soli* comb. nov., *G. taihuensis* as *K. taihuensis* comb. nov. and *G. terrae* as *K. terrae* comb. nov., as well as emended descriptions within *Geodermatophilus*.

MATERIALS AND METHODS

Isolation

Strain YIM M13156^T was isolated from a sample collected from the South China Sea (119° 31.949E, 18° 2.114 N), and was obtained using the serial dilution technique. Sediment sample (1 g) was added to 9 ml sterile distilled water and mixed by vortexing. A 10-fold dilution of this soil suspension was prepared in sterilized distilled water, and 0.1 ml was spread on Fucose-proline agar medium [fucose 5 g; proline 1 g; $(NH_4)_2SO_4$ 1 g; NaCl 1 g;$CaCl_2$ 2 g; K_2HPO_4 1 g; B-Vitamin trace (0.5 mg each of thiamine-HCl (B_1), riboflavin, Niacin, pyridoxin, Ca-pantothenate, inositol, *p*-aminobenzoic acid, and 0.25 mg of biotin); sea salt 30 g; agar 20 g; pH 7.2; distilled water 1 liter]. The plate was then incubated at 28°C for 30 days.

Phenotypic Analysis

Morphological and Physiological Tests

Morphological characteristics of strain YIM M13156^T were determined on GYM *Streptomyces* medium at 28°C. Colony features were observed at 4 and 15 days under a stereo microscope according to Pelczar (1957). Exponentially growing bacterial cultures were observed with an optical microscope (Zeiss AxioScope A1) with a 1000-fold magnification and phase-contrast illumination. Gram reaction was performed using the KOH test described by Gregersen (1978). Oxidase activity was analyzed using filter-paper disks (Sartorius grade 388) impregnated with 1% solution of *N,N,N′,N′*-tetramethyl-*p*-phenylenediamine (Sigma–Aldrich); a positive test was defined by the development of a blue-purple color after applying biomass to the filter paper. Catalase activity was determined based on formation of bubbles following the addition of 1 drop of 3% H_2O_2. Growth rates were determined on plates of GYM *Streptomyces* medium for temperatures from 10°C to 50°C at 5°C increments and for pH values from 4.0 to 12.5 (in increments of 0.5 pH units) on modified ISP2 medium by adding NaOH or HCl, respectively, since the use of a buffer system inhibited growth of the strain. The oxidation of carbon compounds was tested at 28°C using GEN III Microplates in an Omnilog device (BIOLOG Inc., Hayward, CA, United States) in comparison with the reference strains *G. brasiliensis* DSM 44526^T, *G. soli* DSM 45843^T, *G. taihuensis* DSM 45962^T, *G. terrae* DSM 45844^T in parallel assays. The GEN III Microplates were inoculated with cells suspended in a viscous inoculating fluid (IF C) provided by the manufacturer at a cell density of 94% T for *G. brasiliensis* DSM 44526^T, 83% T for *G. soli* DSM 45843^T and *G. terrae* DSM 45844^{T}and 90% T for *G. taihuensis* DSM 45962^T and for the strain YIM M13156^T. Respiration rates were measured yielding a total running time of 5 days in Phenotype Microarray mode. Each strain was studied in two independent technical replicates. Data were exported and analyzed using the *opm* v.1.0.6 package (Vaas et al., 2012, 2013) for the *R* statistical environment (R Core Team, 2017). Reactions with a distinct behavior between the two replicates were regarded as ambiguous.

Chemotaxonomic Tests

Whole-cell amino acids and sugars were prepared according to Lechevalier and Lechevalier (1970), followed by thin-layer chromatography (TLC) analysis (Staneck and Roberts, 1974). Polar lipids were extracted, separated by two-dimensional TLC and identified according to procedures outlined by Minni et al. (1984) with modifications proposed by Kroppenstedt and Goodfellow (2006). For identification the presumed OH-PE spots were manually scraped off from unstained TLC plates, extracted with methanol and evaporated to dryness. The extracts were then dissolved in Reagent 3 (fatty acid extraction) and analyzed by MIDI System. Menaquinones (MK) were extracted from freeze-dried cell material using methanol as described by Collins (1985) and analyzed by high-performance liquid chromatography (HPLC) (Kroppenstedt, 1982). The extraction and analysis of cellular fatty acids was carried out in two independent repetitions from biomass grown on GYM agar plates held at 28°C for 4 days and harvested always from the same

sector (the last quadrant streak). Analysis was conducted using the Microbial Identification System (MIDI) Sherlock Version 4.5 (method TSBA40, ACTIN6 database) as described by Sasser (1990). The annotation of the fatty acids in the ACTIN6 peak naming table is consistent with IUPAC nomenclature. Fatty-acid patterns were visualized as a heatmap using the lipid extension of the *opm* package (Vaas et al., 2013) and clustered using the *pvclust* v.1.2.2 package (Suzuki and Shimodaira, 2006) for the *R* statistical environment. Quantitative analysis of the fatty acids used logit-transformed percentages throughout (after setting zero values to the lowest non-zero number) because proportion data are expected to vary stronger around 50% than close to 0 or 100% (Crawley, 2007). All chemotaxonomic analyses were conducted under standardized conditions for strain YIM M13156^T and the type strains listed above.

Except for the separately stored MIDI measurements, all phenotypic characters were collected in a standardized tabular format that constitutes Supplementary Table S1. Custom scripts developed at DSMZ allow for extracting such data in ways suitable for subsequent phylogenetic or other analysis. For determining features specific for predetermined groups of interest, such as new taxa suggested by phylogenetic analysis, the *randomForest* function from the eponymous *R* package v. 4.6-12 (Breiman, 2001) was applied in classification mode under default settings except for increased values of *ntree* (50,000) and *mtry* (half the number of analyzed features).

Sequence Analysis

For 16S rRNA gene sequencing, genomic DNA extraction, PCR-mediated amplification of the 16S rRNA gene and purification of the PCR product was carried out as described by Rainey et al. (1996). For genome sequencing, the strain YIM M13156^T and sixteen species with validly published *Geodermatophilus* names were cultivated in GYM *Streptomyces* broth at 28°C. The project information is available through the Genomes Online Database (Mukherjee et al., 2017). The draft genomes were generated at the DOE Joint Genome Institute (JGI) as part of Genomic Encyclopedia of Archaeal and Bacterial Type Strains, Phase II (KMG-II): from individual species to whole genera (Kyrpides et al., 2014) following the same protocol as in Nouioui et al. (2017). All genomes were annotated using the DOE-JGI annotation pipeline (Huntemann et al., 2015; Chen et al., 2016) and released through the Integrated Microbial Genomes system (Chen et al., 2017).

Phylogenetic analysis of the 16S rRNA gene sequences from the type strains of all species with effectively published names in *Geodermatophilaceae*, as well as the genome-sequenced strains *Cryptosporangium arvum* DSM 44712^T and *Sporichthya polymorpha* DSM 43042^T for use as outgroup, was conducted as previously described (Göker et al., 2011; Montero-Calasanz et al., 2014b). Pairwise 16S rRNA gene similarities were calculated as recommended by Meier-Kolthoff et al. (2013b) to determine strains with ≥99.0% similarity, between which (digital) DNA:DNA hybridization experiments should be conducted. Genome-scale phylogenies were inferred from the available *Geodermatophilaceae* (and outgroup) whole proteome sequences using the high-throughput version (Meier-Kolthoff et al., 2014a) of the genome BLAST Distance Phylogeny (GBDP) approach (Auch et al., 2010) in conjunction with FastME (Lefort et al., 2015) as described earlier (Hahnke et al., 2016). An additional FastME tree was inferred without the two outgroup genomes to detect potential long-branch attraction to the outgroup, a process called long-branch extraction (Siddall and Whiting, 1999). Additionally, the validity of the rooting was tested by re-estimating the root (Simon et al., 2017) using least-squares dating (To et al., 2015). The GBDP tree restricted to the well-supported branches (≥95% pseudo-bootstrap support) was used as a backbone constraint in a further 16S rRNA gene analysis to integrate information from genome-scale data (Hahnke et al., 2016). Digital DNA:DNA hybridisations were conducted using the recommended settings of the Genome-To-Genome Distance Calculator (GGDC) version 2.1 (Meier-Kolthoff et al., 2013a). The G+C content was calculated from the genome sequences as described by Meier-Kolthoff et al. (2014b).

RESULTS

Sequence Analysis

The phylogenetic tree based on the whole proteomes of the sequenced type strains placed, with maximum support, the group formed by the strain YIM M13156^T, *G. brasiliensis* DSM 44526^T, *G. soli* DSM 45843^T and *G. taihuensis* DSM 45962^T as a sister-group of all other *Geodermatophilaceae* genera (**Figure 1**). Hence *Geodermatophilus* is obviously non-monophyletic according to this tree, as *Modestobacter* as well as *Blastococcus* appeared as more closely related to core *Geodermatophilus* (including the type species *G. obscurus*) than the four deviating *Geodermatophilus* strains. When *C. arvum* and *S. polymorpha* were removed, the resulting unrooted topologically was identical to the (unrooted) ingroup topology shown in **Figure 1**. That is, it was impossible to root the reduced tree in a way that made *Geodermatophilus* monophyletic, indicating that the non-monophyly of *Geodermatophilus* was not caused by a long-branch attraction artifact. Least-squares dating confirmed the rooting in both the full and the reduced tree. As expected, the unconstrained phylogenetic tree based on 16S rRNA gene sequences was not well resolved at its backbone but when applying the constraint derived from the GBDP tree, *Modestobacter, Blastoccocus* and the two distinct groups of *Geodermatophilus* appeared as monophyletic (**Figure 2**); support for their monophyly was strong except in the case of *Blastococcus*. Moderate to strong support was obtained for *Modestobacter* and *Blastoccocus* being more closely related to core *Geodermatophilus* than the group formed by the strain YIM M13156^T as well as the species *G. brasiliensis, G. soli, G. taihuensis* and *G. terrae*.

Within this clade, the 16S rRNA gene sequence of strain YIM M13156^T showed similarities ≥ 99.0% with the type strains of *G. soli* (99.1%) and *G. brasiliensis* (99.0%) only; those with *G. terrae* (98.9%) and *G. taihuensis* (98.0%) were lower. Digital DNA:DNA hybridisations between strain YIM M13156^T and *G. soli* DSM 45843^T and *G. brasiliensis* DSM 44526^T resulted in 29.8 and 29.6% similarity, respectively, clearly below the 70%

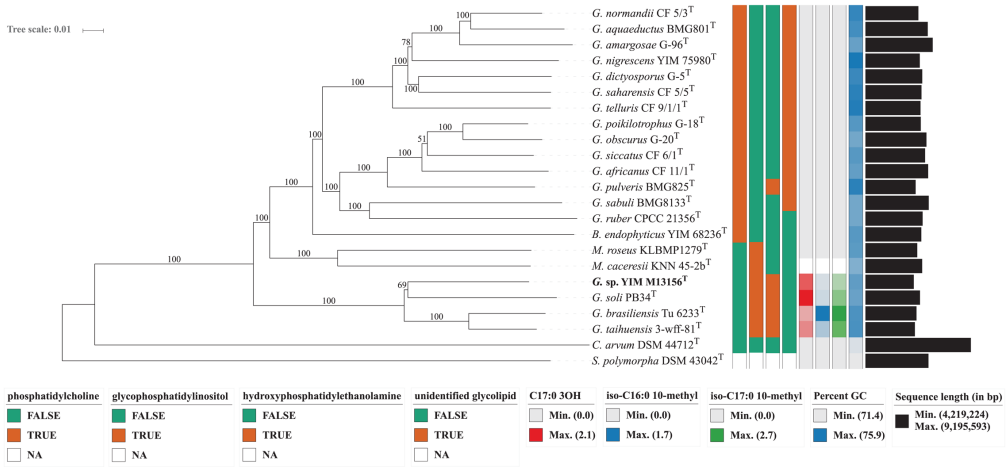


FIGURE 1 | Phylogenomic tree inferred with GBDP. The tree was inferred with FastME from GBDP distances calculated from whole proteomes. The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications. Tip colors indicate chemotaxonomic characters that provide apomorphies for groups of interest, genome sizes and the exact G+C content as calculated from the genome sequences (see the embedded legend for details). *B., Blastococcus*; *C., Cryptosporangium; G., Geodermatophilus; M., Modestobacter; S., Sporichthya*; NA, not applicable.

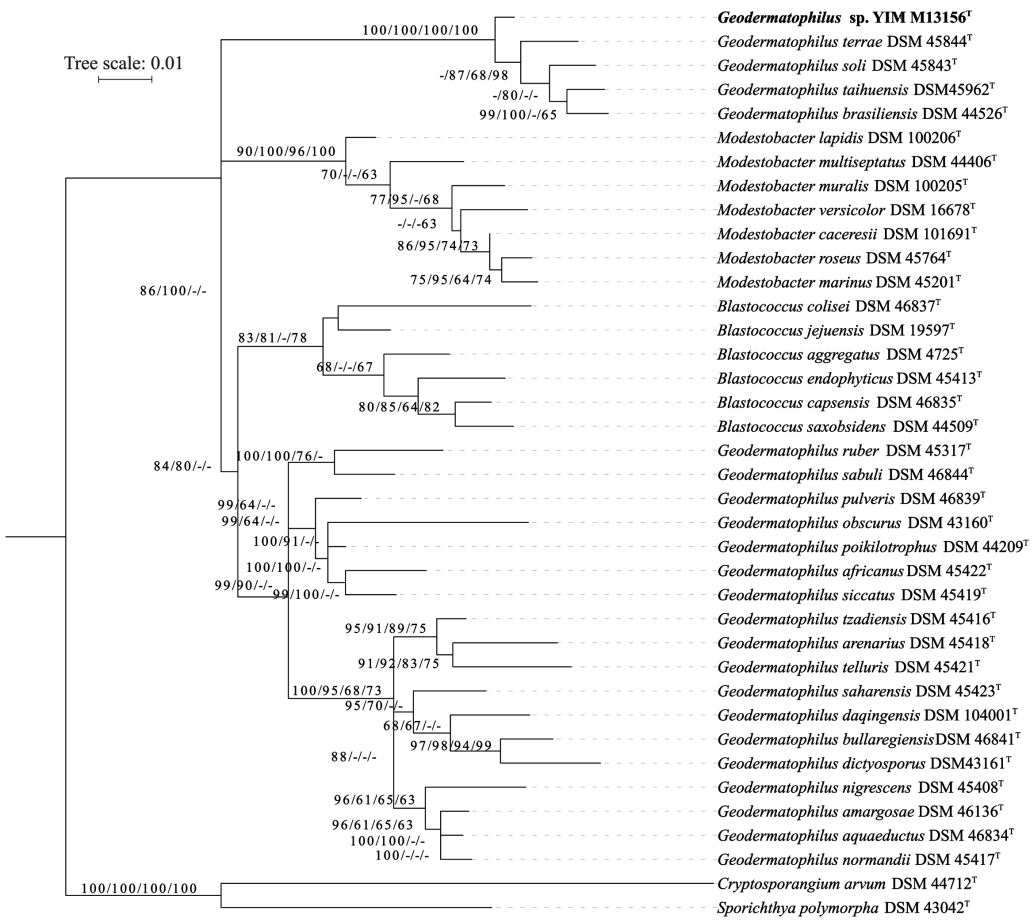


FIGURE 2 | Maximum likelihood phylogenetic tree inferred from 16S rRNA gene sequences, showing the phylogenetic position of the strain YIM M13156^T relative to the type strains within *Geodermatophilaceae*. The branches are scaled in terms of the expected number of substitutions per site (see size bar). Support values from maximum-likelihood constrained (first), maximum-parsimony constrained (second), maximum-likelihood unconstrained (third) and maximum-parsimony unconstrained (fourth) bootstrapping are shown above the branches if equal to or larger than 60%.

threshold recommended by Wayne et al. (1987) to confirm the species status of novel strains. Additional digital DNA:DNA hybridisations were not conducted, based on the observation of Meier-Kolthoff et al. (2013b) that an *Actinobacteria*-specific 16S rRNA threshold of 99.0% yielded a maximum probability of error of only 1% to obtain DNA:DNA hybridization values ≥ 70%.

The genome size range found in the sequenced type strains varied between 4.2 Mbp for the strain YIM M13156^T and 5.9 Mbp for *G. amargosae* DSM 46136^T (**Figure 1**). The genomic G+C content of strain YIM M13156^T was 74.4%. For the other genomes it ranged between 74.0% for *G. obscurus, G. ruber* and *G. sabuli* and 75.9% for *G. nigrescens* (**Figure 1**). Because G+C content values do not differ more than 1% within bacterial species (Meier-Kolthoff et al., 2014b), stronger deviations are due to artifacts in conventionally determined G+C content values. Hence, we accordingly propose to emend the species descriptions of those species for which we observed a deviation from published G+C content values > 1%.

Phenotypic Analysis

Morphology and Physiology

Strain YIM M13156^T showed motile, rod-shaped and Gram-positive cells. These observations are in line with those described by Jin et al. (2013), Qu et al. (2013), and Bertazzo et al. (2014) for *G. soli* and *G. terrae*, *G. taihuensis,* and *G. brasiliensis,* respectively. In contrast to these four species, neither high aggregates forming multilocular sporangia nor zoospores were identified in YIM M13156^T cultures (Supplementary Table S1). Colonies were pink-colored, convex, circular and opaque with a smooth surface and an entire margin, an appearance similar to other *Geodermatophilus* species when cultivated under the same growth conditions. Cell growth ranged from 20 to 35°C (optimal growth temperature 25–30°C) and from pH 6.0 to 8.5 (optimal pH 6.5–8.0). Results from phenotype microarray analysis are shown as a heatmap in the Supplementary Material (Supplementary Figure S1) in comparison to the type strains of the four most closely related species. Differences between species were much more pronounced than between replicates. A summary of selected phenotypic characteristics is presented in **Table 1** (for an overview of phenotypic profiles in *Geodermatophilus* see Supplementary Table S1).

The *randomForest* analysis indicated that among the previously mentioned phenotypic features the cell shape well discriminates between the group formed by strain YIM M13156^T, *G. brasiliensis, G. soli, G. taihuensis* and *G. terrae*, which produce rods, on the one hand and core *Geodermatophilus*, which is characterized by pleomorphic cells (and seldom cocci), on the other hand (Supplementary Figure S2).

Chemotaxonomy

Analysis of whole-cell components revealed the presence of *meso*-diaminopimelic acid (Cell-wall type III), which is consistent with the other representatives of *Geodermatophilaceae* (Normand and Benson, 2012).

Strain YIM M13156^T displayed primarily menaquinone MK-9(H_4) (52.7%), in agreement with values reported for *Geodermatophilaceae* (Normand, 2006), and MK-9(H_0) (39.2%). The presence of a significant amount of MK-9(H_0) was already mentioned in the original descriptions of *G. brasiliensis*, *G. soli, G. taihuensis,* and *G. terrae*, but also in other *Geodermatophilus* species (Montero-Calasanz et al., 2015). In contrast to the already described profiles of isoprenologs of *G. soli, G. taihuensis* and *G. terrae,* traces of MK-9(H_2) (6.6, 1.5, and 2.8%, respectively) were now additionally identified in those type strains. MK-8(H_4) (5.7%) and MK-10(H_4) (3.4%) were also detected in the patterns of *G. taihuensis* DSM 45962^T and *G. terrae* DSM 45844^T, respectively.

The polar lipid pattern of strain YIM M13156^T consisted of diphosphatidylglycerol (DPG), phosphatidylethanolamine (PE), phosphatidylinositol (PI), glycophosphatidylinositol (GPI), an unidentified aminolipid (AL) and traces of hydroxyphosphatidylethanolamine (OH-PE) (**Figure 3a**). It is in accordance with patterns obtained for the closely related species investigated in this study (**Figures 3b–d**) and the phospholipid pattern revealed by Bertazzo et al. (2014) for *G. brasiliensis*. The *randomForest* analysis detected the lack of phosphatidylcholine and the presence of glycophosphatidylinositol as excellently predictive of the prospective new genus, whereas the absence of phosphatidylglycerol was also observed in other *Geodermatophilus* species (Bertazzo et al., 2014; Montero-Calasanz et al., 2014b, 2015; Hezbri et al., 2015a,b,c, 2016a). The unambiguous presence of hydroxyphosphatidylethanolamine is a slightly less relevant taxonomic marker for the prospective new genus; among core *Geodermatophilus* it was only detected in the polar lipid profile of *G. pulveris* DSM 45839^T by Hezbri et al. (2016a).

Even though both absence of phosphatidylcholine and presence of glycophosphatidylinositol were also observed in the polar-lipid profiles of some *Modestobacter* species (Montero-Calasanz et al., in preparation), the additional presence of hydroxyphosphatidylethanolamine forms a unique pattern of strain YIM M13156^T and its four most closely related species. In addition, based on our results and the chromatographic mobility of the polar lipid labeled as phosphatidylmethylethanolamine in the original descriptions of *G. soli* and *G. terrae* by Jin et al. (2013), it is strongly suggested that it was not correctly identified in the original work, since after binding a methyl-group to phosphatidylethanolamine the resultant component would show a higher apolarity and therefore a higher mobility on the plate than phosphatidylethanolamine itself. Hydroxyphosphatidylethanolamine is not known from the outgroup species *Cryptosporangium arvum* (Tamura et al., 1998).

Major fatty acids were the saturated branched-chain iso-$C_{16:0}$ (36.8 ± 1.1%), the monounsaturated $C_{17:1}\omega8c$ (13.4 ± 0.4%) and the saturated branched-chain iso-$C_{15:0}$ (11.5 ± 0.5%) complemented by iso-$C_{16:1}$ H (5.0 ± 0.4%), $C_{17:0}$ (4.5 ± 0.3%) and $C_{18:1}9\omega c$ (5.0 ± 0.2%) in agreement with the closest related species (**Table 1**; for an overview of fatty-acid profiles in *Geodermatophilus* see Supplementary Table S2). In addition, the occurrence of 2-hydroxy fatty acids (mainly iso-$C_{17:0}$ 2OH) is also worth mentioning as it supports the presence of OH-PE observed in the polar lipids profiles of strain YIM M13156^T and its four most closely related species (for an overview of fatty-acids profiles in *Geodermatophilus* see Supplementary Table S2). The

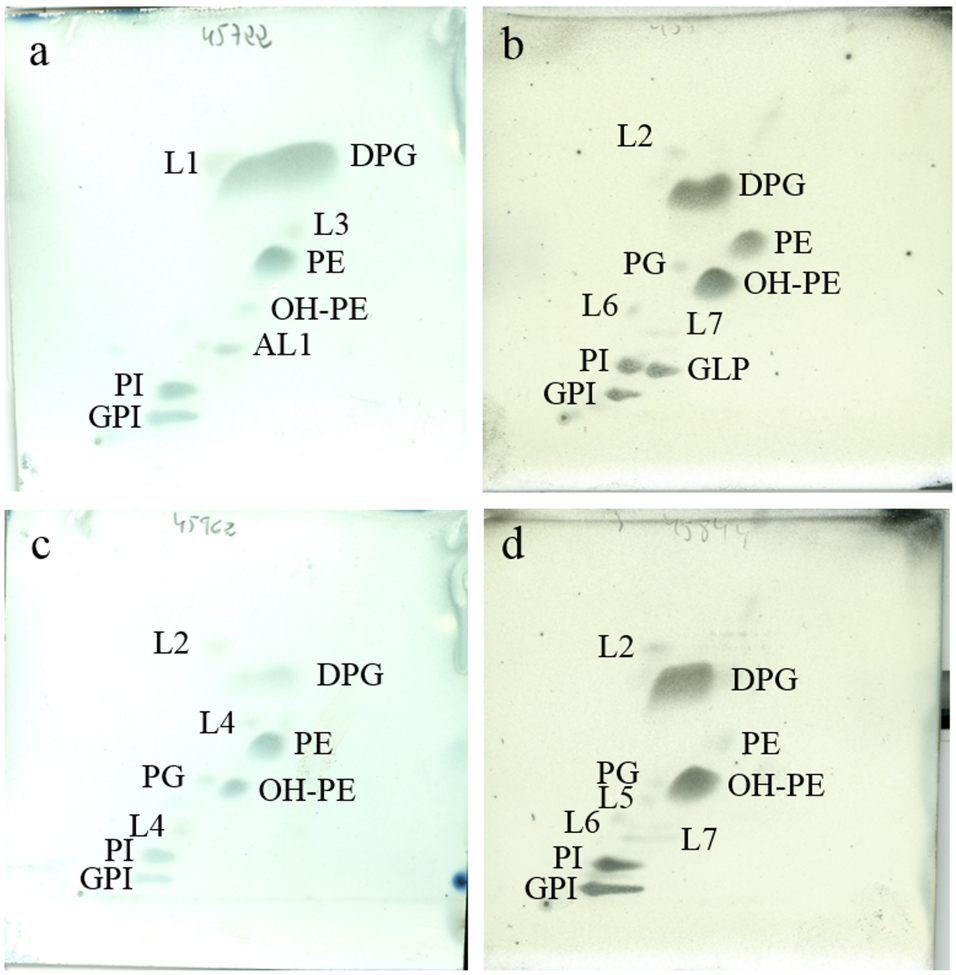


FIGURE 3 | Polar lipids profile of strain YIM M13156^T **(a)**, *G. soli* DSM 45843^T **(b)**, *G. taihuensis* DSM 45962^T **(c)** and *G. terrae* DSM 45844^T **(d)** after separation by two-dimensional TLC using the solvents chloroform:methanol:water (65:25:4; v:v:v) in the first dimension and chloroform:methanol:acetic acid:water (80:12:15:4; v:v:v:v) in the second one. Plates were sprayed with molybdatophosphoric acid (3.5%; Merck^{TM}) for detection of the total polar lipids. DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; OH-PE, hydroxyphosphatidylethanolamine; PI, phosphatidylinositol; GPI, glycophosphatidylinositol; GPL, unidentified glycophospholipid; AL, aminolipid; L1-7, unidentified lipids. All data are from this study.

2-hydroxy fatty acids are a pre-requisite for the synthesis of the hydroxylated polar lipid (Kämpfer et al., 2010).

The clustering analysis of the logit-transformed fatty-acid profiles revealed that those of YIM M13156^T, *G. brasiliensis, G. soli, G. taihuensis,* and *G. terrae* separated first. Hence the profiles of the other *Geodermatophilus* species were more similar to the ones of *Blastococcus* and *Modestobacter* (**Figure 4**). Accordingly, the *randomForest* analysis identified three minor components (iso-$C_{17:0}$ 10-methyl, $C_{17:0}$ 3OH and iso-$C_{16:0}$ 10-methyl) highly predictive of the group formed by strain YIM M13156^T, *G. brasiliensis, G. soli, G. taihuensis,* and *G. terrae*. The fatty-acid profiles thus even independently supported the assignment of these five species to a new genus (Supplementary Figures S3, S4).

Whole-cell sugar analysis revealed rhamnose, ribose, mannose, glucose and an unidentified sugar showing a similar chromatographic mobility than the unidentified sugar found in *G. normandii* DSM 45417^T by Montero-Calasanz et al. (2013b). On the other hand, *G. soli* DSM 45843^T, *G. taihuensis* DSM 45962^T and *G. terrae* DSM 45844^T showed the same sugar profile as *G. brasiliensis* DSM 45426^T (Bertazzo et al., 2014), consisting of ribose, mannose, glucose and galactose (Lechevalier and Lechevalier, 1970). The absence of galactose in the profile of strain YIM M13156^T might differentiate this species from others within the group.

In order to standardize the phenotypic data available for the genus *Geodermatophilus,* analyses of polar lipids, whole-cell sugars and menaquinones were also carried out for the species *G. obscurus* DSM 43160^T, *G. ruber* DSM 45317^T and *G. nigrescens* DSM 45408^T. The polar lipid and menaquinone profiles of those species were already specified in the **Table 1** of the original description of *G. arenarius* DSM 45418^T by Montero-Calasanz et al. (2012), nevertheless they were never properly described nor the species emended. The three species showed the typical

TABLE 1 | Phenotypic characteristics of strain YIM M13156^T in comparison to those of the type strains of the most closely related *Geodermatophilus* species.

Characteristics	1	2	3	4	5
Motility	+	−	+	+	+
Oxidation of:					
Stachyose	−	−	+	+/−	−
α-D-Lactose	−	+	+	+/−	+
D-Melibiose	−	+/−	+	+	−
N-Acetyl-D-Galactosamine	−	+	−	−	−
D-Mannose	+	+	+	−	+
L-Rhamnose	−-	+	+	−	+
L-Aspartic acid	−	−	−	−	+
Quinic acid	−	+	+	+	+
L-Lactic acid	−	−	+	−	+
Citric acid	+	−	+	−	−
α-Keto-Glutaric acid	+	−	−	−	+
D-Malic acid	+	−	+	−	+
L-Malic acid	+	−	+	−	+
Polar lipids	DPG, PE, OH-PE, PI, GPI	DPG, PE, OH-PE, PI, GPI	DPG, PG, PE, OH-PE, PI, GPI, GPL	DPG, PG, PE, OH-PE, PI, GPI	DPG, PG, PE, OH-PE, PI, GPI
Sugars	Rham, Rib, Man, Gluc, US	Rib, Man, Gluc, Gal	Rib, Man, Gluc, Gal	Rib, Man, Gluc, Gal	Rib, Man, Gluc, Gal
Menaquinones[a#]	MK-9(H_4), MK-9(H_0)	MK-9(H_4), MK-9(H_2), MK-9(H_0), MK-10(H_4)	MK-9(H_4), MK-9(H_2), MK-9(H_0)	MK-9(H_4), MK-9(H_2), MK-9(H_0), MK-8(H_4)	MK-9(H_4), MK-9(H_2), MK-9(H_0), MK-10(H_4)
Fatty acids[b#]	iso-$C_{16:0}$, $C_{17:1}\omega8c$, iso-$C_{15:0}$	iso-$C_{15:0}$, iso-$C_{16:0}$	iso-$C_{15:0}$, iso-$C_{16:0}$, iso-$C_{17:0}$	iso-$C_{15:0}$, iso-$C_{16:0}$	iso-$C_{16:0}$, iso-$C_{15:0}$, $C_{18:1}\omega9c$

Strains: 1, strain YIM M13156^T; 2, G. brasiliensis DSM 44526^T; 3, G. soli DSM 45843^T; 4, G. taihuensis DSM 45962^T; 5, G. terrae DSM 45844^T. All data are from this study. +, positive reaction; −, negative reaction; +/−, ambiguous; DPG, diphosphatidylglycerol; PE, phosphatidylethanolamine; PE-OH, hydroxy-phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; GPL, unidentified glycophospholipid; APL, unidentified amino-phospholipid; Rham, rhamnose; Rib, ribose; Man, mannose; Gluc, glucose; US, unidentified sugar; MK, menaquinones; iso-, iso-branched. [a]*only components making up ≥ 1% peak area ratio are shown;* [b]*only components making up ≥ 10% peak area ratio are shown;* [#]*, the components are listed in decreasing order of quantity.*

polar-lipid profile observed in *Geodermatophilus* consisting of diphosphatidylglycerol (DPG), phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylinositol (PI) and minor amounts of phosphatidylglycerol (PG) (see Supplementary Figure S5). Similar to the group formed by *G. brasiliensis* DSM 44526^T, *G. soli* DSM 45843^T, *G. taihuensis* DSM 45962^T and *G. terrae* DSM 45844^T, *G. ruber* DSM 45317^T displayed glycophosphatidylinositol (GPI) in addition to an unidentified phospholipid (PL). The polar-lipid profiles of *G. obscurus* DSM 43160^T and *G. nigrescens* DSM 45408^T conversely revealed the typical unidentified glycolipid already described for most *Geodermatophilus* species (See supplementary Table S1). Regarding the MK pattern of *G. obscurus* DSM 43160^T, apart from MK-9(H_4) (64.3%) already indicated by Zhang et al. (2011), MK-9(H_2) (8.7%), MK-9(H_0) (4.9%) and MK-8(H_4) (4.3%) were also revealed. Nevertheless, in contrast to Zhang et al. (2011) and in addition to MK-9(H_4), our studies did not identify MK-9(H_0) in the profile of *G. ruber* DSM 45317^T. The whole-cell sugar patterns of *G. obscurus* DSM 43160^T revealed the presence of ribose, xylose, mannose, glucose and galactose. *G. ruber* DSM 45317^T displayed a profile consisting of ribose and glucose. Differently from Nie et al. (2012) who identified galactose, arabinose and glucosamine as the whole-cell sugar patters of *G. nigrescens* DSM 45408^T, our results showed a profile comprising mannose, glucose, galactose and traces of rhamnose and ribose. These profiles are consistent with those previously described in the genus, although it is worth mentioning the absence of galactose in the profiles of *G. ruber* DSM 45317^T, a feature shared, as mentioned previously, with the strain YIM M13156^T. The presence of xylose was already described for *G. saharensis* DSM 45423^T (Montero-Calasanz et al., 2013c).

DISCUSSION

Phylogenetic analysis based on whole genome and 16S rRNA gene sequences revealed with strong support that strain YIM M13156^T and the species *G. brasiliensis* DSM 44526^T, *G. soli* DSM 45843^T, *G. taihuensis* DSM 45962^T and *G. terrae* DSM 45844^T formed a separate lineage within *Geodermatophilaceae*, hence *Geodermatophilus* is not monophyletic. Since the main goal of phylogenetic systematics is to obtain monophyletic taxa (Hennig, 1965; Wiley and Lieberman, 2011) taxonomic consequences are necessary. A lumping approach would require merging all *Geodermatophilaceae* genera into *Geodermatophilus*, which has priority, and thus the generation of 13 new names (i.e., new combinations for *Blastococcus* and *Modestobacter*). In contrast, placing the aberrant *Geodermatophilus* species into a separate

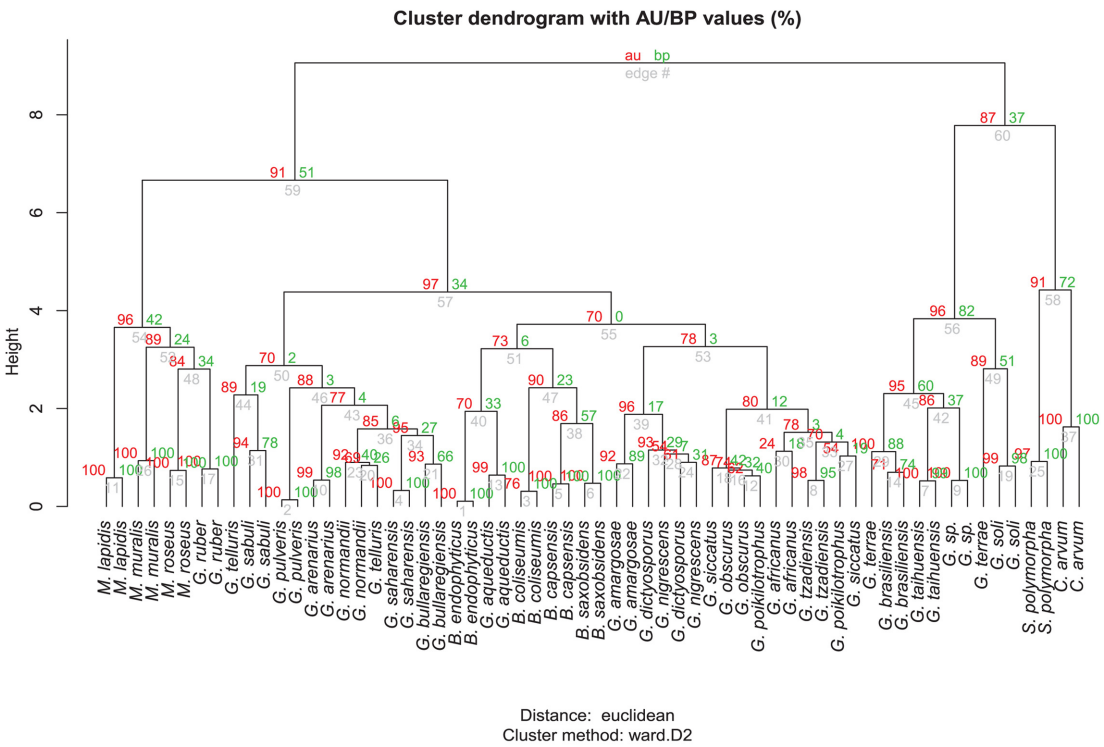


FIGURE 4 | Fatty-acids dendrogram estimated from the logit-transformed percent values as measured with the MIDI system using the Ward algorithm for agglomerative hierarchical clustering and euclidean distances using the *pvclust* package for the *R* statistical environment. Approximately unbiased (AU, left) and bootstrapping (BP, right) support values are shown above the branches.

genus require the generation of only four new names, one for the new genus and four new combinations. Taxonomic conservatism, most easily be measured as inversely proportional to the number of new names to be created (Breider et al., 2014), thus clearly favors the splitting solution. These two arguments alone justify the need to introduce a new genus of *Geodermatophilaceae*.

Nevertheless, the phenotype also provided rich information on the interrelationships of the envisaged new genus. Strain YIM M13156^T and its four neighboring species were distinguished from other genera in the family *Geodermatophilaceae* by cell morphology, the lack of spores, the absence of phosphatidylcholine and the typical unidentified glycolipid found ins *Geodermatophilus*, the presence of glycophosphatidylinositol and hydroxyphosphatidylethanolamine in their polar lipids profiles (for an overview of the characteristics that differentiate the strain YIM M13156^T, *G. brasiliensis, G. soli, G. taihuensis* and *G. terrae* from closely related *Geodermatophilaceae* genera see Supplementary Table S1 and *randomForest* predictions Supplementary Figure S2) and the occurrence of iso-$C_{17:0}$ 2OH and other minor compounds (Supplementary Figure S4 and Supplementary Table S2) in their fatty-acids patterns.

However, in phylogenetic systematics diagnostic features for a group are insufficient to establish it as a taxon because when these features were plesiomorphic (ancestral) instead of apomorphic (derived) then they could well diagnose a paraphyletic group (Hennig, 1965; Wiley and Lieberman, 2011); reptiles are a classical example. For this reason, we studied the distribution of the above listed features among *Geodermatophilaceae* and the outgroup species *Cryptosporangium arvum* (Tamura et al., 1998) and *Sporichthya polymorpha* (Supplementary Figure S4 and Supplementary Table S2) for determining with maximum-parsimony reconstructions which character state was apomorphic for which group (**Figure 1**). Accordingly, presence of hydroxyphosphatidylethanolamine, iso-$C_{17:0}$ 10-methyl, $C_{17:0}$ 3OH and iso-$C_{16:0}$ 10-methyl appeared as synapomorphies of YIM M13156^T and its four neighboring species; presence of glycophosphatidylinositol appeared as autapomorphy of *Geodermatophilaceae*; secondary absence of glycophosphatidylinositol and presence of phosphatidylcholine as synapomorphy of core *Geodermatophilus* and *Blastococcus*; and presence of the unidentified glycolipid as autopomorphies of core *Geodermatophilus*. Hydroxyphosphatidylethanolamine is present in *G. pulveris*, too, but was gained independently; the unidentified glycolipid is missing in *G. ruber* but was secondarily lost (**Figure 1**). The status of cell morphology and spore formation was unclear due to missing data (Supplementary Table S1) but the already assembled evidence clearly supports the envisaged reclassification.

The currently still dominating practice of polyphasic taxonomy (Vandamme et al., 1996) in microbial systematics has increasingly been called into question in recent years (Sutcliffe et al., 2012; Vandamme and Peeters, 2014; Sutcliffe,

2015; Thompson et al., 2015). Critics mainly emphasize that more genomic information should be incorporated and that some of the nowadays routinely conducted phenotypic tests might actually be unnecessary. It was also obvious in the present study that genome-scale data yielded high resolution (**Figure 1**), which via a backbone constraint (Hahnke et al., 2016) could also inform a more comprehensively sampled 16S rRNA gene analysis (**Figure 2**).

Whereas phylogenomics is expected to yield more strongly resolved trees, these might in theory also yield more conflict between distinct analyses (Jeffroy et al., 2006; Klenk and Göker, 2010). Horizontal gene transfer is a known cause of topological conflict between analyses of single genes that has even been be used to argue against hierarchical classification (Bapteste and Boucher, 2009; Klenk and Göker, 2010). However, the increase of support in phylogenomic analyses after adding genes up to virtually all available genes indicates a strong hierarchical signal (Breider et al., 2014), whereas the selection of pre-defined set of few genes does not yield genome-scale data and relies on *a priori* assumptions about the relative suitability of genes for analysis (Lienau and DeSalle, 2009; Klenk and Göker, 2010). Methods such as GBDP, which infer trees rather directly from complete genomes, are more promising for obtaining a truly genome-based classification, but conflict between single genes raises the question of how to not overestimate phylogenetic confidence (Taylor and Piel, 2004). This issue can hardly be overestimated because in phylogenetic systematics taxa must be as well supported as monophyletic as possible (Vences et al., 2013), which rules out all phylogenomic methods that do not even yield statistical support values. Instead of standard bootstrapping, the partition bootstrap, which resamples entire genes, is supposed to reduce conflict and provide more reliable support values (Siddall, 2010; Simon et al., 2017). Within the GBDP pseudo-bootstrapping framework, the greedy-with-trimming algorithm (Meier-Kolthoff et al., 2014a) as applied in the present study is the equivalent of the partition bootstrap (Hahnke et al., 2016).

Properly analyzed genome-scale data thus address the current shortcoming of polyphasic taxonomy that its starting point is an often poorly resolved 16S rRNA gene tree. After choosing taxon boundaries from such a tree, the polyphasic approach would then proceed with determining diagnostic features for the new taxa. Tools such as *randomForest* as used here can assist in the task to select features predictive for a certain group of interest from larger numbers of characters.

However, a more serious problem with the currently dominating polyphasic approach is that such diagnostic features cannot provide independent evidence for taxon boundaries when these boundaries were already used to choose the features. Independent evidence can instead be obtained by detecting the same groups independently when analyzing the additional features (**Figure 3**). Moreover, in phylogenetic systematics diagnostic features are insufficient for justifying a taxon because taxa must be monophyletic, whereas diagnostic character states can be plesiomorphic and thus diagnose a paraphyletic group (Hennig, 1965; Wiley and Lieberman, 2011). To the best of our knowledge, publications applying polyphasic taxonomy hardly ever address these two issues, even though phylogenetic systematics is the appropriate paradigm for microbial taxonomy, too (Klenk and Göker, 2010). Outgroup and ingroup comparisons might sometimes be difficult because of incomplete character sampling but in the present study succeeded in determining that some character states were apomorphies for the envisaged new taxa. We thus believe that microbial taxonomy would not only benefit from incorporating genomic information but also from adhering to the principles of phylogenetic systematics.

Taxonomic Consequences

Based on phenotypic and genotypic data presented, we propose that strain YIM M13156^T represents a novel species of a new genus of *Geodermatophilaceae,* for which the name *Klenkia marina* gen. nov., sp. nov. is proposed. In addition we propose the reclassification of *Geodermatophilus brasiliensis* as *Klenkia brasiliensis* comb. nov., *Geodermatophilus soli* as *Klenkia soli* comb. nov., *Geodermatophilus taihuensis* as *Klenkia taihuensis* comb. nov. and *Geodermatophilus terrae* as *Klenkia terrae* comb. nov. The emendation of the genus *Geodermatophilus* and the species *G. africanus, G. amargosae, G. aquaeductus, G. dictyosporus, G. nigrescens, G. normandii, G. obscurus, G. poikilotrophus, G. pulveris, G. ruber, G. sabuli, G. saharensis, G. siccatus,* and *G. telluris* are also proposed in this study.

Description of *Klenkia* gen. nov.

Klen'ki.a (N. L. fem. n. *Klenkia*, named in honor of Hans-Peter Klenk, Professor at Newcastle University (United Kingdom) in recognition of his contributions to bacterial systematics including the promotion of studies in *Geodermatophilaceae*).

Cells are motile, rod-shaped and Gram-reaction-positive. The peptidoglycan in the cell-wall contains *meso*-diaminopimelic acid. The predominant menaquinones are MK-9(H_4) and MK-9(H_0) but MK-9(H_2), MK-8(H_4) and MK-10(H_4) may also be present in minor amounts. The basic polar-lipids profile includes diphosphatidylglycerol, phosphatidylethanolamine, hydroxyphosphatidylethanolamine, phosphatidylinositol and glycophosphoinositol. In some species an unidentified glycophospholipid may be present. Phosphatidylcholine is absent. Major cellular fatty acids are iso-$C_{16:0}$ and iso-$C_{15:0}$. The basic whole-cell sugar pattern includes ribose, mannose and glucose. The presence of galactose is frequent. Rhamnose may occur in some species. The genomic G+C content is 74.0–75.0%. The type species of *Klenkia* is *Klenkia marina*, sp. nov.

Description of *Klenkia marina* sp. nov.

K. ma.ri'na. (L. fem. adj. *marina*, of the sea, marine).

Colonies are pink-colored, convex, circular and opaque with a smooth surface and an entire margin. Cells are motile, rod-shaped and Gram-reaction-positive. According to the BIOLOG System: dextrin, D-maltose, D-trehalose, D-cellobiose, sucrose, turanose, β-methyl-D-glucoside, *N*-acetyl-D-glucosamine, D-glucose, D-mannose, D-fructose, D-galactose, inosine, D-mannitol, glycerol, L-alanine, L-glutamic acid, L-pyroglutamic acid, pectin, methyl pyruvate, citric acid, α-keto-glutaric acid,

D-malic acid, L-malic acid, bromo-succinic acid, nalidixic acid, lithium chloride, potassium tellurite, α-hydroxy-butyric acid, β-hydroxy-butyric acid, α-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid and aztreonam are positive but stachyose, D-raffinose, alpha-D-lactose, D-melibiose, *N*-acetyl-D-galactosamine, *N*-acetyl-neuraminic acid, 3-*O*-methyl-D-glucose, D-fucose, L-fucose, L-rhamnose, fusidic acid, D-sorbitol, D-arabitol, D-fructose-6-phosphate, D-aspartic acid, D-serine, troleandomycin, rifamycin SV, minocycline, gelatin, L-arginine, L-aspartic acid, L-histidine, lincomycin, guanidinehydrochloride, niaproof, D-galacturonic acid, L-galactonic acid-γ-lactone, D-glucuronic acid, glucuronamide, mucic acid, quinic acid, D-saccharic acid, vancomycin, tetrazolium violet, tetrazolium blue, *p*-hydroxy-phenylacetic acid, L-lactic acid, tween 40, γ-amino-*n*-butyric acid, sodium formate and butyric acid are negative. Cell growth ranges from 20 to 30°C (optimal growth temperature is 25–30°C), from pH 6.0–8.5 (optimal range 6.5–8.0) and 0–4% NaCl. The peptidoglycan in the cell-wall contains *meso*-diaminopimelic acid as diamino acid. The whole-cell sugars are rhamnose, ribose, mannose, glucose and an unidentified sugar. The predominant menaquinones are MK-9(H_4) and MK-9(H_0). The main polar lipids are diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, glycophosphatidylinositol, an unidentified aminolipid and traces of hydroxyphosphatidylethanolamine. Cellular fatty acids consist mainly of iso-$C_{16:0}$, $C_{17:1}\omega9$ and iso-$C_{15:0}$. The type strain has a genomic G+C content of 74.4%. The genome size is 4.2 Mbp.

The INSDC accession number for the 16S rRNA gene sequences of the type strain YIM M13156^T (=DSM 45722^T = CCTCC AB 2012057^T) is LT746188. The accession number for the whole genome sequence of strain YIM M13156^T is FMUH01000001.

Description of *Klenkia brasiliensis* comb. nov.

K. bra.si.li.en'sis. (N. L. fem. adj. *brasiliensis*, referring to Brazil, the country from where the type strain was isolated).

Basonym: *Geodermatophilus brasiliensis* Bertazzo et al. (2014)

The description is as given by Bertazzo et al. (2014) with the following modification. The genomic G+C content is 74.8%. The genome size is 4.5 Mbp.

The accession number for the whole genome sequence of strain DSM 44526^T is FNCF00000000.

The type strain Tü 6233^T (= DSM 44526^T = CECT 8402^T) was isolated from soil collected in São José do Rio Preto, São Paulo (20°46′39″ S, 49°21′35″ W, altitude 530 m above mean see level), Brazil.

Description of *Klenkia soli* comb. nov.

K. so'li. (L. gen. n. *soli*, of soil)

Basonym: *Geodermatophilus soli* Jin et al. (2013)

The properties are as given in the species description by Jin et al. (2013) with the following emendation. In addition to diphosphatidylglycerol, phosphatidylethanolamine and phosphatidylinositol, the polar lipids pattern consists in phosphatidylglycerol, hydroxyphosphatidylethanolamine, an unidentified glycophospholipid and glycophosphatidylinositol (the chromatographic mobility of which is documented in Fig. 1b). Phosphatidylcholine and phosphatidylmethylethanolamine are absent. The whole-cell sugars are ribose, mannose, glucose and galactose. MK-9(H_4) is the predominant menaquinone but also contains MK-9(H_0) (as listed by Jin et al., 2013) and MK-9(H_2). The genomic G+C content is 74.2%. The genome size is 4.8 Mbp.

The accession number for the whole genome sequence of strain DSM 45843^T is FNIR00000000.

The type strain, PB34^T (=DSM 45843^T = KCTC 19880^T = JCM 17785^T), was isolated from grass soil in Korea.

Description of *Klenkia taihuensis* comb. nov.

K. tai.hu.en'sis. (N. L. fem. adj. *taihuensis*, of or pertaining to Taihu Lake, the source of the sediment from which the type strain was isolated).

Basonym: *Geodermatophilus taihuensis* Qu et al. (2013)

The properties are as given in the species description by Qu et al. (2013) with the following emendation. In addition to diphosphatidylglycerol, phosphatidylethanolamine and phosphatidylinositol, the polar lipids pattern consists of phosphatidylglycerol, hydroxyphosphatitylethanolamine and glycophosphatidylinositol (the chromatographic mobility of which is documented in **Figure 1C**). Phosphatidylcholine is absent. The whole-cell sugars are ribose, mannose, glucose and galactose. Meso-diaminopimelic acid is present. MK-9(H_4) is the predominant menaquinone but the strain also contains MK-9(H_0), MK-9(H_6) (as listed by Qu et al., 2013), MK-9(H_2) and MK-8(H_4). The genomic G+C content is 74.9%. The genome size is 4.3 Mbp.

The accession number for the whole genome sequence of strain DSM 45962^T is FOMD00000000.

The type strain is 3-wff-81T (=DSM 45962^T = CGMCC 1.12303^T = NBRC 109416^T), isolated from the superficial sediment of Taihu Lake in Jiangsu Province, China.

Description of *Klenkia terrae* comb. nov.

K. ter'ra.e. (L. gen. n. *terrae*, of the earth).

Basonym: *Geodermatophilus terrae* Jin et al. (2013)

The properties are as given in the species description by Jin et al. (2013) with the following emendation. In addition to diphosphatidylglycerol, phosphatidylethanolamine and phosphatidylinositol, the polar lipids pattern consists in phosphatidylglycerol, hydroxyphosphatitylethanolamine and glycophosphatidylinositol (the chromatographic mobility of which is documented in **Figure 1D**). Phosphatidylcholine and phosphatidylmethylethanolamine are absent. The whole-cell sugars are ribose, mannose, glucose and galactose. MK-9(H_4) is the predominant menaquinone but also contains MK-9(H_0) (as listed by Jin et al., 2013), MK-9(H_2) and MK-10(H_4).

The type strain, PB261^T (=DSM 45844^T = KCTC 19881^T = JCM 17786^T), was isolated from grass soil in Korea.

Emended Description of the Genus *Geodermatophilus* Luedemann (1968)

The properties are as given by Luedemann (1968) with the following modifications. The peptidoglycan in the whole-cell contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4) but MK-9(H_0), MK-9(H_2), MK-9(H_6), MK-8(H_4) and MK-10(H_4) may also be present in significant or minor amounts. The basic polar lipids profile involves diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine, phosphatidylinositol and an unidentified glycolipid. The presence of phosphatidylglycerol is frequent. Major cellular fatty acids are iso-$C_{16:0}$ and iso-$C_{15:0}$. The whole-cell sugar pattern frequently includes ribose, mannose, glucose and galactose as diagnostic sugar. The genomic G+C content is 74.0–76.0%.

Emended Description of *Geodermatophilus africanus* Montero-Calasanz et al. (2013e)

The properties are as given in the species description by Montero-Calasanz et al. (2013e) with the following modification. The genomic G+C content is 74.3%. The genome size is 5.5 Mbp.

The accession number for the whole genome sequences of the type strain DSM 45422^T is FNOT00000000.

Emended Description of *Geodermatophilus amargosae* Montero-Calasanz et al. (2014a)

The properties are as given in the species description by Montero-Calasanz et al. (2014a) with the following modification. The genomic G+C content is 74.2%. The genome size is 5.9 Mbp.

The accession number for the whole genome sequence of the type strain DSM 46136^T is FPBA00000000.

Emended Description of *Geodermatophilus aquaeductus* Hezbri et al. (2015c)

The properties are as given in the species description by Hezbri et al. (2015c) with the following modification. The genomic G+C content is 75.0%. The genome size is 5.4 Mbp.

The ENA accession numbers for the whole genome sequence of the type strain DSM 46834^T are FXTJ01000001 -FXTJ01000028.

Emended Description of *Geodermatophilus dictyosporus* Montero-Calasanz et al. (2015)

The properties are as given in the species description by Montero-Calasanz et al. (2015) with the following modification. The DNA G+C content is 75.3% (genome sequence). The genome size is 5.0 Mbp.

The ENA accession numbers for the whole genome sequence of the type strain DSM 43161^T are FOWE01000001 -FOWE01000022.

Emended Description of *Geodermatophilus nigrescens* Nie et al. (2012)

The properties are as given in the species description by Nie et al. (2012) with the following emendation. It grows well on GYM and GPHF media but poor on TSA and not on R2A, GEO, Luedemann and PYGV media. The following enzymatic activities according to API ZYM strips are present: alkaline phosphatase, esterase lipase, leucine arylamidase, valine arylamidase and α-glucosidase. In addition to diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine and phosphatidylcholine the polar lipids profile contains phosphatidylinositol and two unidentified glycolipids (the chromatographic mobility of which is documented in Supplementary Figure S5A). The whole-cell sugars are mannose, glucose, galactose and traces of rhamnose and ribose. Arabinose and glucosamine as listed by Nie et al. (2012) are absent. The genomic G+C content is 75.9%. The genome size is 4.7 Mbp.

The accession number for the whole genome sequences of strain DSM 45408^T is FQVX00000000.

Emended Description of *Geodermatophilus normandii* Montero-Calasanz et al. (2013b)

The properties are as given in the species description by Montero-Calasanz et al. (2013b) with the following modification. The DNA G+C content is 75.3% (genome sequence). The genome size is 4.6 Mbp.

The IMG accession number for the whole genome sequence of the type strain DSM 45417^T is 2585427554.

Emended Description of *Geodermatophilus obscurus* Luedemann (1968)

The properties are as given in the species description by Luedemann (1968) with the following emendation. The temperature range is from 15.0°C to 40.0 with an optimum range from 28°C to 37°C. pH range is 6.0–9.0 (optimal range 6.5–8.5). It grows well on GYM and regular on GPHF media but not on R2A, GEO, TSA, Luedemann and PYGV media. It degrades starch but not tyrosine, xanthine, casein, hypoxanthine. The following enzymatic activities according to API ZYM strips are present: alkaline phosphatase, esterase lipase, leucine arylamidase. Catalase positive but oxidase negative. Hydrolisis of aesculin and gelatine. The polar lipids profile consists in diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine, phosphatidylinositol and unidentified glycolipid (the chromatographic mobility of which is documented in Supplementary Figure S5B). Whole-cell sugars are ribose, xylose, mannose, glucose and galactose. *Meso*-diaminopimelic acid is present in the cell-wall. MK-9(H_4) is the predominant menaquinone, but MK-9(H_2), MK-9(H_0) and MK-8(H_4) are present as minor components. The genome size is 5.3 Mbp. The genomic G+C content is 73.9% (Ivanova et al., 2010).

Emended Description of *Geodermatophilus poikilotrophus* corrig. Montero-Calasanz et al. (2015)

The properties are as given in the species description by Montero-Calasanz et al. (2014b) with the following modification. The genomic G+C content is 74.6%. The genome size is 4.8 Mbp.

The accession number for the whole genome sequence of the type strain DSM 44209^T is FOIE00000000.

Emended Description of *Geodermatophilus pulveris* Hezbri et al. (2016a)

The properties are as given in the species description by Hezbri et al. (2016a) with the following modification. The genomic G+C content is 75.6%. The genome size is 4.4 Mbp.

The accession number for the whole genome sequence of the type strain DSM 46839^T is FZOO00000000.

Emended Description of *Geodermatophilus ruber* Zhang et al. (2011)

The properties are as given in the species description by Zhang et al. (2011) with the following emendation. It grows well on GYM, TSA and R2A media but not on GEO, Luedemann, PYGV and GPHF media. In addition to diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol and two unidentified phospholipids, the polar lipids pattern consists in phosphatidylglycerol, phosphatidylcholine and an unidentified glycolipid (the chromatographic mobility of which is documented in Supplementary Figure S5C). The whole-cell sugars are ribose and glucose. Oxidase and catalase positive. Aesculin hydrolysis present. MK-9(H_4) is the predominant menaquinone. MK-9(H_0) as listed by Zhang et al. (2011) is absent. The genomic G+C content is 74.0%. The genome size is 5.0 Mbp.

The accession number for the whole genome sequence of strain DSM 45317^T is FOSW00000000.

Emended Description of *Geodermatophilus sabuli* Hezbri et al. (2015a)

The properties are as given in the species description by Hezbri et al. (2015a) with the following modification. The genomic G+C content is 74.0%. The genome size is 5.5 Mbp.

The accession number for the whole genome sequence of the type strain DSM 46844^T is OBDO00000000.

Emended Description of *Geodermatophilus saharensis* Montero-Calasanz et al. (2013c)

The properties are as given in the species description by Montero-Calasanz et al. (2013c) with the following modification. The genomic G+C content is 75.6%. The genome size is 4.9 Mbp.

The accession number for the whole genome sequence of the type strain DSM 45423^T is FZOH00000000.

Emended Description of *Geodermatophilus siccatus* Montero-Calasanz et al. (2013f)

The properties are as given in the species description by Montero-Calasanz et al. (2013f) with the following modification. The genomic G+C content is 74.6%. The genome size is 5.2 Mbp.

The accession number for the whole genome sequence of the type strain DSM 45419^T is FNHE00000000.

Emended Description of *Geodermatophilus telluris* Montero-Calasanz et al. (2013d)

The properties are as given in the species description by Montero-Calasanz et al. (2013d) with the following modification. The genomic G+C content is 75.7%. The genome size is 4.8 Mbp.

The accession number for the whole genome sequences of the type strain DSM 45421^T is FMZF00000000.

AUTHOR CONTRIBUTIONS

MCM-C, W-JL, and MG designed the study. MCM-C, D-FZ, AY, MR, and PS performed experiments. JM-K and MG performed bioinformatics analysis. MG, TW, and NK sequenced genomes. MCM-C, PS, and MG wrote the manuscript. All authors read and approved the manuscript.

FUNDING

MCM-C was the recipient of a DSMZ postdoctoral fellowship 2013–2015. The work conducted by the Joint Genome Institute, a United States Department of Energy Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. W-JL was supported by Guangdong Province Higher Vocational Colleges and Schools Pearl River Scholar Funded Scheme (2014).

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the help of Brian J. Tindall for his guidance in the chemotaxonomic analyses and Cathrin Spröer and Bettina Sträubler (all at DSMZ, Braunschweig) for preliminary DNA:DNA hybridization analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02501/full#supplementary-material

REFERENCES

Ahrens, R., and Moll, G. (1970). Ein neues knospendes bakterium aus der ostsee. *Arch. Mikrobiol.* 70, 243–265.

Auch, A. F., von Jan, M., Klenk, H.-P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Bapteste, E., and Boucher, Y. (2009). Epistemological impacts of horizontal gene transfer on classification in microbiology. *Methods Mol. Biol.* 532, 55–72. doi: 10.1007/978-1-60327-853-9

Bertazzo, M., Montero-Calasanz, M. C., Martinez-Garcia, M., Spröer, C., Schumann, P., Kroppenstedt, R. M., et al. (2014). *Geodermatophilus brasiliensis* sp. nov., isolated from Brazilian soil. *Int. J. Syst. Evol. Microbiol.* 64, 2841–2848. doi: 10.1099/ijs.0.062844-0

Breider, S., Scheuner, C., Schumann, P., Fiebig, A., Petersen, J., Pradella, S., et al. (2014). Genome-scale data suggest reclassifications in the *Leisingera-Phaeobacter* cluster including proposals for *Sedimentitalea* gen. nov. and *Pseudophaeobacter* gen. nov. *Front. Microbiol.* 5:416. doi: 10.3389/fmicb.2014.00416

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.

Chen, I. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., et al. (2017). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 45, D507–D516. doi: 10.1093/nar/gkw929

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

Collins, M. D. (1985). "Analysis of isoprenoid quinones," in *Methods in Microbiology*, Vol. 18, ed. G. Gottschalk (London: Academic Press), 329–366.

Crawley, M. J. (2007). *The R Book*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9780470515075

Giongo, A., Favet, J., Lapanje, A., Gano, K. A., Kennedy, S., Davis-Richardson, A. G., et al. (2013). Microbial hitchhikers on intercontinental dust: high-throughput sequencing to catalogue microbes in small sand samples. *Aerobiologia* 29, 71–84. doi: 10.1007/S10453-012-9264-0

Göker, M., Cleland, D., Saunders, E., Lapidus, A., Nolan, M., Lucas, S., et al. (2011). Complete genome sequence of *Isosphaera pallida* type strain (IS1B^T). *Stand. Genomic Sci.* 4, 63–71. doi: 10.4056/sigs.1533840

Gregersen, T. (1978). Rapid method for distinction of gram-negative from positive bacteria. *Appl. Microbiol. Biotechnol.* 5, 123–127.

Gtari, M., Essoussi, I., Maaoui, R., Sghaier, H., Boujmil, R., Gury, J., et al. (2012). Contrasted resistance of stone-dwelling *Geodermatophilaceae* species to stresses known to give rise to reactive oxygen species. *FEMS Microbiol. Ecol.* 80, 566–577. doi: 10.1111/j.1574-6941.2012.01320.x

Hahnke, R. L., Meier-Kolthoff, J. P., García-López, M., Mukherjee, S., Huntemann, M., Ivanova, N. N., et al. (2016). Genome-based taxonomic classification of *Bacteroidetes*. *Front. Microbiol.* 7:2003. doi: 10.3389/fmicb.2016.02003

Hennig, W. (1965). Phylogenetic systematics. *Annu. Rev. Ecol. Syst.* 10, 97–116.

Hezbri, K., Ghodhbane-Gtari, F., Montero-Calasanz, M. C., Sghaier, H., Rohde, M., Schumann, P., et al. (2015a). *Geodermatophilus sabuli* sp. nov., a γ-radiation-resistant actinobacterium isolated from desert limestone in Sahara desert. *Int. J. Syst. Evol. Microbiol.* 65, 3365–3372. doi: 10.1099/ijsem.0.000422

Hezbri, K., Ghodhbane-Gtari, F., Montero-Calasanz, M. C., Sghaier, H., Rohde, M., Spröer, C., et al. (2015b). Description of *Geodermatophilus bullaregiensis* sp. nov. *Antonie Van Leeuwenhoek* 108, 415–425. doi: 10.1007/s10482-015-0494-3

Hezbri, K., Ghodhbane-Gtari, F., Montero-Calasanz, M. C., Sghaier, H., Rohde, M., Spröer, C., et al. (2015c). *Geodermatophilus aquaeductus* sp. nov., isolated from the ruins from Hadrian's aqueduct. *Antonie Van Leeuwenhoek* 108, 41–50. doi: 10.1007/s10482-015-0461-z

Hezbri, K., Ghodhbane-Gtari, F., Montero-Calasanz, M. C., Nouioui, I., Rohde, M., Spröer, C., et al. (2016a). *Geodermatophilus pulveris* sp. nov., a γ-radiation-resistant actinobacterium isolated from desert limestone in Sahara desert. *Int. J. Syst. Evol. Microbiol.* 66, 3828–3834. doi: 10.1099/ijsem.0.001272

Hezbri, K., Louati, M., Nouioui, I., Gtari, M., Rohde, M., Spröer, C., et al. (2016b). *Blastococcus capsensis* sp. nov., isolated from an archaeological Roman pool and emended description of the genus *Blastococcus*, *B. aggregatus*, *B. saxobsidens*, *B. jejuensis* and *B. endophyticus*. *Int. J. Syst. Bacteriol.* 66, 4864–4872. doi: 10.1099/ijsem.0.001443

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.

Ivanova, N., Sikorski, J., Jando, M., Munk, C., Lapidus, A., Glavina Del Rio, T., et al. (2010). Complete genome sequence of *Geodermatophilus obscurus* type strain (G-20T). *Stand. Genomic Sci.* 2, 158–167. doi: 10.4056/sigs.711311

Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231. doi: 10.1016/j.tig.2006.02.003

Jin, L., Lee, H. G., Kim, H. S., Ahn, C. Y., and Oh, H. M. (2013). *Geodermatophilus soli* sp. nov. and *Geodermatophilus terrae* sp. nov., two novel actinobacteria isolated from grass soil. *Int. J. Syst. Bacteriol.* 63, 2625–2629. doi: 10.1099/ijs.0.048892-0

Kämpfer, P., Busse, H.-J., Tindall, B. J., Nimtz, M., and Grün-Wollny, I. (2010). *Nonomuraea rosea* sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 1118–1124. doi: 10.1099/ijs.0.014845-0

Klenk, H.-P., and Göker, M. (2010). En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst. Appl. Microbiol.* 33, 175–182. doi: 10.1016/j.syapm.2010.03.003

Kroppenstedt, R. M. (1982). Separation of bacterial menaquinones by HPLC using reverse phase (RP18) and a silver loaded ion exchanger as stationary phases. *J. Liq. Chromatogr.* 5, 2359–2387.

Kroppenstedt, R. M., and Goodfellow, M. (2006). "The family *Thermomonosporaceae*: *Actinocorallia*, *Actinomadura*, *Spirillispora* and *Thermomonospora*," in *The Prokaryotes, Archaea and Bacteria: Firmicutes, Actinomycetes*, 3rd Edn, Vol. 3, eds M. Dworkin, S. Falkow, K. H. Schleifer, and E. Stackebrandt (New York, NY: Springer), 682–724.

Kyrpides, N. C., Hugenholtz, P., Eisen, J. A., Woyke, T., Göker, M., Parker, C. T., et al. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLOS Biol.* 12:e1001920. doi: 10.1371/journal.pbio.1001920

Lechevalier, M. P., and Lechevalier, H. A. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. *Int. J. Syst. Bacteriol.* 20, 435–443.

Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800. doi: 10.1093/molbev/msv150

Lienau, K. E., and DeSalle, R. (2009). Evidence, content and corroboration and the tree of life. *Acta Biotheor.* 57, 187–199. doi: 10.1007/s10441-008-9066-5

Luedemann, G. M. (1968). *Geodermatophilus*, a new genus of the *Dermatophilaceae* (Actinomycetales). *J. Bacteriol.* 96, 1848–1858.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013a). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Göker, M., Spröer, C., and Klenk, H.-P. (2013b). When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* 195, 413–418. doi: 10.1007/s00203-013-0888-4

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2014a). Highly parallelized inference of large genome-based phylogenies. *Concurr. Comput. Pract. Exp.* 26, 1715–1729. doi: 10.1002/cpe.3112

Meier-Kolthoff, J. P., Klenk, H.-P., and Göker, M. (2014b). Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol.* 64, 352–356. doi: 10.1099/ijs.0.056994-0

Mevs, U., Stackebrandt, E., Schumann, P., Gallikowski, C. A., and Hirsch, P. (2000). *Modestobacter multiseptatus* gen. nov., sp. nov., a budding actinomycete from soils of the Asgard Range (Transantarctic Mountains). *Int. J. Syst. Evol. Microbiol.* 50, 337–346.

Minni, D. E., Odonnell, T., Goodfellow, M., Alderson, G., Athalye, M., Schaal, A., et al. (1984). An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids. *J. Microbiol. Methods* 2, 233–241.

Montero-Calasanz, M. C., Göker, M., Broughton, W. J., Cattaneo, A., Favet, J., Pötter, G., et al. (2013a). *Geodermatophilus tzadiensis* sp. nov., a UV radiation-resistant bacterium isolated from sand of the Saharan desert. *Syst. Appl. Microbiol.* 36, 177–182. doi: 10.1016/j.syapm.2012.12.005

Montero-Calasanz, M. C., Göker, M., Pötter, G., Rohde, M., Spröer, C., Schumann, P., et al. (2012). *Geodermatophilus arenarius* sp. nov., a xerophilic actinomycete isolated from Saharan desert sand in Chad. *Extremophiles* 16, 903–909. doi: 10.1007/s00792-012-0486-4

Montero-Calasanz, M. C., Göker, M., Pötter, G., Rohde, M., Spröer, C., Schumann, P., et al. (2013b). *Geodermatophilus normandii* sp. nov., isolated from Saharan desert sand. *Int. J. Syst. Evol. Microbiol.* 63, 3437–3443. doi: 10.1099/ijs.0.051201-0

Montero-Calasanz, M. C., Göker, M., Pötter, G., Rohde, M., Spröer, C., Schumann, P., et al. (2013c). *Geodermatophilus saharensis* sp. nov., isolated from sand of the Saharan desert in Chad. *Arch. Microbiol.* 195, 153–159. doi: 10.1007/s00203-012-0860-8

Montero-Calasanz, M. C., Göker, M., Pötter, G., Rohde, M., Spröer, C., Schumann, P., et al. (2013d). *Geodermatophilus telluris* sp. nov., a novel actinomycete isolated from Saharan desert sand in Chad. *Int. J. Syst. Evol. Microbiol.* 63, 2254–2259. doi: 10.1099/ijs.0.046888-0

Montero-Calasanz, M. C., Göker, M., Pötter, G., Rohde, M., Spröer, C., Schumann, P., et al. (2013e). *Geodermatophilus africanus* sp. nov., an halotolerant actinomycete isolated from Saharan desert sand. *Antonie Van Leeuwenhoek* 104, 207–216. doi: 10.1007/s10482-013-9939-8

Montero-Calasanz, M. C., Göker, M., Rohde, M., Schumann, P., Pötter, G., Spröer, C., et al. (2013f). *Geodermatophilus siccatus* sp. nov., isolated from arid sand of the Saharan desert in Chad. *Antonie Van Leeuwenhoek* 10, 449–456. doi: 10.1007/s10482-012-9824-x

Montero-Calasanz, M. C., Göker, M., Rohde, M., Spröer, C., Schumann, P., Mayilraj, S., et al. (2014a). Description of *Geodermatophilus amargosae* sp. nov., to accommodate the not validly named *Geodermatophilus obscurus* subsp. *amargosae* (Luedemann, 1968). *Curr. Microbiol.* 68, 365–371. doi: 10.1007/s00284-013-0479-2

Montero-Calasanz, M. C., Hezbri, K., Göker, M., Sghaier, H., Rohde, M., Spröer, C., et al. (2015). Description of gamma-radiation-resistant *Geodermatophilus dictyosporus* sp. nov., to accommodate the not validly named *Geodermatophilus obscurus* subsp. *dictyosporus* (Luedemann, 1968). *Extremophiles* 19, 77–85.

Montero-Calasanz, M. C., Hofner, B., Göker, M., Rohde, M., Spröer, C., Hezbri, K., et al. (2014b). *Geodermatophilus poikilotrophi* sp. nov., a multi-tolerant actinomycete isolated from dolomitic marble. *Biomed Res. Int.* 2014:914767. doi: 10.1155/2014/914767

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., et al. (2017). Genomes online database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 45, D446–D456. doi: 10.1093/nar/gkw992

Neilson, J. W., Legatzki, A., Wing, R. A., Quade, J., Tian, F., Ortiz, M., et al. (2012). Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* 16, 553–566. doi: 10.1007/s00792-012-0454-z

Nie, G. X., Ming, H., Li, S., Zhou, E. M., Cheng, J., Yu, T. T., et al. (2012). *Geodermatophilus nigrescens* sp. nov., isolated from a dry-hot valley. *Antonie Van Leeuwenhoek.* 101, 811–817. doi: 10.1007/s10482-012-9696-0

Normand, P. (2006). *Geodermatophilaceae* fam. nov., a formal description. *Int. J. Syst. Evol. Microbiol.* 56, 2277–2278.

Normand, P., and Benson, D. R. (2012). "Genus I. *Geodermatophilus* Luedemann, 1968. 1994," in *Bergey's Manual of Systematic Bacteriology: The Actinobacteria Part 1*, 2nd Edn, Vol. 5, eds M. Goodfellow, P. Kämpfer, H. J. Busse, M. E. Trujillo, K. I. Suzuki, W. Ludwig, et al. (New York, NY: Springer), 528–530.

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., et al. (1996). Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family *Frankiaceae*. *Int. J. Syst. Bacteriol.* 46, 1–9.

Nouioui, I., Göker, M., Carro, L., Montero-Calasanz, M. C., Rohde, M., Woyke, T., et al. (2017). High quality draft genome of *Nakamurella lactea* type strain, a rock actinobacterium, and emended description of *Nakamurella lactea*. *Stand. Genomic Sci.* 12:4. doi: 10.1186/s40793-016-0216-0

Parte, A. C. (2014). LPSN – list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.* 42, D613–D616.

Pelczar, M. J. Jr. (ed.). (1957). *Manual of Microbiological Methods*. New York, NY: McGraw-Hill Book Company.

Qu, J.-H., Hui, M., Qu, J.-Y., Wang, F.-F., Li, H.-F., Hu, Y.-S., et al. (2013). *Geodermatophilus taihuensis* sp. nov., isolated from the interfacial sediment of a eutrophic lake. *Int. J. Syst. Evol. Microbiol.* 63, 4108–4112. doi: 10.1099/ijs.0.049460-0

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rainey, F. A., Ward-Rainey, N., Kroppenstedt, R. M., and Stackebrandt, E. (1996). The genus *Nocardiopsis* represents a phylogenetically coherent taxon and a distinct actinomycete lineage: proposal of *Nocardiopsiaceae* fam. nov. *Int. J. Syst. Bacteriol.* 46, 28–96.

Ramasamy, D., Mishra, A. K., Lagier, J. C., Padhmanabhan, R., Rossi, M., Sentausa, E., et al. (2014). A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *J. Syst. Bacteriol.* 64, 384–391. doi: 10.1099/ijs.0.057091-0

Sasser, M. (1990). Identification of bacteria by gas chromatography of cellular fatty acids. *USFCC Newsl.* 20:16.

Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A. M., and Normand, P. (2014). Phylogeny of the class *Actinobacteria* revisited in the light of complete genomes. The orders '*Frankiales*' and *Micrococcales* should be split into coherent entities: proposal of *Frankiales* ord. nov., *Geodermatophilales* ord. nov., *Acidothermales* ord. nov. and *Nakamurellales* ord. nov. *Int. J. Syst. Evol. Microbiol.* 64(Pt 11), 3821–2332. doi: 10.1099/ijs.0.063966-0

Sghaier, H., Hezbri, K., Gohdhbane-Gtari, F., Pujic, P., Sen, A., Daffonchio, D., et al. (2016). Stone-dwelling actinobacteria *Blastococcus saxobsidens*, *Modestobacter marinus* and *Geodermatophilus obscurus* proteogenomes. *ISME J.* 10, 21–29. doi: 10.1038/ismej.2015.108

Siddall, M. E. (2010). Unringing a bell: metazoan phylogenomics and the partition bootstrap. *Cladistics* 26, 444–452. doi: 10.1111/j.1096-0031.2009.00295.x

Siddall, M. E., and Whiting, M. F. (1999). Long-branch abstractions. *Cladistics* 15, 9–24.

Simon, M., Scheuner, C., Meier-Kolthoff, J. P., Brinkhoff, T., Wagner-Döbler, I., Ulbrich, M., et al. (2017). Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 11, 1483–1499. doi: 10.1038/ismej.2016.198

Skerman, V. B. D., McGowan, V., and Sneath, P. H. A. (eds). (1980). *Approved Lists of Bacterial Names (Amended)*. Washington, DC: ASM Press.

Stackebrandt, E., Rainey, F. A., and Ward-Rainey, N. L. (1997). Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int. J. Syst. Bacteriol.* 47, 479–491.

Staneck, J. L., and Roberts, G. D. (1974). Simplified approach to identification of aerobic actinomycetes by thin-layer chromatography. *Appl. Microbiol.* 28, 226–231.

Sutcliffe, I. C. (2015). Challenging the anthropocentric emphasis on phenotypic testing in prokaryotic species descriptions: Rip it up and start again. *Front. Genet.* 6:218. doi: 10.3389/fgene.2015.00218

Sutcliffe, I. C., Trujillo, M. E., and Goodfellow, M. (2012). A call for arms to systematists: revitalising the purpose and practices underpinning the description of novel microbial taxa. *Antonie Van Leeuwenhoek* 101, 13–20. doi: 10.1007/s10482-011-9644-0

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542.

Tamura, T., Hayakawa, M., and Hatano, K. (1998). A new genus of the order *Actinomycetales*, *Cryptosporangium* gen. nov., with descriptions of *Cryptosporangium arvum* sp. nov. and *Cryptosporangium japonicum* sp. nov. *Int. J. Syst. Bacteriol.* 48, 995–1005.

Taylor, D. J., and Piel, W. H. (2004). An assessment of accuracy, error, and conflict with support values from genome-scale phylogenetic data. *Mol. Biol. Evol.* 21, 1534–1537. doi: 10.1093/molbev/msh156

Thompson, C. C., Amaral, G. R., Campeão, M., Edwards, R. A., Polz, M. F., Dutilh, B. E., et al. (2015). Microbial taxonomy in the post-genomic era: rebuilding from scratch? *Arch. Microbiol.* 197, 359–370. doi: 10.1007/s00203-014-1071-2

To, T. H., Jung, M., Lycett, S., and Gascuel, O. (2015). Fast dating using least-squares criteria and algorithms. *Syst. Biol.* 65, 82–97. doi: 10.1093/sysbio/syv068

Urzì, C., La Cono, V., and Stackebrandt, E. (2004). Design and application of two oligonucleotide probes for the identification of *Geodermatophilaceae* strains using fluorescence *in situ* hybridization (FISH). *Environ. Microbiol.* 6, 678–685.

Vaas, L. A. I., Sikorski, J., Hofer, B., Fiebig, A., Buddruhs, N., Klenk, H.-P., et al. (2013). opm: an R package for analysing OmniLog phenotype microarray data. *Bioinformatics* 29, 1823–1824. doi: 10.1093/bioinformatics/btt291

Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M., and Klenk, H.-P. (2012). Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLOS ONE* 7:e34846. doi: 10.1371/journal.pone.0034846

Vandamme, P., and Peeters, C. (2014). Time to revisit polyphasic taxonomy. *Antonie Van Leeuwenhoek* 106, 57–65. doi: 10.1007/s10482-014-0148-x

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev* .60, 407–438.

Vences, M., Guayasamin, J. M., Miralles, A., and De La Riva, I. (2013). To name or not to name: criteria to promote economy of change in Linnaean classification schemes. *Zootaxa* 3636, 201–244. doi: 10.11646/zootaxa.3636.2.1

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsly, M. I., et al. (1987). Report of the Ad Hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37, 463–464. doi: 10.1099/00207713-37-4-463

Wiley, E. O., and Lieberman, B. B. S. (2011). *Phylogenetics: Theory and Practice of Phylogenetic Systematics*, 2nd Edn. Hoboken, NJ: John Wiley & Sons. doi: 10.2307/3280934

Zhang, Y.-Q., Chen, J., Liu, H.-Y., Zhang, Y.-Q., Li, W.-J., and Yu, L.-Y. (2011). *Geodermatophilus ruber* sp. nov., isolated from rhizosphere soil of a medicinal plant. *Int. J. Syst. Evol. Microbiol.* 61, 190–193. doi: 10.1099/ijs.0.020610-0

Zhi, X. Y., Li, W. J., and Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *Actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int. J. Syst. Evol. Microbiol.* 59, 589–608. doi: 10.1099/ijs.0.65780-0

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 © 2017 Montero-Calasanz, Meier-Kolthoff, Zhang, Yaramis, Rohde, Woyke, Kyrpides, Schumann, Li and Göker. 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 February 2018
doi: 10.3389/fmicb.2018.00067

Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera

Radhey S. Gupta*, Brian Lo and Jeen Son

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, CA, Canada

OPEN ACCESS

Edited by:
Antonio Ventosa,
Universidad de Sevilla, Spain

Reviewed by:
Stephanus Nicolaas Venter,
University of Pretoria, South Africa
Alice Rebecca Wattam,
Virginia Tech, United States

****Correspondence:***
Radhey S. Gupta
gupta@mcmaster.ca

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

Received: *16 September 2017*
Accepted: *11 January 2018*
Published: *13 February 2018*

Citation:
Gupta RS, Lo B and Son J (2018) Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus Mycobacterium into an Emended Genus Mycobacterium and Four Novel Genera. Front. Microbiol. 9:67. doi: 10.3389/fmicb.2018.00067

The genus *Mycobacterium* contains 188 species including several major human pathogens as well as numerous other environmental species. We report here comprehensive phylogenomics and comparative genomic analyses on 150 genomes of *Mycobacterium* species to understand their interrelationships. Phylogenetic trees were constructed for the 150 species based on 1941 core proteins for the genus *Mycobacterium*, 136 core proteins for the phylum Actinobacteria and 8 other conserved proteins. Additionally, the overall genome similarity amongst the *Mycobacterium* species was determined based on average amino acid identity of the conserved protein families. The results from these analyses consistently support the existence of five distinct monophyletic groups within the genus *Mycobacterium* at the highest level, which are designated as the "*Tuberculosis-Simiae,*" "*Terrae,*" "*Triviale,*" "*Fortuitum-Vaccae,*" and "*Abscessus-Chelonae*" clades. Some of these clades have also been observed in earlier phylogenetic studies. Of these clades, the "*Abscessus-Chelonae*" clade forms the deepest branching lineage and does not form a monophyletic grouping with the "*Fortuitum-Vaccae*" clade of fast-growing species. In parallel, our comparative analyses of proteins from mycobacterial genomes have identified 172 molecular signatures in the form of conserved signature indels and conserved signature proteins, which are uniquely shared by either all *Mycobacterium* species or by members of the five identified clades. The identified molecular signatures (or synapomorphies) provide strong independent evidence for the monophyly of the genus *Mycobacterium* and the five described clades and they provide reliable means for the demarcation of these clades and for their diagnostics. Based on the results of our comprehensive phylogenomic analyses and numerous identified molecular signatures, which consistently and strongly support the division of known mycobacterial species into the five described clades, we propose here division of the genus *Mycobacterium* into an emended genus *Mycobacterium* encompassing the "*Tuberculosis-Simiae*" clade, which includes all of the major human

pathogens, and four novel genera viz. *Mycolicibacterium* gen. nov., *Mycolicibacter* gen. nov., *Mycolicibacillus* gen. nov. and *Mycobacteroides* gen. nov. corresponding to the "*Fortuitum-Vaccae,*" "*Terrae,*" "*Triviale,*" and "*Abscessus-Chelonae*" clades, respectively. With the division of mycobacterial species into these five distinct groups, attention can now be focused on unique genetic and molecular characteristics that differentiate members of these groups.

Keywords: *Mycobacterium* classification, slow-growing and fast-growing mycobacteria, conserved signature indels and signature proteins, phylogenomic analysis, fortuitum-vaccae clade, abscessus-chelonae clade, terrae clade, triviale clade

INTRODUCTION

The genus *Mycobacterium* encompasses a large group of Gram-positive, rod-shaped, acid-fast organisms in the phylum Actinobacteria (Hartmans et al., 2006; Gao and Gupta, 2012; Magee and Ward, 2012). Many members are well-known human pathogens, most notably *Mycobacterium tuberculosis* and *Mycobacterium leprae* are causative agents of tuberculosis and leprosy, respectively (Medjahed et al., 2010; Magee and Ward, 2012; Lory, 2014). In addition, *Mycobacterium* species are found to inhabit a diverse range of environments including water bodies, soil, and metalworking fluids (Hartmans et al., 2006; Brzostek et al., 2009; Falkinham, 2009; Tortoli, 2012). At the time of writing, the genus *Mycobacterium* consists of 188 species with validly published names (www.namesforlife.com) (Parte, 2014). In view of the large numbers of both clinically important as well as environmental species present in a single genus, an understanding of the relationships between these organisms is of much importance (Gao and Gupta, 2012; Magee and Ward, 2012; Tortoli, 2012; Lory, 2014; Fedrizzi et al., 2017). Current understanding of the relationships within the genus *Mycobacterium* is primarily based on analysis of the 16S rRNA gene sequences and other physical and chemotaxonomic characteristics of the species (Runyon, 1965; Rogall et al., 1990; Stahl and Urbance, 1990; Goodfellow and Magee, 1998; Hartmans et al., 2006; Magee and Ward, 2012). Besides the 16S rRNA, the relationships among the mycobacterial species has also been examined using the 16S-23S spacer sequences (Roth et al., 1998) and several housekeeping genes including *hsp65* (Kim et al., 2005; Tortoli et al., 2015), *gyrB* (Kasai et al., 2000), *rpoB* (Tortoli, 2012), and *gyrA* (Guillemin et al., 1995). A number of studies have also been performed on a limited number of mycobacterial species using multilocus sequence analysis based on concatenated sequences of nucleotides or amino acid fragments from several gene sequences viz. 16S rRNA, *rpoB*, and *hsp65* (Kim and Shin, 2017); 16S rRNA, *hsp65*, *sodA*, *recA, rpoB* (Adékambi and Drancourt, 2004) and *hsp65*, *tuf*, *rpoB*, *smpB*, 16S rRNA, *sodA*, tmRNA (Mignard and Flandrois, 2008). The results of these studies have provided useful insights into the relationships between members of the genus *Mycobacterium*.

An important difference observed among the mycobacterial species very early was the differences in their growth rates (Tsukamura, 1967a; Wayne and Kubica, 1986; Magee and Ward, 2012). Based on their rates of growth, *Mycobacterium* species, in general, can be roughly divided into two groups; one group consists of slow-growing bacteria (i.e., requiring *more* than 7 days to form colonies), while the second group is comprised of rapid-growing bacteria which require <7 days to form colonies (Tsukamura, 1967a; Wayne and Kubica, 1986; Magee and Ward, 2012; Lory, 2014). The clades encompassing most of the slow-growing mycobacteria also branches distinctly from the fast-growing species in the 16S rRNA trees (Rogall et al., 1990; Stahl and Urbance, 1990; Goodfellow and Magee, 1998), and also in phylogenetic trees based on some other genes/proteins sequences (Adékambi and Drancourt, 2004; Kim et al., 2005; Adékambi et al., 2006a; Mignard and Flandrois, 2008; Tortoli, 2012; Tortoli et al., 2015). Although a broad separation of the slow-growing mycobacteria from the rapid-growing species is generally supported, the reliability of the methods used to discern these two groups, particularly the cohesiveness of the rapid-growing mycobacteria, remains of concern (Magee and Ward, 2012; Tortoli, 2012). Recent studies have also identified some distinct groupings within the slow- or rapid-growing mycobacteria. For example, a clade consisting of *Mycobacterium terrae* and its closely related members, which exhibits slow to intermediate rate of growth, can be differentiated from other slow-growing members by a characteristic 14 nt insert in the helix 18 of 16S rRNA gene and by means of phylogenetic analysis (Mignard and Flandrois, 2008; Kim et al., 2012; Tortoli, 2012; Tortoli et al., 2013; Ngeow et al., 2015; Vasireddy et al., 2016). Another clade of mycobacterial species closely related to *Mycobacterium abscessus*, can also be differentiated from other rapid-growing members based on phylogenetic branching and unique pathogenicity profile of its members (Adékambi and Drancourt, 2004; Medjahed et al., 2010; Tortoli, 2012; Wee et al., 2017). In light of the increased awareness of the diversity that exists within the mycobacterial species as well the clinical importance of many of the members from this genus, the need for more robust methods of delineation of different groups that exists within this important group of bacteria is warranted (Fedrizzi et al., 2017).

Due to rapid advances in genome sequencing technology, genome sequences for 150 members from the genus *Mycobacterium* are now publicly available in the NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/). The analysis of whole genome sequences allows for construction of more robust phylogenetic trees providing greater resolution in identifying the relationships at various taxonomic levels (Wu et al., 2009; Segata et al., 2013; Gupta et al., 2015; Adeolu et al., 2016). A number of recent studies have reported phylogenomic

analyses based on large datasets of core genes/proteins from the genomes of 28–47 *Mycobacterium* species in order to elucidate their relationships (Prasanna and Mehra, 2013; Wang et al., 2015; Fedrizzi et al., 2017; Wee et al., 2017). Based on genome sequences, the genomic relatedness among the organisms can also be determined and this approach is now widely applied in taxonomic studies (Konstantinidis and Tiedje, 2005; Thompson et al., 2013; Qin et al., 2014). In addition, the genome sequences provide a unique resource for comparative genomic studies in identifying molecular markers or signatures that are specifically shared by an evolutionarily related group of organisms and are useful in the demarcation of different taxa and for understanding interrelationships (Gao and Gupta, 2012; Gupta, 2014, 2016a; Adeolu et al., 2016). Of the two types of molecular markers that have proven particularly useful for evolutionary/taxonomic studies, conserved signature indels (CSIs) are amino acid insertions or deletions of fixed lengths that are present at a specific position within a conserved region in an evolutionarily related group of species (Gupta, 2014, 2016b; Naushad et al., 2014). Likewise, conserved signature proteins (CSPs) are proteins, whose homologs are exclusively found in a related-group of organisms (Gao et al., 2006; Gao and Gupta, 2012; Gupta et al., 2015; Gupta, 2016b). The presence of these clade-specific marker gene sequences (or synapomorphies) is most parsimoniously accounted by their initial introduction in a common ancestor of the group followed by vertical inheritance (Gupta, 1998, 2016b; Gao and Gupta, 2012; Naushad et al., 2014).

To reliably understand the relationships within the genus *Mycobacterium*, we have carried out comprehensive phylogenomic and comparative genomic studies on 150 mycobacterial species, whose genome sequences are now available. Based on genome sequences, robust phylogenetic trees have been constructed based on different large datasets of concatenated protein sequences including two trees based on 1941 and 136 core proteins for the genus *Mycobacterium* and the phylum Actinobacteria, respectively. Based on genome sequences, the pairwise average amino acid identity (AAI) was also determined for the mycobacterial species. Lastly, our detailed comparative genomic studies on mycobacterial genomes have identified 172 highly specific molecular markers in the forms of CSIs and CSPs, which are either uniquely shared by all members of the genus *Mycobacterium* or for a number of distinct clades within this genus at multiple phylogenetic levels. Based on the results of these comprehensive analyses, it is now possible to reliably divide the species from the genus *Mycobacterium* into five main monophyletic clades, which are referred to here as the "*Tuberculosis-Simiae*" clade, the "*Terrae*" clade, the "*Triviale*" clade, the "*Fortuitum-Vaccae*" clade, and the "*Abscessus-Chelonae*" clade. Based on the large body of evidence presented here which consistently and strongly supports the existence of these five clades, a proposal is made here to divide the genus *Mycobacterium* into an emended genus *Mycobacterium* encompassing the members of the "*Tuberculosis-Simiae*" clade and four new genera *Mycolicibacter* gen. nov. ("*Terrae*" clade), *Mycolicibacillus* gen. nov. ("*Triviale*" clade), *Mycolicibacterium* gen. nov. ("*Fortuitum-Vaccae*" clade), and *Mycobacteroides* gen. nov. ("*Abscessus-Chelonae*" clade).

METHODS

Phylogenetic and Genomic Analyses of the Genus *Mycobacterium*

Phylogenetic trees were constructed for 150 members of the genus *Mycobacterium* whose genomes are now sequenced (some characteristics of these genomes are listed in Supplementary Table 1) and six members from the order *Corynebacteriales* (viz. *Corynebacterium diphtheriae* NCTC 11397, *Gordonia bronchialis* DSM 43247, *Nocardia farcinica* NCTC 11134, *Rhodococcus erythropolis* PR4, *Segniliparus rotundus* DSM 44985 and *Tsukamurella paurometabola* DSM 20162), which served as outgroups. The first of these trees was based on 1941 core proteins from the genomes of *Mycobacterium* species and its construction was carried out by using a software pipeline (Adeolu et al., 2016). Briefly, the CD-HIT program was used (Li and Godzik, 2006; Fu et al., 2012) to identify protein families sharing a minimum of 50% in sequence identity and sequence length and which were found in at least 80% of the input genomes. The Clustal Omega (Sievers et al., 2011) algorithm was used to generate multiple sequence alignment (MSA) of these protein families. The aligned protein families were trimmed with TrimAl (Capella-Gutierrez et al., 2009) to remove poorly aligned regions (Talavera and Castresana, 2007) before concatenation to the other core proteins. The concatenated sequence alignment of 1941 core proteins consisted of 624,360 aligned amino acids. Another comprehensive phylogenetic tree was constructed based on concatenated sequences for 136 proteins, which comprise the phyloeco markers set for the phylum Actinobacteria (Wang and Wu, 2013). Information regarding these proteins is provided in Supplementary Table 2. The profile Hidden Markov Models of these protein families were used for the identification of members of these protein families in the input genomes using HMMer 3.1 (Eddy, 2011). The sequence alignments were trimmed using TrimAl (Capella-Gutierrez et al., 2009) before their concatenation into a single file. The combined sequence from the phyloeco set of proteins consisted of a total of 44,976 aligned amino acids. Maximum likelihood (ML) trees based on both these sequence alignments were constructed using the Whelan and Goldman model of protein sequence evolution (Whelan and Goldman, 2001) in FastTree 2 (Price et al., 2010) and the Le and Gascuel model of protein sequence evolution (Le and Gascuel, 2008) in RAxML 8 (Stamatakis, 2014). Optimization of the robustness of the tree was completed by conducting SH tests (Guindon et al., 2010) in RAxML 8 (Stamatakis, 2014). The identification of the conserved protein families and the construction of phylogenetic trees were completed using an internal software pipeline (Adeolu et al., 2016).

In addition to these two comprehensive trees, another phylogenetic tree was constructed based on concatenated sequences for 8 conserved housekeeping proteins (viz. RpoA, RpoB, RpoC, GyrA, GyrB, Hsp65, EF-Tu and RecA). After removal of non-conserved regions, the concatenated sequence alignment in this case consisted of 6052 aligned amino acids. A maximum likelihood phylogenetic tree based on this sequence was constructed as described above.

The sequence alignments of the 1941 core proteins identified by the above methods were also used to measure genome relatedness. Using the amino acid sequences from these conserved protein families, the amino acid sequence identity between each pair of *Mycobacterium* genomes was calculated (Thompson et al., 2013).

Information regarding branching of all type species from the genus *Mycobacterium* in a tree based on 16S rRNA sequences was obtained from the SILVA All Species Tree of Life Project 128 (Quast et al., 2013).

Identification of Conserved Signature Indels (CSIs)

The identification of CSIs was carried out as described in earlier work (Gao and Gupta, 2005; Bhandari et al., 2012; Gupta, 2014; Naushad et al., 2014; Sawana et al., 2014). All annotated proteins from the genomes of *M. tuberculosis* H37Rv and *M. sinense* JDM601 were used in these analyses. BLASTp (Altschul et al., 1997) searches were conducted on all protein sequences >100 amino acids in length against the NCBI non-redundant (nr) database. Multiple sequence alignments were generated by obtaining 15–25 homologs from diverse *Mycobacterium* species and 8–10 homologs from other groups of bacteria. The alignments were visually inspected for sequence gaps of fixed lengths which were flanked on both sides by at least 5 conserved amino acids in the neighboring 30–40 amino acids, and appeared to be shared by either some or all mycobacterial homologs. Query sequences encompassing the potential indel and flanking regions (60–100 amino acids long) were collected and subjected to a more detailed BLASTp search (500 or more hits) to determine the group specificities of the observed indels. Signature files for all CSIs of interest were created using SIG_CREATE and SIG_STYLE programs in the GLEANS software package (available on Gleans.net). Unless otherwise noted, the described CSIs are specific for the indicated groups of species.

Identification of Conserved Signature Proteins (CSPs)

The identification of conserved signature proteins was carried out using the protocol described in earlier work (Gao et al., 2006; Adeolu and Gupta, 2014; Naushad et al., 2014). BLASTp (Altschul et al., 1997) searches were conducted on all sequenced proteins from the genomes of *M. tuberculosis* H37Rv, *M. aurum* (LSHTM), *M. sinense* JDM601 (Zhang et al., 2011), *M. triviale* DSM 44153 (Fedrizzi et al., 2017), and *M. abscessus* ATCC 19977 (Ripoll et al., 2009) against the NCBI nr database. Proteins of interest were those where either all significant hits were limited to the genus *Mycobacterium* or the indicated groups/clades of mycobacteria, or where a large increase in E value was observed from the last hit belonging to these groups and the first hit from any other bacteria, and the *E*-values for the latter hits were $>1e^{-3}$ (Gao et al., 2006; Gao and Gupta, 2012; Naushad et al., 2014). However, in some cases, a few proteins where an isolated significant hit from an unrelated group of bacteria was observed were also retained as CSPs specific for the group of interest.

RESULTS

Phylogenomic Analysis of the Genus *Mycobacterium*

In the present work, two comprehensive phylogenomic trees were constructed based on the genome sequences of 150 *Mycobacterium* species. The first of these trees was a core genome tree of 1941 proteins, whose homologs are present in at least 80% of the input mycobacterial genomes as well as the outgroup species. The second genome sequence tree was based on 136 proteins, which are part of the phyloeco set for the phylum Actinobacteria. The trimmed concatenated sequence alignments for the two sets of core proteins, which were employed for phylogenetic analyses, consisted of 624,360 and 44,976 aligned amino acids, respectively. Although phylogenetic trees based on core genes/proteins for mycobacterial species have also been constructed in earlier studies (Prasanna and Mehra, 2013; Fedrizzi et al., 2017; Wee et al., 2017), they were based only on a small number (between 28 and 47) of *Mycobacterium* species. In contrast, the trees produced in this work include information for ∼80% (150 of the 188) of all known mycobacterial species and thus constitute the most comprehensive phylogenetic trees constructed for the genus *Mycobacterium*. In addition to the two core genome protein trees, a maximum-likelihood tree was also constructed based on concatenated sequences of 8 conserved housekeeping proteins.

The ML trees based on the core proteins from mycobacterial genomes and for the phylum Actinobacteria are shown in **Figures 1A,B**, respectively. The tree based on the 8 conserved proteins is provided as Supplementary Figure 1. In all of these phylogenetic trees, which were rooted using the sequences from the *Corynebacteriales* species, nearly all of the observed nodes were supported with high (100%) bootstrap scores or SH-values. Further, the majority of the interrelationships among the *Mycobacterium* species were highly similar and consistent in all constructed trees. In all of these trees, members of the genus *Mycobacterium* consistently grouped into four main clades and a clade consisting of the *M. triviale—M. koreense*, as indicated in **Figure 1**. Three of these clades are comprised of the slow-growing species, whereas the other two clades are mostly made up of the fast-growing species. Of the two clades of fast-growing species, the first clade referred to as the "*Abscessus-Chelonae*" clade, forms the earliest branching lineage within the genus *Mycobacterium.* The second clade of the fast-growing species designated as the "*Fortuitum-Vaccae*" clade, encompasses most of the other fast-growing species including those related to *M. fortuitum, M. vaccae, M. parafortuitum,* and *M. mucogenicum* (Hartmans et al., 2006; Magee and Ward, 2012; Lory, 2014). Of the three clades of slow-growing mycobacteria, the clade designated as "*Tuberculosis-Simiae,*" encompasses most of the clinically important *Mycobacterium* species including those related to *M. tuberculosis, M. avium, M. gordonae, M, kansasii* and *M. simiae* (Magee and Ward, 2012). The two other clades

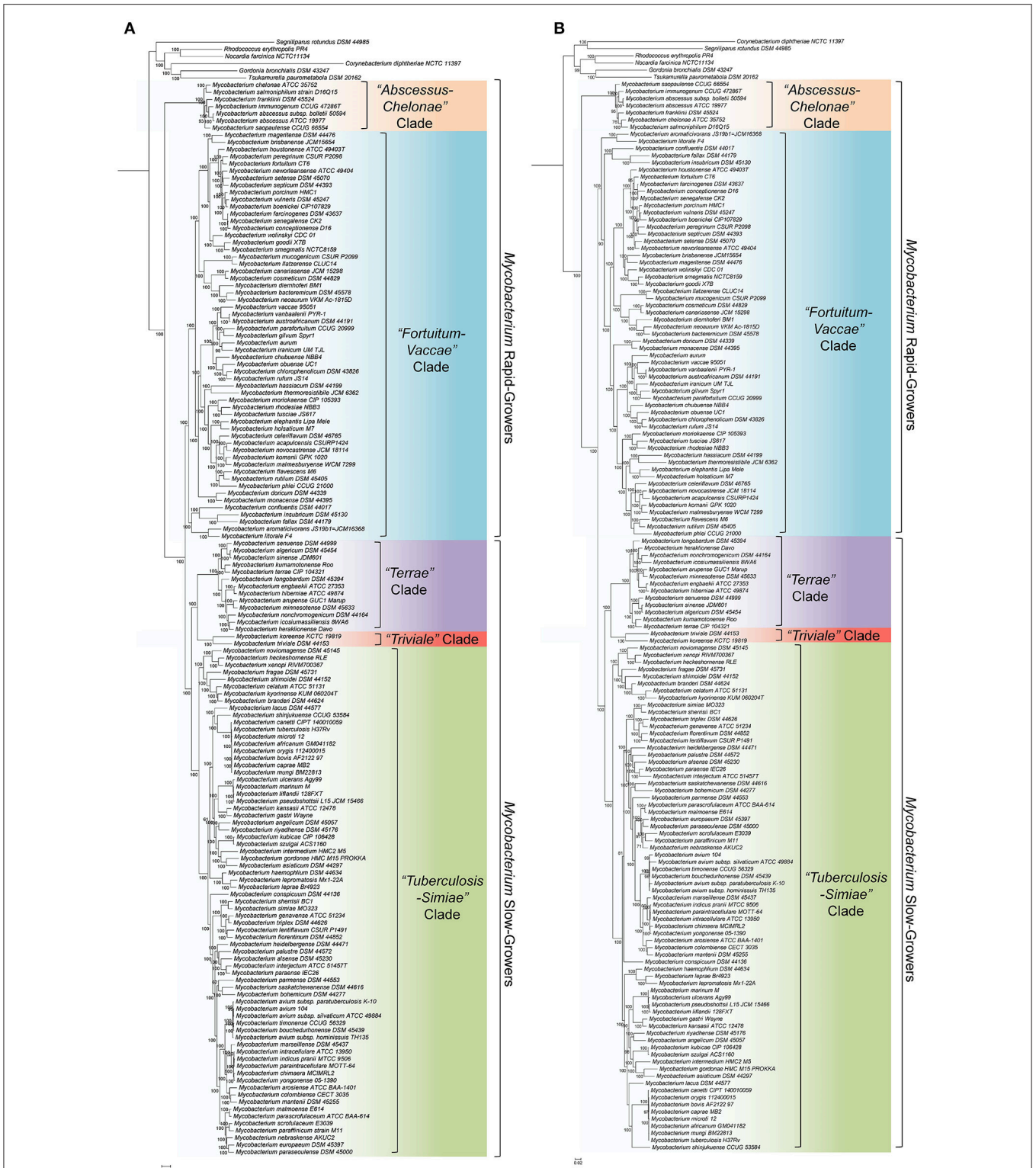


FIGURE 1 | (A) Maximum-likelihood phylogenetic tree for 150 *Mycobacterium* species based on the concatenated sequence of 1941 core proteins from the genus *Mycobacterium*. **(B)** A maximum-likelihood phylogenetic tree based on the 136 proteins consistuting the phyloeco set for the phylum Actinobacteria. Both of these trees were rooted using the sequences from the *Corynebacteriales* species. Trees were constructed as described in the Methods section. SH-like statistical support values and the bootstrap value are marked on the nodes. The major clades as well as the clusters of slow-growing and fast-growing *Mycobacterium* species are labeled. Some slow-growing species, which branched within the rapid-growing species are marked with*.

of the slow-growing species, often referred to as part of the "*M. terrae* complex," group together and they form a sister clade to the "*Tuberculosis-Simiae*" clade. Of the two clades which form the "*M. terrae* complex," most of the species closely related to *M. terrae* are part of a clade that is designated here as the "*Terrae*" clade (Magee and Ward, 2012; Tortoli, 2012; Ngeow et al., 2015). Adjacent to the "*Terrae*" clade, the species *M. koreense* and *M. triviale* form a distinct clade (designated here as the "*Triviale*" clade), which is separated from members of the "*Terrae*" clade by a long branch. It is important to note that in the phylogenetic trees shown in **Figure 1**, the two clades of fast-growing species do not form a monophyletic grouping, whereas the clades corresponding to the slow-growing mycobacteria group together and form a monophyletic lineage.

We have also compared the relationships observed in the aforementioned phylogenetic trees with the relationships observed in a tree based on 16S rRNA gene sequences, which was extracted from the SILVA Tree of Life Project 128 (Yarza et al., 2008; Quast et al., 2013). This tree is shown in Supplementary Figure 2 with the analogous groups labeled. Overall, in concordance with the core protein-based phylogenetic trees and the tree based on 8 conserved proteins, the slow-growing mycobacterial species corresponding to the "*Tuberculosis-Simiae*" clade formed a distinct clade in the 16S rRNA tree. The species corresponding to the "*Terrae*" clade also branched in the immediate proximity of the "*Tuberculosis-Simiae*" clade, with members of the "*Triviale*" clade (viz. *M. triviale, M. koreense,* and *M. parakoreense*) forming a deeper-branching lineage. However, in contrast to the different trees based on protein sequences, the rapid-growing *Mycobacterium* species exhibited extensive polyphyly and their interrelationships were poorly resolved. In particular, the members of the "*Abscessus-Chelonae*" clade formed a monophyletic lineage within the other rapid-growing *Mycobacterium* species, whereas the relationships among the other rapid growing species were difficult to discern.

Genome Relatedness of the Members of the Genus *Mycobacterium*

Based on genome sequences, the average amino acid identity between different species can be calculated to determine the overall genome relatedness of the species (Konstantinidis and Tiedje, 2007; Richter and Rossello-Mora, 2009; Thompson et al., 2013; Qin et al., 2014; Yarza et al., 2014). Pairwise amino acid identity was calculated based on the conserved protein families between each genome used in the analysis and the results of these analyses are presented in the form of a matrix in **Figure 2**. An expanded version of this matrix is provided in Supplementary Figure 3. As seen from the AAI matrix (**Figure 2**), the members of the four main clades observed in the phylogenetic trees (**Figure 1** and Supplementary Figure 1) showed higher amino acid identity to members within each clade than to the other *Mycobacterium* species. Further, members of the "*Triviale*" clade could be clearly distinguished from the "*Terrae*" clade, based on their much lower amino acid identity to the members of this latter clade. In addition, members of the "*Abscessus-Chelonae*" clade exhibited a high degree of amino acid identity (Avg. 92%) to other members of this clade, but significantly lower similarity to members of the "*Fortuitum-Vaccae*" or the "*Tuberculosis-Simiae*" clades (Avg. 62%). The results of the genome relatedness analysis support the existence of the four main clades observed in the phylogenetic trees and also the distinctness of the "*Triviale*" clade from members of the "*Terrae*" clade.

Molecular Signatures Specific for the Genus *Mycobacterium* and Its Main Clades

The results of phylogenomic studies and genomic similarity analysis indicated that the known mycobacterial species can be divided into five main groups including the "*Triviale*" clade. However, as the branching of species in phylogenetic trees can be affected by a large number of variables (Stackebrandt, 1992; Ludwig and Klenk, 2005; Klenk and Goker, 2010; Gupta, 2016b), it is important to confirm the genetic cohesiveness of the observed clades by independent means not involving phylogenetic analysis. Rare genetic changes, such as insertions and deletions in genes/proteins as well as novel genes/proteins (viz. CSIs and CSPs) which are uniquely shared by an evolutionary related group of organisms constitute synapomorphic characteristics, whose shared presence in a given group of organisms generally results from the occurrence of the genetic changes in a common ancestor of the group (Gupta, 1998, 2014, 2016b; Rokas and Holland, 2000; Dutilh et al., 2008). In our earlier work on Actinobacteria, we described large numbers of CSIs and CSPs which were distinctive characteristics of either the entire phylum or a number of different clades within this phylum at multiple phylogenetic/taxonomic levels (Gao and Gupta, 2005, 2012; Gao et al., 2006; Gupta et al., 2013b). Although the focus of this earlier study was not on mycobacteria, a limited number of CSIs and CSPs which were then specific for the genus *Mycobacterium* were also identified (Gao et al., 2006; Gao and Gupta, 2012). Since these earlier studies, genome sequences for a large number of other mycobacterial species have become available (Supplementary Table 1). In the present work, we have carried out comprehensive comparative genomic studies on members of the genus *Mycobacterium*, to identify molecular markers (CSIs and CSPs) that are specific characteristics of either all mycobacterial species or of the identified main clades within this genus. The results of these analyses have identified 172 molecular markers (CSIs and CSPs) that are uniquely found in either all mycobacteria or by the members of different main clades identified by phylogenomic studies. Brief descriptions of the characteristics of the identified molecular markers and their group specificities are provided below.

Molecular Signatures (CSIs and CSPs) Specific for the Genus *Mycobacterium*

Our analysis has identified 10 CSIs in proteins involved in diverse functions that are uniquely found in all available mycobacterial homologs. An example of a CSI that is specific for the genus *Mycobacterium* is shown in **Figure 3**. In the partial sequence alignment of the protein EgtB (ergothioneine biosynthesis protein), a two amino acid insertion in a conserved

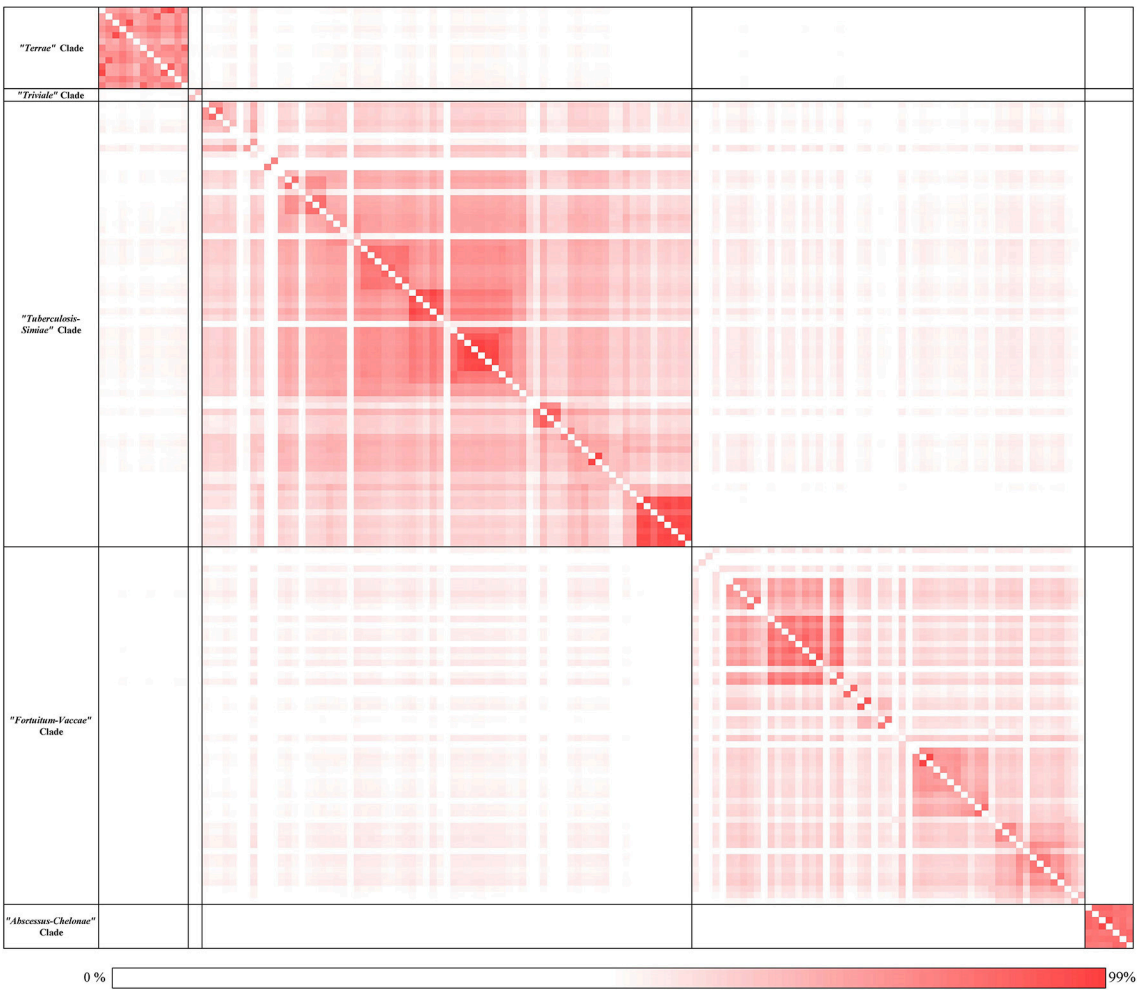


FIGURE 2 | A matrix of the average Amino Acid Identity for the 150 *Mycobacterium* genomes used in this study. A darker shade represents higher similarity between the pair of genomes. The average amino acid identity between each pair of genomes was calculated as described in the Methods section. The numerical values underlying this matrix are provided in Supplementary Figure 3.

region is exclusively found in all members of the genus *Mycobacterium*, but it is not present in the top 500 homologs of this protein sequence in other bacteria. Ergothionine is a naturally occurring amino acid (thiourea derivative of histidine), whose synthesis is uniquely carried out by only certain groups of actinobacteria as well as some cyanobacteria and fungi (Fahey, 2001). More detailed sequenced information for this CSI as well as sequence information for 9 other CSIs in important proteins, which are also specific for the genus *Mycobacterium* is provided in Supplementary Figures 4–13 and their main characteristics are summarized in **Table 1**. Of the described CSIs, the CSI in the protein orotidine 5'-phosphate decarboxylase (Supplementary Figure 7) was identified in our earlier work (Gao and Gupta, 2012). Although the number of sequenced mycobacterial genomes has increased many folds, this CSI is still found only in members of the genus *Mycobacterium*.

We have previously described a number of CSPs, whose homologs were uniquely found in the then sequenced mycobacterial species (Gao et al., 2006; Gao and Gupta, 2012). In light of the large increase in the number of sequenced mycobacterial genomes, the group specificities of the previously described CSPs were re-examined. Results of these analyses reveal that despite >20-fold increase in the number of sequenced mycobacterial genomes since these CSPs were first identified (Gao et al., 2006), 9 of the CSPs reported in our earlier work are still specific for members of the genus *Mycobacterium* and no homologs showing significant similarities to these proteins are present in other bacteria (**Table 2**). In view of the unique shared presence of these 10 CSIs and 9 CSPs by either all or most members of the genus *Mycobacterium* (except for an isolated exception), the genetic changes leading to these genetic markers most likely initially occurred in a common ancestor of the genus *Mycobacterium* and then retained by all descendant species.

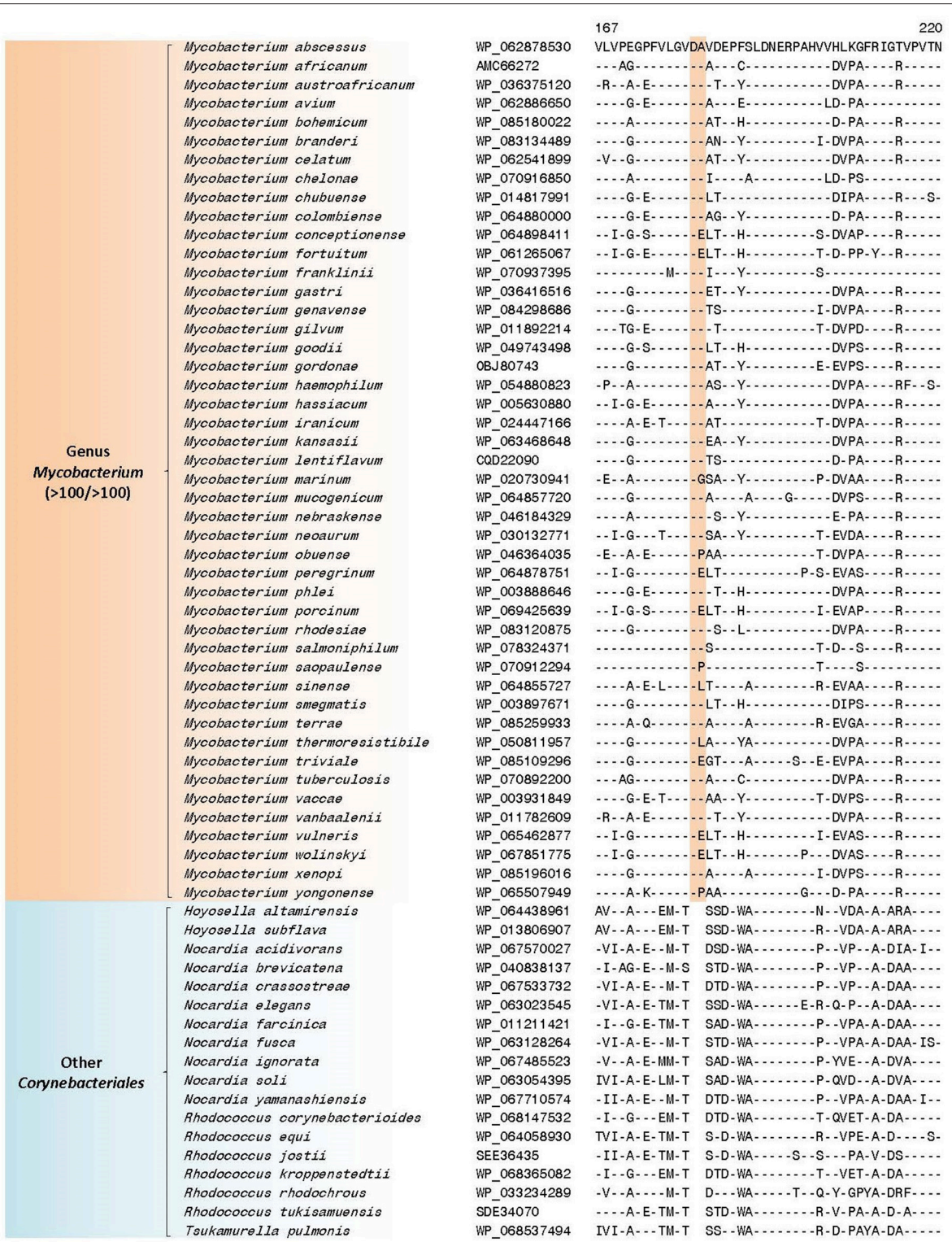


FIGURE 3 | Partial sequence alignment of a conserved region of the ergothioneine biosynthesis protein EgtB showing a two amino acid insertion (boxed) exclusively found in members of the genus *Mycobacterium* and not present in other *Corynebacteriales*. Dashes (-) in all alignments denote identity with the amino acid shown in the top sequence. Sequence information for only limited numbers of species is presented in this figure; a detailed alignment for this CSI is shown in Supplementary Figure 4. Information for additional CSIs specific for the genus *Mycobacterium* are provided in Supplementary Figures 4–13 and summarized in **Table 1**.

TABLE 1 | Conserved signature indels (CSIs) that are specific for different members of the genus *Mycobacterium* and those which are lacking only in members of the "*Abscessus-Chelonae*" clade.

Protein name	Accession number	Figure number	Indel size	Indel position	Specificity
Ergothioneine biosynthesis protein EgtB[a]	WP_062878530	**Figure 3** Supplementary Figure 4	2aa ins	167–220	Genus *Mycobacterium*
Precorrin-4 C(11)-methyltransferase	WP_078061976	Supplementary Figure 5	1aa ins	159–206	
NAD(P)H-quinone dehydrogenase	WP_062879231	Supplementary Figure 6	3aa ins	9–56	
Orotidine 5′-phosphate decarboxylase	WP_062879058	Supplementary Figure 7	1aa del	39–71	
Deoxyribonuclease IV	WP_031666830	Supplementary Figure 8	4aa del	109–149	
Serine hydrolase	WP_062879883	Supplementary Figure 9	3aa ins	122–166	
Peptidase C69[a,b]	WP_070410295	Supplementary Figure 10	1aa del	279–327	
SGNH/GDSL hydrolase family protein[a]	WP_031743956	Supplementary Figure 11	1aa del	95–143	
Succinate dehydrogenase[a]	WP_062880084	Supplementary Figure 12	1aa ins	27–56	
N-dimethylarginine dimethylaminohydrolase[a,b]	SGA93253	Supplementary Figure 13	1aa del	109–153	
Nif3-like dinuclear metal center hexameric protein	WP_066808468	**Figure 4** Supplementary Figure 14	2aa del	31–75	Genus *Mycobacterium* except the "*Abscessus-Chelonae*" Clade
Phosphoribosylamine-glycine ligase	CKM81105	Supplementary Figure 15	5 aa ins	106–161	
D-alanyl-D-alanine carboxypeptidase/D-alanyl-D-alanine-endopeptidase	WP_083039002	Supplementary Figure 16	1aa del	418–449	
Heat-inducible transcriptional repressor HrcA	WP_031668340	Supplementary Figure 17	2aa ins	193–228	

[a] *Only in comparison to other Corynebacteriales.*
[b] *Homologues of Hoyosella species were absent in BLASTp searches.*

Molecular Signatures Specific for the "*Abscessus-Chelonae*" Clade and Supporting the Deep Branching of this Group within the Genus *Mycobacterium*

The "*Abscessus-Chelonae*" clade, also referred to as *M. chelonae* or *M. abscessus* complex (Adékambi and Drancourt, 2004; Medjahed et al., 2010; Tortoli, 2012; Wee et al., 2017), consists of six members and it has recently gained clinical attention in light of its emerging pathogenicity to humans (Medjahed et al., 2010; Tortoli, 2014). In the phylogenetic trees constructed in our work, members of this clade form a monophyletic grouping which comprises the deepest branching lineage among the *Mycobacterium* species (**Figures 1A,B** and Supplementary Figure 1). The deep branching of the "*Abscessus-Chelonae*" clade in comparison to the other *Mycobacterium* species is also independently supported by 4 CSIs in four different proteins which are commonly shared by the homologs of all other mycobacterial species except those from the "*Abscessus-Chelonae*" clade. One example of a CSI depicting this pattern is presented in **Figure 4**, where in the partial sequence alignment of Nif3-like dinuclear metal center hexameric protein, a two amino acid deletion in a conserved region is present in all members of the genus *Mycobacterium* except members of the "*Abscessus-Chelonae*" clade. Additional information for this CSI and the sequence information for the three other CSIs exhibiting similar species distributions is provided in Supplementary Figures 14–17 and their main characteristics are summarized in **Table 1**. Based upon the species distributions of these CSIs, the genetic changes leading to them have likely occurred in a common ancestor of the other *Mycobacterium* species after the divergence of the "*Abscessus-Chelonae*" clade.

Our analyses have also identified 27 CSIs in proteins involved in diverse functions that are uniquely shared by members of the "*Abscessus-Chelonae*" clade providing strong evidence of the genetic cohesiveness and distinctness of this group of mycobacteria. Two examples of the CSIs specific for the "*Abscessus-Chelonae*" clade are shown in **Figure 5**. **Figure 5A** shows a partial sequence alignment of the protein uracil phosphoribosyltransferase, where a six amino acid insertion in a conserved region is present in all members of the "*Abscessus-Chelonae*" clade but absent in the homologs from all other *Mycobacterium* species as well as other groups of bacteria. Likewise, **Figure 5B** shows a four amino acid deletion in the sequence alignment of protein L-histidine N(alpha)-methyltransferase, which is also specific for the "*Abscessus-Chelonae*" clade. More detailed information for these CSIs and the 25 other identified CSIs, which are also specific for the "*Abscessus-Chelonae*" clade, is provided in Supplementary Figures 15, 18–43 and their main characteristics are summarized in **Table 3**. In addition to these CSIs, our work has also identified 24 CSPs listed in **Table 2**, for which homologs exhibiting significant similarity are only found in members of the "*Abscessus-Chelonae*" clade. Thus, the distinctness of the "*Abscessus-Chelonae*" clade from all other mycobacteria is strongly supported by 51 highly-specific molecular signatures identified in this work.

TABLE 2 | Conserved signature proteins (CSPs) specific for the genus *Mycobacterium* and members of the *"Abscessus-Chelonae"* clade.

Gene or protein	Accession number	Function	Length	Specificity
Hypothetical protein[a,c]	WP_011723520.1	Hypothetical	277	Genus *Mycobacterium* (Gao and Gupta, 2012)
Hypothetical protein[a]	WP_011723901.1	Hypothetical	129	
Hypothetical protein[a]	WP_011723955.1	Hypothetical	220	
Membrane protein[a,c]	WP_011724283.1	Atrophin-1	253	
PE-PPE domain-containing protein[a,b]	WP_011724324.1	Hypothetical	376	
DUF2561 domain-containing protein[a,c]	WP_011724709.1	Hypothetical	210	
Membrane protein[a,c]	WP_009976570.1	Actinobacterial Holin-x	131	
Hypothetical protein[a]	WP_003876314.1	Hypothetical	61	
Hypothetical protein[a]	WP_003874755.1	Hypothetical	116	
MAB_0188c	YP_001700942.1	Hypothetical	60	*"Abscessus-Chelonae"* Clade
MAB_0375	YP_001701128.1	Hypothetical	99	
MAB_0601	YP_001701353.1	Hypothetical	98	
MAB_2852c	YP_001703585.1	Hypothetical	108	
MAB_3058	YP_001703790.1	Hypothetical	127	
MAB_3079c	YP_001703811.1	Hypothetical	193	
MAB_1107c	YP_001701850.1	Hypothetical	74	
MAB_1519	YP_001702259.1	tRNA synthetase class II	127	
MAB_1642	YP_001702381.1	Hypothetical	60	
MAB_0008	YP_001700765.1	Hypothetical	75	
MAB_0245c	YP_001700999.1	Hypothetical	74	
MAB_2487[b]	YP_001703222.1	Hypothetical	75	
MAB_3020c	YP_001703752.1	Hypothetical	55	
MAB_1440c	YP_001702180.1	Hypothetical	76	
MAB_0014	YP_001700771.1	Hypothetical	74	
MAB_0015	YP_001700772.1	Hypothetical	95	
MAB_0345	YP_001701098.1	Hypothetical	170	
MAB_0448c	YP_001701201.1	Hypothetical	67	
MAB_0456	YP_001701209.1	Hypothetical	94	
MAB_0460	YP_001701213.1	Hypothetical	146	
MAB_2549	YP_001703284.1	Hypothetical	69	
MAB_1765	YP_001702504.1	Hypothetical	98	
MAB_1767	YP_001702506.1	Hypothetical	81	
MAB_1806[b]	YP_001702544.1	Mycobacterial 2 TMS Phage Holin (M2) Holin Family	138	

[a] *Previously identified by Gao and Gupta (2012).*
[b] *Some exceptions are present.*
[c] *A significant BLASTp hit was also observed for 1 to 2 other species of the genus Klebsiella.*

Molecular Signatures Specific for the *"Fortuitum-Vaccae"* Clade

The "*Fortuitum-Vaccae*" clade as designated here (see **Figure 1**) encompasses all rapid-growing mycobacterial species, except those from the "*Abscessus-Chelonae*" clade. In the present work, 4 CSIs and 10 CSPs have been identified that are specific for either all or most members of the "*Fortuitum-Vaccae*" clade and support the monophyletic clustering of these species as observed in the phylogenomic trees (**Figure 1**). One of the identified CSIs, which are specific for the "*Fortuitum-Vaccae*" clade, is found in the *LacI* family transcriptional regulator. In the partial sequence alignment of this protein shown in **Figure 6**, a five amino acid insert in a conserved region is exclusively found in different members of the "*Fortuitum-Vaccae*" clade but it is not found in any other mycobacteria. Three other CSIs showing similar species specificities are present in three other proteins. Detailed sequence information for all of these CSIs is provided in the Supplementary Figures 44–47 and the main characteristics of all CSIs specific for the "*Fortuitum-Vaccae*" clade are summarized in **Table 4**.

BLASTp searches on the protein sequences from the genome of *Mycobacterium aurum* (LSHTM) have also identified 10 CSPs, whose homologs, except for rare exceptions, are only found in the "*Fortuitum-Vaccae*" clade of *Mycobacterium* species. Most of these CSPs are hypothetical proteins and their characteristics are summarized in **Table 5**. For the first four CSPs listed in

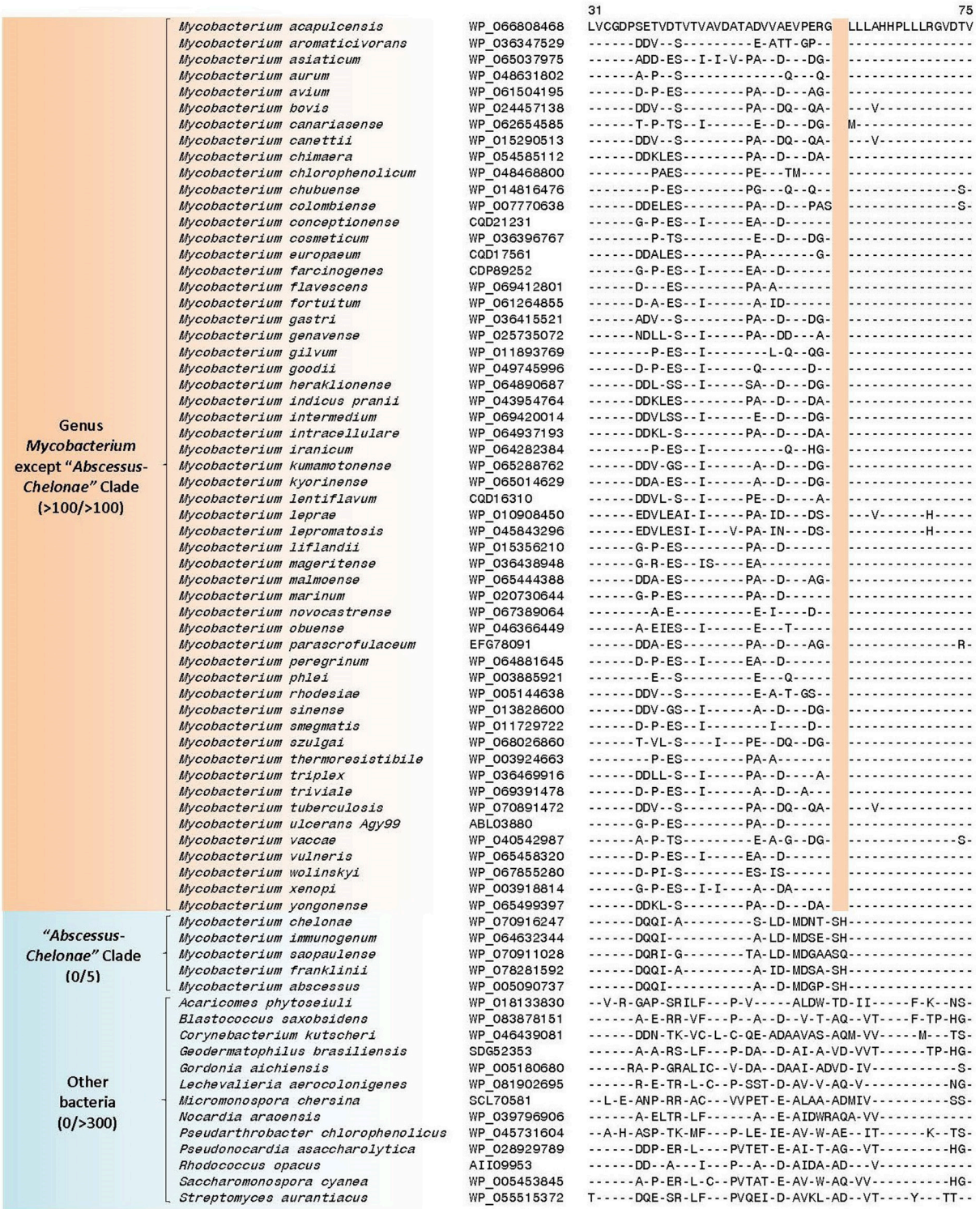


FIGURE 4 | A partial sequence alignment of a conserved region of Nif3-like protein exhibiting a two amino acid deletion that is specific for members of the genus *Mycobacterium* except members of the *"Abscessus-Chelonae"* clade; a detailed alignment for this CSI is shown in Supplementary Figure 14. Information for additional CSIs specific for the genus *Mycobacterium* are provided in Supplementary Figures 14–17 and summarized in **Table 1**. Dashes (-) in all alignments denote identity with the amino acid shown in the top sequence.

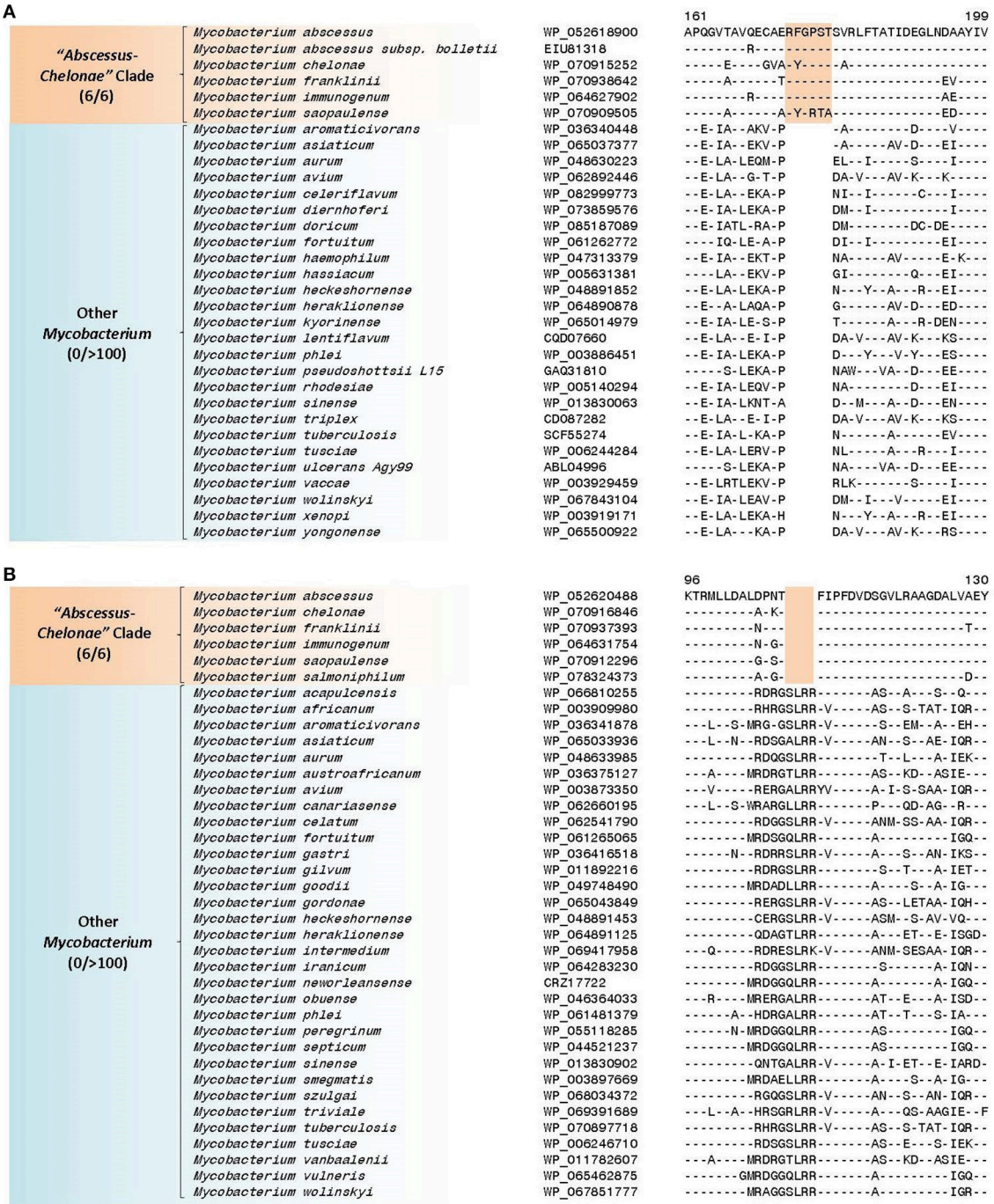


FIGURE 5 | (A) Partial sequence alignment of the protein uracil phosphoribosyltransferase showing a six amino acid insertion that is specific for the *"Abscessus-Chelonae"* clade; **(B)** Sequence alignment of L-histidine N(alpha)-methyltransferase showing a four amino acid deletion that is also specific for the *"Abscessus-Chelonae"* clade. More detailed alignments for these CSIs are shown in Supplementary Figures 18, 19 respectively. Additional CSIs that are specific for this clade are summarized in **Table 3** and sequences of these are provided in Supplementary Figures 15, 18–43.

Table 5, the homologs are present in different members of the "*Fortuitum-Vaccae*" clade, while for the remaining six CSPs, although they are specific for the "*Fortuitum-Vaccae*" clade, homologs were not detected in some members of this clade. In all, our identification of 14 molecular markers (4 CSIs and 10 CSPs), which are uniquely shared by members of the "*Fortuitum-Vaccae*" clade support its monophyletic origin and genetic cohesiveness.

TABLE 3 | Conserved signature indels (CSIs) specific to members of the "*Abscessus-Chelonae*" clade.

Protein name	Accession number	Figure number	Indel size	Indel position	Specificity
Uracil phosphoribosyltransferase	WP_052618900	**Figure 5A** Supplementary Figure 18	6aa ins	161–199	*"Abscessus-Chelonae"* Clade
L-histidine N(alpha)-methyltransferase	WP_052620488	**Figure 5B** Supplementary Figure 19	4aa del	94–130	
DUF58 domain-containing protein	WP_074245867	Supplementary Figure 20	10aa ins	364–407	
NADH-quinone oxidoreductase subunit G	WP_062878914	Supplementary Figure 21	4aa ins	724–762	
ATP-dependent helicase	WP_052624897	Supplementary Figure 22	3aa ins	169–208	
tRNA (cytidine(34)-2′-O)-methyltransferase	WP_005056099	Supplementary Figure 23	1aa del	13–52	
Glutamine-fructose-6-phosphate transaminase (isomerizing)	WP_052618678	Supplementary Figure 24	1aa del	49–81	
Error prone DNA polymerase	WP_052620523	Supplementary Figure 25	1aa ins	674–715	
2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase	WP_052622963	Supplementary Figure 26	2aa del	39–78	
DEAD/DEAH box helicase	WP_052619503	Supplementary Figure 27	1aa del	253–295	
Anion transporter	WP_052620306	Supplementary Figure 28	1aa del	51–90	
Membrane protein	WP_005081027	Supplementary Figure 29	6aa ins	331–379	
Nicotinate-nucleotide adenylyltransferase	WP_005074554	Supplementary Figure 30	2aa del	124–163	
CoA ester lyase	WP_052529870	Supplementary Figure 31	2aa del	172–207	
Hypothetical protein	WP_052613689	Supplementary Figure 32	6aa ins	342–382	
Hypothetical protein	WP_052613689	Supplementary Figure 33	2aa del	808–844	
Hypothetical protein	WP_057138049	Supplementary Figure 34	2aa ins	359–399	
Hypothetical protein	WP_052618664	Supplementary Figure 35	1aa del	199–235	
Bifunctional ADP-dependent (S)-NAD(P)H-hydrate dehydratase/ NAD(P)H-hydrate epimerase	WP_052543860	Supplementary Figure 36	3aa ins	318–354	
Hypothetical protein	WP_052621243	Supplementary Figure 37	1aa del	181–208	
Pyridoxal phosphate-dependent aminotransferase	WP_057138073	Supplementary Figure 38	4aa ins	157–194	
Carotenoid oxygenase	WP_062880095	Supplementary Figure 39	1aa ins	221–259	
Hypothetical protein	WP_062879314	Supplementary Figure 40	1aa del	139–168	
Hypothetical protein	WP_062879407	Supplementary Figure 41	3aa ins	340–367	
SAM-dependent methyltransferase	WP_062879423	Supplementary Figure 42	2aa ins	92–128	
SAM-dependent methyltransferase	WP_062879423	Supplementary Figure 43	1aa ins	12–47	
Phosphoribosylamine-glycine ligase	CKM81105	Supplementary Figure 15	2aa ins	106–161	

Molecular Signatures that Are Specific for the Slow-Growing *Mycobacterium*

The slow-growing *Mycobacterium* species generally form a monophyletic clade in most phylogenetic trees based on protein sequences (see **Figure 1** and Supplementary Figure 1) as well as those based on the 16S rRNA gene sequences (see Supplementary Figure 2) (Devulder et al., 2005; Kim et al., 2005; Hartmans et al., 2006; Mignard and Flandrois, 2008; Magee and Ward, 2012; Tortoli, 2012; Quast et al., 2013; Lory, 2014; Wang et al., 2015; Wee et al., 2017). The monophyly of the slow-growing *Mycobacterium* clade is also supported by 3 CSIs and 4 CSPs that have been identified in this study. One example of a CSI that is largely specific for the slow-growing *Mycobacterium* clade is shown in **Figure 7**. In the sequence alignment of alkyl-aryl sulfatase protein, a one amino acid insert in a conserved region is present in all of the homologs from slow-growing *Mycobacterium* species, but it is not found in the homologs of other *Mycobacterium* species. Detailed sequence information for this CSI and the two other CSIs showing similar specificities is provided in Supplementary Figures 48–50 and their main characteristics are summarized in **Table 4**. As noted above, the homologs for four of the identified CSPs (Accession numbers: YP_177721.1, YP_178025.1, WP_011725130.1, WP_003874405.1) are also specifically found in slow-growing *Mycobacterium* species (**Table 5**). The last two of these CSPs were identified by our earlier work based on limited number of genomes (Gao and Gupta, 2012) and they continue to be specific for this large clade of mycobacteria. Further, of the identified CSPs, which are specific for the slow-growing mycobacterial clade, three of the CSPs correspond to the PE or PPE family of proteins, which are often involved in mycobacterial virulence (Mukhopadhyay and Balaji, 2011).

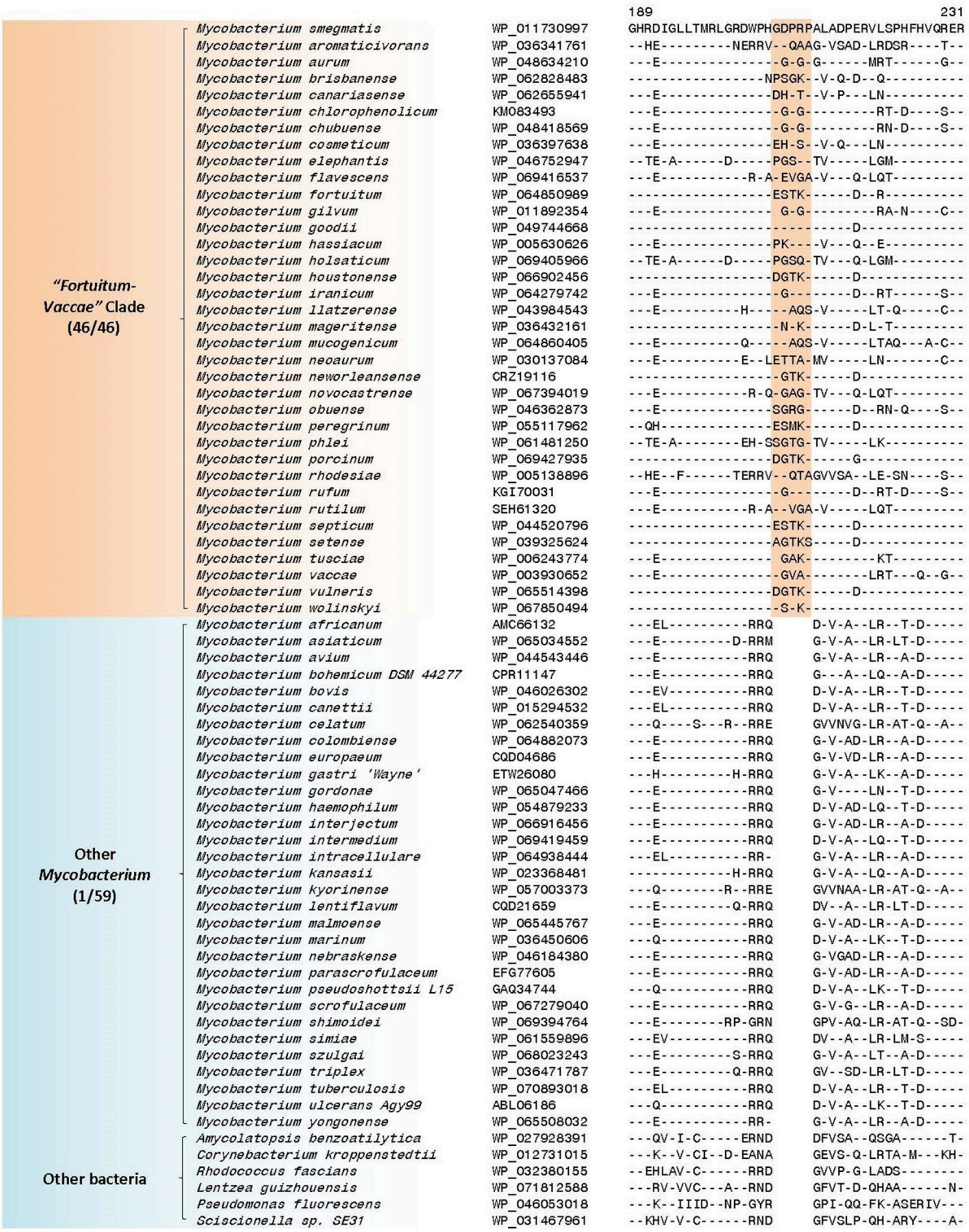


FIGURE 6 | A partial sequence alignment of a conserved region of LacI family transcriptional regulator showing a five amino acid insertion that is specific for the *"Fortuitum-Vaccae"* clade; a more detailed alignment of this CSI is shown in Supplementary Figure 44. Sequence information for additional CSIs that are specific for this clade is shown in Supplementary Figures 44–47 and summarized in **Table 4**.

TABLE 4 | Conserved Signature Indels (CSIs) specific for members of the *"Fortuitum-Vaccae"* clade, Slow-Growing *Mycobacterium* (*"Tuberculosis-Simiae"* + *"Terrae"* clades), and *"Tuberculosis-Simiae"* clade.

Protein name	Accession number	Figure number	Indel size	Indel position	Specificity
LacI family transcriptional regulator	WP_036341761	**Figure 6** Supplementary Figure 44	5aa ins	189–231	*"Fortuitum-Vaccae"* Clade
Cyclase	WP_066808156	Supplementary Figure 45	2aa ins	243–280	
CDP-x	WP_036344961	Supplementary Figure 46	1aa ins	49–75	
CDP-diacylglycerol–serine O-phosphatidyltransferase	WP_066811333	Supplementary Figure 47	1aa del	112–160	
Alkyl/aryl sulfatase	WP_083113621	**Figure 7** Supplementary Figure 48	1aa ins	123–155	*Mycobacterium* Slow-Growers
Succinate dehydrogenase iron-sulfur subunit	WP_083139296	Supplementary Figure 49	4aa ins	30–68	
Hypothetical protein	WP_009976218	Supplementary Figure 50	1aa del	51–100	
Hypothetical protein	WP_031701648	**Figure 8** Supplementary Figure 51	1aa del	81–119	*"Tuberculosis-Simiae"* Clade
Aldehyde dehydrogenase family protein	WP_080699385	Supplementary Figure 52	2aa del	403–450	
23S rRNA (guanosine(2251)-2′-O)-methyltransferase RlmB	WP_083139967	Supplementary Figure 53	1aa del	136–176	

In our phylogenetic trees, the slow-growing mycobacterial species form three main clades including a clade consisting of *M. triviale* and *M. koreense* ("*Triviale*" clade). The genetic cohesiveness of these clades of slow-growing mycobacteria is also supported by a large number of molecular signatures that are described below.

Molecular Signatures for the "*Tuberculosis-Simiae*" Clade

The "*Tuberculosis-Simiae*" clade in our work is comprised of all other slow-growing mycobacteria except those from the "*Terrae*" and "*Triviale*" clades. This clade encompasses various pathogenic *Mycobacterium* species including those from the *M. tuberculosis* complex, *M. avium* complex, *M. gordonae* clade, *M. kansasii* clade, *M. simiae* clade, as well as several other slow-growing species (Magee and Ward, 2012; Lory, 2014). We have identified a total of 3 CSIs that are specific for the "*Tuberculosis-Simiae*" clade (**Table 4**, Supplementary Figures 51–53). One example of a CSI specific for this clade, which is found in a protein of unknown function is shown in **Figure 8**, where a single amino acid deletion is found in all members of the "*Tuberculosis-Simiae*" clade, but it is not present in any other mycobacterial homolog. In addition to these CSIs, BLASTp searches on the proteins found in the genome of *Mycobacterium tuberculosis* H37Rv have identified 3 CSPs, whose homologs are only found in either all or most members of the "*Tuberculosis-Simiae*" clade. A summary of the CSPs which are specific for the "*Tuberculosis-Simiae*" clade is provided in **Table 5** and of these CSPs, one protein (Genbank Accession Number NP_218369.1) is annotated as a histone-like protein.

Molecular Signatures Demarcating the "*Terrae*" and "*Triviale*" Clades of Mycobacteria

The members of the "*M. terrae* complex" (Tortoli, 2012; Ngeow et al., 2015) has drawn attention recently as some members of this clade are opportunistic pathogens (Mignard and Flandrois, 2008; Kim et al., 2012, 2013; Tortoli, 2012, 2014; Tortoli et al., 2013; Ngeow et al., 2015; Vasireddy et al., 2016). In the core-genome protein trees and the tree based on 8 conserved proteins, members of the "*M. terrae* complex" form a monophyletic lineage consisting of two distinct subclades: a larger "*Terrae*" clade encompassing most of the species from the "*M. terrae* complex" and a deeper branching "*Triviale*" clade consisting of *M. triviale* and *M. koreense* (*M. parakoreense* also branches with these species in the 16S rRNA tree, Supplementary Figure 2). The phylogenetic distinctness of this larger "*Terrae*" + "*Triviale*" clade is also supported by a number of identified molecular signatures. In this work, we have identified 6 CSIs, which are specific for the larger "*Terrae* complex" consisting of the "*Terrae*" + "*Triviale*" clades (**Table 6**). Sequence information for one of the CSIs specific for the larger "*Terrae* complex" is presented in **Figure 9A**. In this case a four amino acid insertion in the protein ATP-dependent helicase is specifically present in all members of the "*Terrae* complex," but it is not present in any other bacteria. Detailed sequence information for this CSI as well as other CSIs specific for this clade is presented in Supplementary Figures 54–59 and summarized in **Table 6**. In addition to these CSIs, which are commonly shared by the "*Terrae*" + "*Triviale*" clades, our analyses have also identified 26 other CSIs listed in **Table 6**, which are specifically shared by only the members of the "*Terrae*" clade and not present in *M. triviale* and *M. koreense*. An example of such a CSI consisting of a four amino acid insertion found in

TABLE 5 | Conserved signature proteins (CSPs) specific for members of the *"Fortuitum-Vaccae"* clade, Slow-Growing *Mycobacterium* (*"Tuberculosis-Simiae"* + *"Terrae"* + "*Triviale*" clades), and *"Tuberculosis-Simiae"* clade.

Gene or protein	Accession number	Function	Length	Specificity
Hypothetical protein	WP_048630777.1	Hypothetical	91	*"Fortuitum-Vaccae"* Clade
Hypothetical protein[a,b]	WP_048632025.1	Hypothetical	124	
Hypothetical protein	WP_048632497.1	Hypothetical	79	
Hypothetical protein[b]	WP_048634851.1	Hypothetical	75	
Hypothetical protein[c]	WP_048633467.1	Hypothetical	200	
Hypothetical protein[c]	WP_048633322.1	Hypothetical	151	
Hypothetical protein[b,c]	WP_048631132.1	Ribonuclease E	320	
Hypothetical protein[c]	WP_048634509.1	Hypothetical	93	
Hypothetical protein[b,c]	WP_048630657.1	Hypothetical	153	
Hypothetical protein,[c]	WP_048632441.1	Prolipoprotein diacylglyceryl transferase	338	
PPE Family protein	YP_177721.1	Hypothetical	3,300	*Mycobacterium* Slow-Growers
PE Family protein PE36	YP_178025.1	Hypothetical	103	
PE Family protein[a,b]	WP_011725130.1	Hypothetical	99	
MAP_RS07685[a,b]	WP_003874405.1	WXG100 family	94	
Histone-like protein HNS	NP_218369.1	Histone-like protein	134	*"Tuberculosis-Simiae"* Clade
Rv4010	YP_004837050.1	Hypothetical Protein	83	
Membrane protein	NP_217322.1	Hypothetical Protein	63	

[a] *Previously identified by Gao and Gupta (2012).*
[b] *Some exceptions are present.*
[c] *Homologues from all species were not observed in BLASTp searches.*

the protein UDP-N-acetylmuramate–L-alanine ligase is shown in **Figure 9B**. Sequence information for all the "*Terrae*" clade CSIs is presented in Supplemntary Figures 35, 60–84 and summarized in **Table 6**. These CSIs serve to indicate the distinctness of the species from the "*Terrae*" clade from the deeper branching *M. triviale* and *M. koreense* species, which are part of the "*Triviale*" clade.

Our BLASTp searches on the protein sequences from the genome of *M. sinense* JDM601 (Zhang et al., 2011) and *M. triviale* DSM 44153 (Fedrizzi et al., 2017) have also identified many CSPs whose homologs are found specifically in either members of the larger "*Terrae* complex" or uniquely by species which are part of either the "*Terrae*" clade or the "*Triviale*" clade. A summary of these CSPs is provided in **Table 7**. Of the identified CSPs, two CSPs (viz. accession numbers WP_013830140.1 and WP_013827845.1) are uniquely found in most members of the "*Terrae*" + "*Triviale*" clades. However, a large number of the other identified CSPs are specific for only either members of the "*Terrae*" clade (15 CSPs) or members of the "*Triviale*" clade (22 CSPs) and their homologs are not detected in other mycobacteria. Four of the CSPs specific for the "*Triviale*" clade included in **Table 7** were also previously identified by Ngeow et al. (2015). The identification of a large number of CSPs, which are uniquely found in either all/most members of the "*Terrae*" clade or those from the "*Triviale*" clade again serve to clearly differentiate these two groups of mycobacteria and demarcate them in molecular terms.

DISCUSSION

The genus *Mycobacterium* comprises a large group of species (currently 188 species have validly published names), which includes some of the most impactful human pathogens (viz. *M. tuberculosis* and *M. leprae*) as well as large numbers of species found in diverse environments (Magee and Ward, 2012; Lory, 2014). In view of the immense clinical importance of certain *Mycobacterium* species, it is of much interest to have a reliable understanding as to how different species within this large group are related (Tsukamura, 1967a; Rogall et al., 1990; Stahl and Urbance, 1990; Goodfellow and Magee, 1998; Magee and Ward, 2012; Tortoli, 2012; Lory, 2014). However, despite much work (reviewed in Introduction), all known mycobacterial species are currently part of a single genus and their interrelationships are generally poorly understood (Magee and Ward, 2012; Tortoli, 2012; Lory, 2014; Fedrizzi et al., 2017). Genome sequences are now available for 150 of the 188 known mycobacterial species providing a unique opportunity for reliably understanding the relationships among the *Mycobacterium* species through genomic approaches. Using genome sequences, comprehensive phylogenetic and comparative genome analyses were carried out on *Mycobacterium* species using multiple independent approaches. In the first approach, phylogenomic trees were constructed for *Mycobacterium* species based on several large datasets of protein sequences including 1941 core proteins for the genus *Mycobacterium*, 136 core proteins for the phylum Actinobacteria, and another set of 8 highly conserved essential proteins found in all mycobacteria. Based on the core proteins in mycobacterial genomes, pairwise amino acid identity was also determined amongst different *Mycobacterium* species, providing a measure of the overall genetic relatedness of the species. In the third approach, exhaustive comparative genomic analyses were carried out on protein sequences of mycobacterial genomes to identify highly specific markers in the forms of CSIs and CSPs that are distinctive

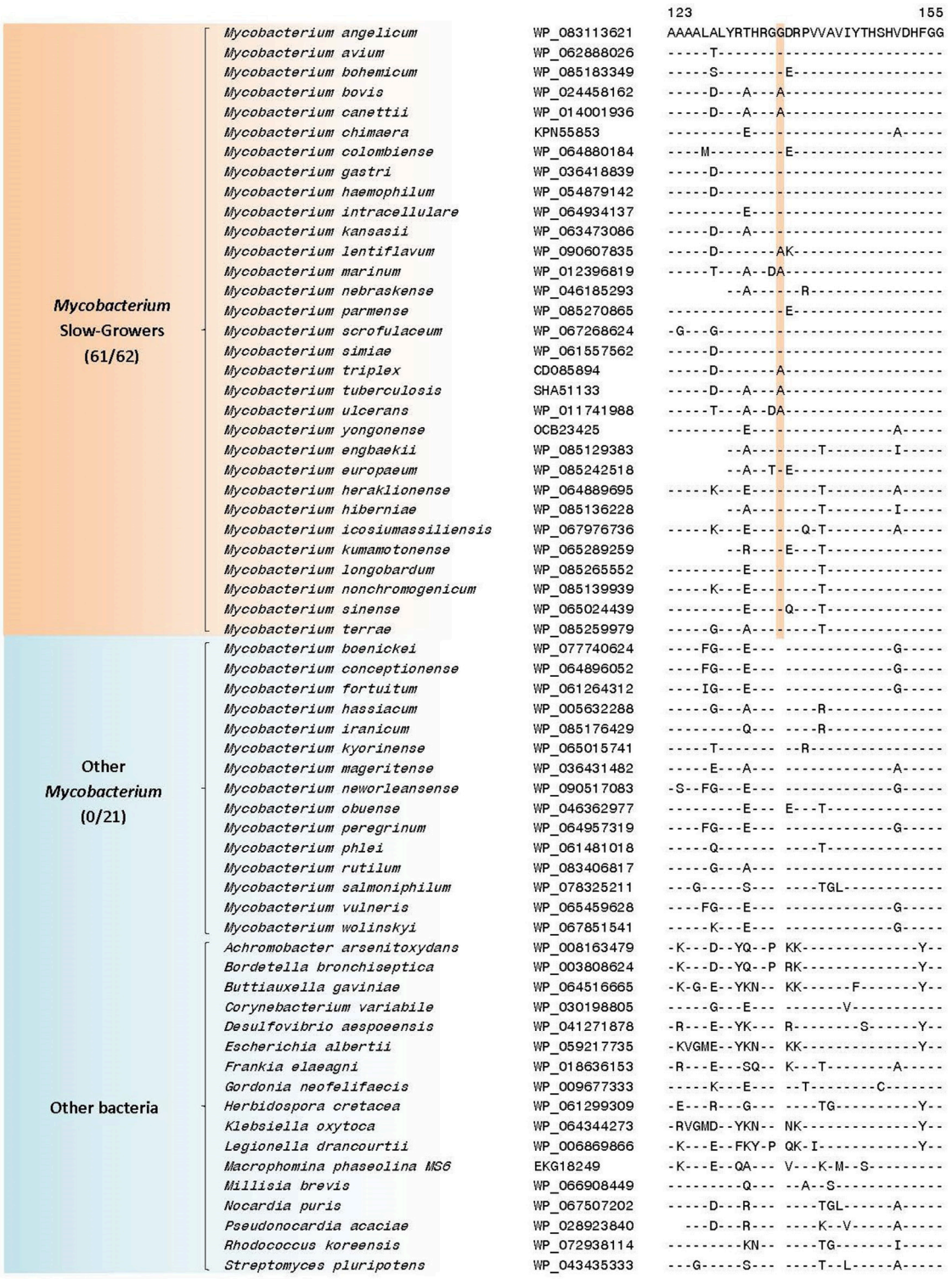


FIGURE 7 | A partial sequence alignment of a conserved region of the protein alkyl/aryl sulfatase showing a one amino acid insertion that is specific for the *Mycobacterium* slow-growers (i.e., *"Tuberculosis-Simiae"* + *"Terrae"*) clade; a detailed alignment of this CSI is shown in Supplementary Figure 48. Additional CSIs that are specific for this clade are summarized in **Table 4** and their sequence alignments are shown in Supplementary Figures 48–50.

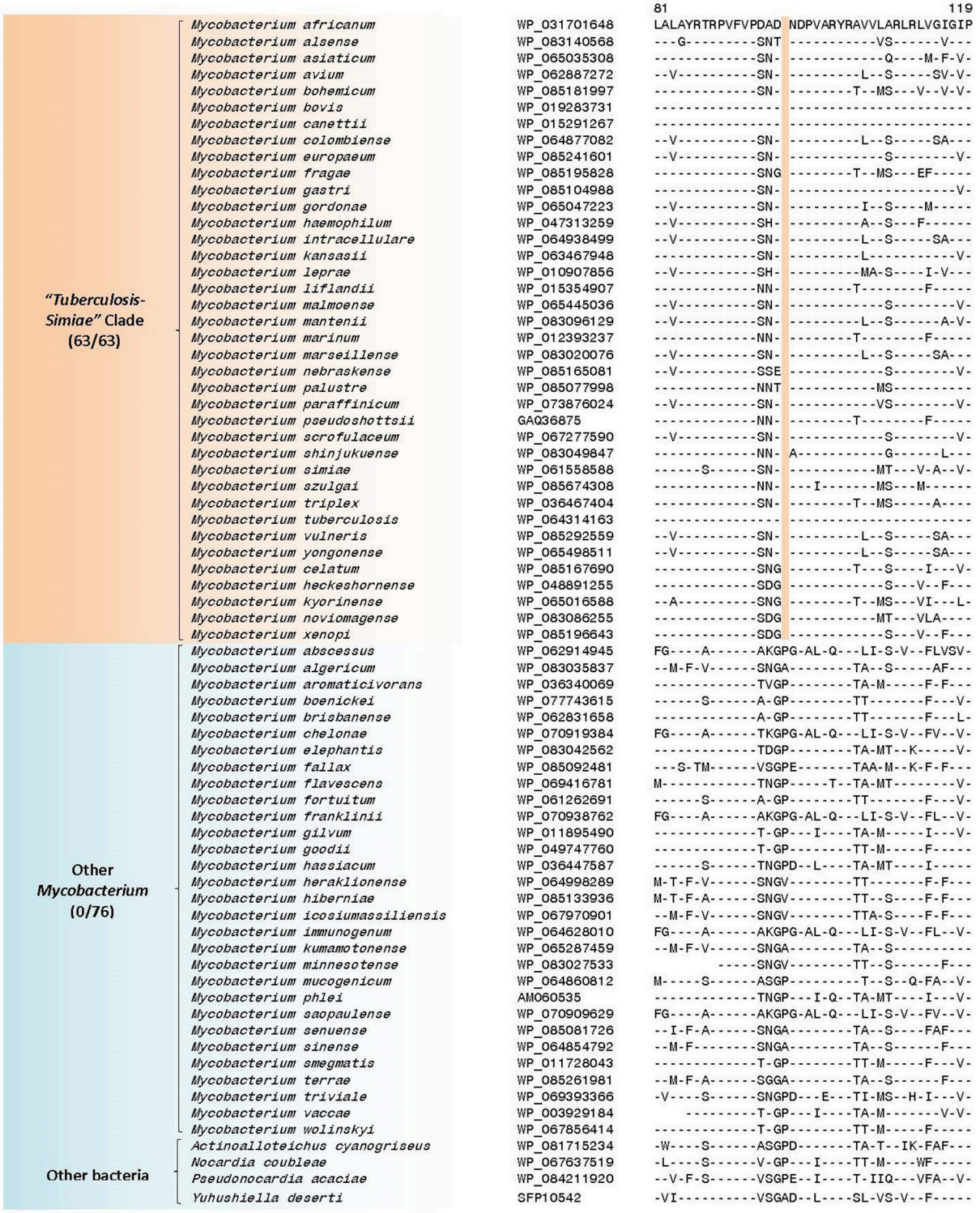


FIGURE 8 | Partial sequence alignment of a conserved region of a hypothetical protein showing a one amino acid deletion exclusively found in members of the *"Tuberculosis-Simiae"* clade; a detailed alignment of this CSI is shown in Supplementary Figure 51. Additional CSIs that are specific for this clade are shown in Supplementary Figures 51–53 and information for them is summarized in **Table 4**.

TABLE 6 | Conserved Signature Indels (CSIs) specific for members of the *"Terrae"* + *"Triviale"* clades or only the *"Terrae"* clade.

Protein name	Accession number	Figure number	Indel size	Indel position	Specificity
ATP-dependent helicase	WP_083036071	**Figure 9A** Supplementary Figure 54	4aa ins	1,144–1,184	*"Terrae"* + *"Triviale"* Clades
PDZ domain-containing protein	WP_083035840	Supplementary Figure 55	1aa del	290–324	
Ferredoxin reductase	WP_083035874	Supplementary Figure 56	3aa ins	199–263	
DUF2236 domain-containing protein	WP_083036515	Supplementary Figure 57	4aa del	37–80	
Hypothetical protein	WP_083040170	Supplementary Figure 58	1aa ins	100–141	
DUF4185 domain-containing protein	WP_083070918	Supplementary Figure 59	3aa ins	286–324	
Non-ribosomal peptide synthetase[a]	WP_083036306	Supplementary Figure 60	1aa del	497–532	*"Terrae"* Clade
Nucleoside hydrolase[a]	WP_085126548	Supplementary Figure 61	2aa del	134–174	
TetR/AcrR family transcriptional regulator[a]	WP_083037632	Supplementary Figure 62	1aa del	165–203	
Carbon starvation protein A[a]	WP_083035732	Supplementary Figure 63	4aa ins	593–639	
Error-prone DNA polymerase[a]	WP_046686430	Supplementary Figure 64	5aa ins	98–124	
TetR/AcrR family transcriptional regulator[a]	WP_083040593	Supplementary Figure 65	1aa ins	153–192	
TetR family transcriptional regulator[a]	WP_085129961	Supplementary Figure 66	1aa ins	157–199	
Hypothetical protein[a]	WP_083037591	Supplementary Figure 67	1aa del	47–92	
Amidohydrolase[a]	WP_083037811	Supplementary Figure 68	1aa del	4–32	
Carboxymuconolactone decarboxylase family protein[a]	WP_083037652	Supplementary Figure 69	1aa ins	1–29	
Polyketide cyclase[a]	WP_085128055	Supplementary Figure 70	3aa ins	107–152	
Spirocyclase AveC family protein[a]	WP_085128375	Supplementary Figure 71	1aa del	113–242	
Hypothetical protein[a]	WP_083036336	Supplementary Figure 72	3aa del	362–404	
TobH protein[a]	WP_083035918	Supplementary Figure 73	3aa ins	37–83	
UDP-N-acetylmuramate–L-alanine ligase	WP_083038184	**Figure 9B** Supplementary Figure 74	4aa ins	131–177	
DUF2236 domain-containing protein	WP_085129074	Supplementary Figure 75	1aa ins	145–190	
Cobaltochelatase subunit CobN	WP_083037938	Supplementary Figure 76	2aa ins	209–255	
Alpha/beta hydrolase	WP_083040154	Supplementary Figure 77	1aa ins	232–276	
Potassium transporter Kef	WP_083035846	Supplementary Figure 78	1aa ins	141–180	
Bifunctional tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase I	WP_083040227	Supplementary Figure 79	1aa del	330–357	
Membrane protein	KKB98129	Supplementary Figure 80	1aa del	104–137	
DUF222 domain-containing protein	WP_083036231	Supplementary Figure 81	5aa ins	57–100	
MFS transporter	WP_083036343	Supplementary Figure 82	2aa del	232–267	
Adenylate/guanylate cyclase domain-containing protein	WP_083036337	Supplementary Figure 83	1aa ins	369–407	
DUF2029 domain-containing protein	WP_083037148	Supplementary Figure 84	3aa ins	154–187	
Hypothetical protein	WP_052618664	Supplementary Figure 35	1aa del	199–235	

[a]*Homologues of M. triviale and M. koreense were absent in BLASTp searches.*

characteristics of the genus *Mycobacterium* as a whole or of different major clades within this genus. The results from all of these comprehensive genomic approaches reveal a consistent picture of the overall evolutionary relationships among the mycobacterial species, a summary of which is presented in **Figure 10**.

In phylogenetic trees constructed based on different large datasets of protein sequences, the *Mycobacterium* consistently grouped into four main strongly supported clades at the highest level. Within the larger "*Terrae* complex," the species *M. triviale* and *M. koreense* also consistently formed a deeper branching "*Triviale*" clade. The existence of these five clades is also supported by the high degree of genome relatedness amongst the members of each clade, as indicated by the results of average amino acid identity analysis. More importantly, our analyses of protein sequences from *Mycobacterium* species have resulted in the identification of a total of 172 novel molecular markers (CSIs and CSPs) that are distinctive characteristics of either the entire genus *Mycobacterium* or of the five clades identified within this genus at various phylogenetic levels. A graphical schematic of the identified molecular markers and the mycobacterial clades for which they are specific for is shown in **Figure 10**. Thus, the existence as well as the distinctness of the five main clades within the genus *Mycobacterium* is supported not only by comprehensive phylogenomic studies and by genome relatedness analysis, but also by the identification

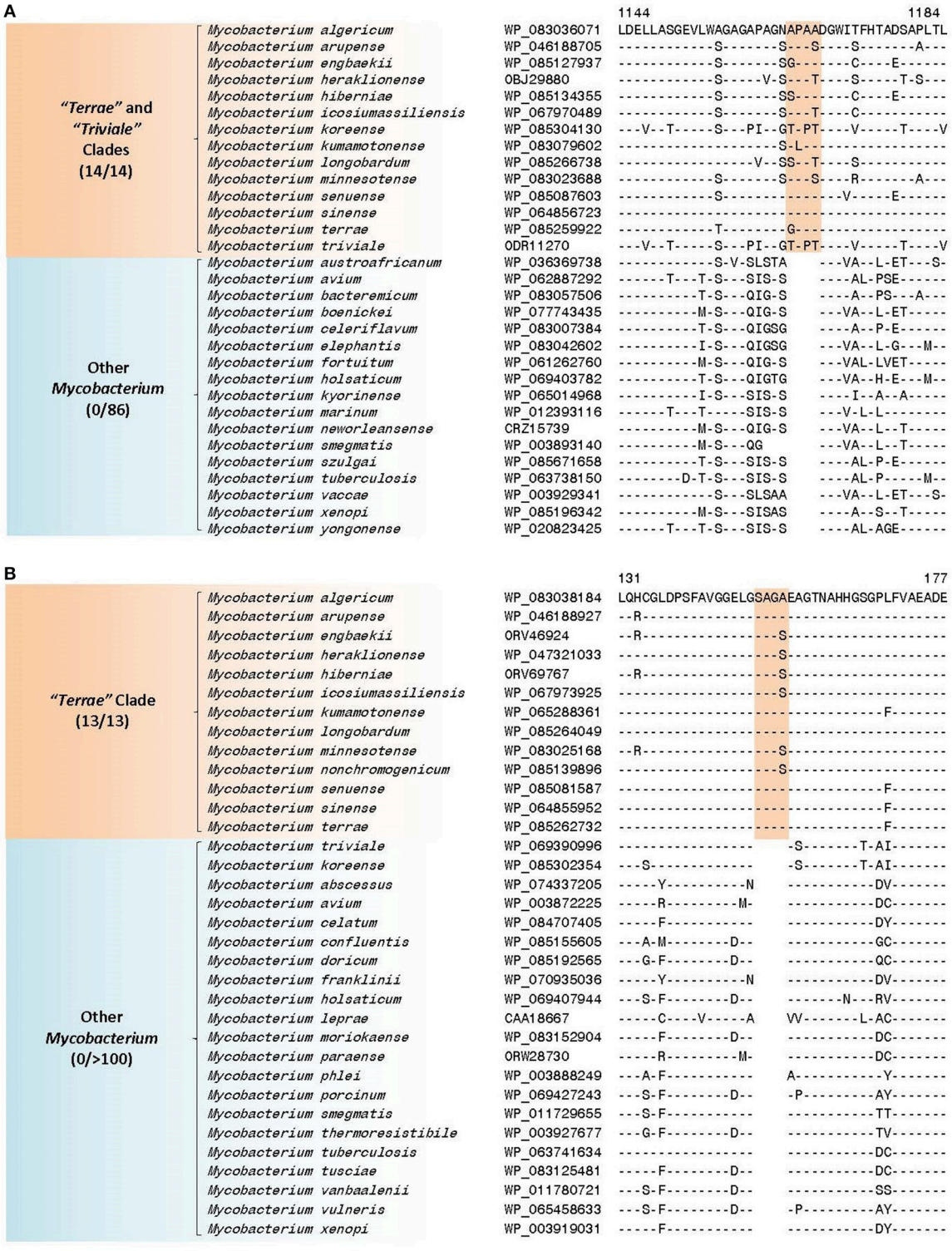


FIGURE 9 | Partial sequence alignment of a conserved region of **(A)** ATP-dependent helicase showing a four amino acid insertion that is specific for the *"Terrae"* + *"Triviale"* clades and **(B)** UDP-N-acetylmuramate—L-alanine ligase showing a four amino acid insertion that is specific for only the members of the *"Terrae"* clade but lacking in members of the *"Triviale"* clade as well as other mycobacteria. More detailed alignments of these CSIs are shown in Supplementary Figures 54 and 74, respectively. Additional CSIs that are specific for this clade are shown in Supplementary Figures 35, 54–84 and summarized in **Table 6**.

TABLE 7 | Summary of Conserved Signature Proteins (CSPs) that are specific for members of both "*Terrae*" + "*Triviale*" clades or only the "*Terrae*" clade or the "*Triviale*" clade.

Gene or protein	Accession number	Function	Length	Specificity
Hypothetical Protein	WP_013830140.1	Hypothetical	73	"*Terrae*" + "*Triviale*" Clades
Hypothetical protein	WP_013827845.1	Hypothetical	147	
Hypothetical Protein	WP_013828100.1	Hypothetical	267	"*Terrae*" Clade
CHAP domain-containing protein	WP_013830932.1	Amidase	209	
Hypothetical Protein	WP_013828443.1	Hypothetical	192	
Hypothetical Protein[a]	WP_013828919.1	MotB of proton channel complex MotA/MotB	159	
Hypothetical Protein	WP_013829267.1	Hypothetical	126	
Hypothetical Protein	WP_041317168.1	Hypothetical	172	
DUF732 domain-containing protein	WP_013827978.1	Hypothetical	179	
Hypothetical Protein[a]	WP_041318963.1	Hypothetical	231	
Hypothetical Protein	WP_013830185.1	Hypothetical	84	
Hypothetical Protein	WP_013828762.1	Hypothetical	1369	
Hypothetical Protein[a]	WP_013827315.1	Hypothetical	165	
Hypothetical Protein	WP_041318191.1	Hypothetical	69	
Hypothetical Protein[a,b]	WP_013829648.1	Glypican	207	
Hypothetical Protein[b]	WP_013829864.1	Hypothetical	131	
Hypothetical Protein[b]	WP_041317804.1	Hypothetical	133	
Hypothetical protein	WP_069390591.1	Hypothetical	199	"*Triviale*" Clade
Hypothetical protein	WP_069390644.1	Hypothetical	106	
Hypothetical protein	WP_069390667.1	Hypothetical	63	
Hypothetical protein	WP_069390717.1	Hypothetical	182	
Hypothetical protein	WP_069391089.1	Hypothetical	178	
Hypothetical protein	WP_069391367.1	PQQ enzyme repeat	152	
Hypothetical protein	WP_069391463.1	Hypothetical	68	
Hypothetical protein	WP_069391521.1	Hypothetical	441	
Hypothetical protein	WP_069391698.1	Hypothetical	63	
Hypothetical protein	WP_069391782.1	Hypothetical	180	
Hypothetical protein	WP_069391793.1	Hypothetical	188	
Hypothetical protein	WP_069392105.1	Hypothetical	129	
Hypothetical protein	WP_069392126.1	NT_Pol-beta-like Superfamily	272	
Hypothetical protein	WP_069392251.1	Hypothetical	319	
Hypothetical protein	WP_069392420.1	Hypothetical	116	
Hypothetical protein	WP_069392510.1	Hypothetical	71	
Hypothetical protein	WP_069392884.1	Hypothetical	126	
Hypothetical protein	WP_069392982.1	Hypothetical	79	
Hypothetical protein	WP_069392983.1	Hypothetical	104	
Hypothetical protein	WP_069393100.1	Hypothetical	105	
Hypothetical protein	WP_069393493.1	Hypothetical	128	
Hypothetical protein	WP_069393844.1	Hypothetical	128	

[a] *Previously also identified by Ngeow et al. (2015).*
[b] *Some exceptions are present.*

of large numbers of highly specific molecular markers, which serve to clearly demarcate these clades. Although it is difficult to specify how many characters are sufficient to divide a given taxon into more than one group, as this will depend upon the genetic diversity as well as phylogenetic depth of a taxon, in cases where the monophyly and distinctness of the described clades are strongly supported by multiple genome-scale phylogenetic trees as well as other independent approaches (e.g., AAI or ANI analysis), even 1–2 reliable molecular characters such as the CSIs and CSPs are sufficient for separation of a given group into distinct taxa (Gao and Gupta, 2012; Bhandari et al., 2013; Gupta et al., 2013a,b, 2015; Adeolu and Gupta, 2014; Bhandari and Gupta, 2014; Sawana et al., 2014; Adeolu et al., 2016; Alnajar and Gupta, 2017; Barbour et al., 2017).

It should be noted that molecular markers such as CSIs and CSPs represent synapomorphic characteristics and they provide important means for reliable identification/demarcation of different monophyletic clades of organisms (Baldauf and

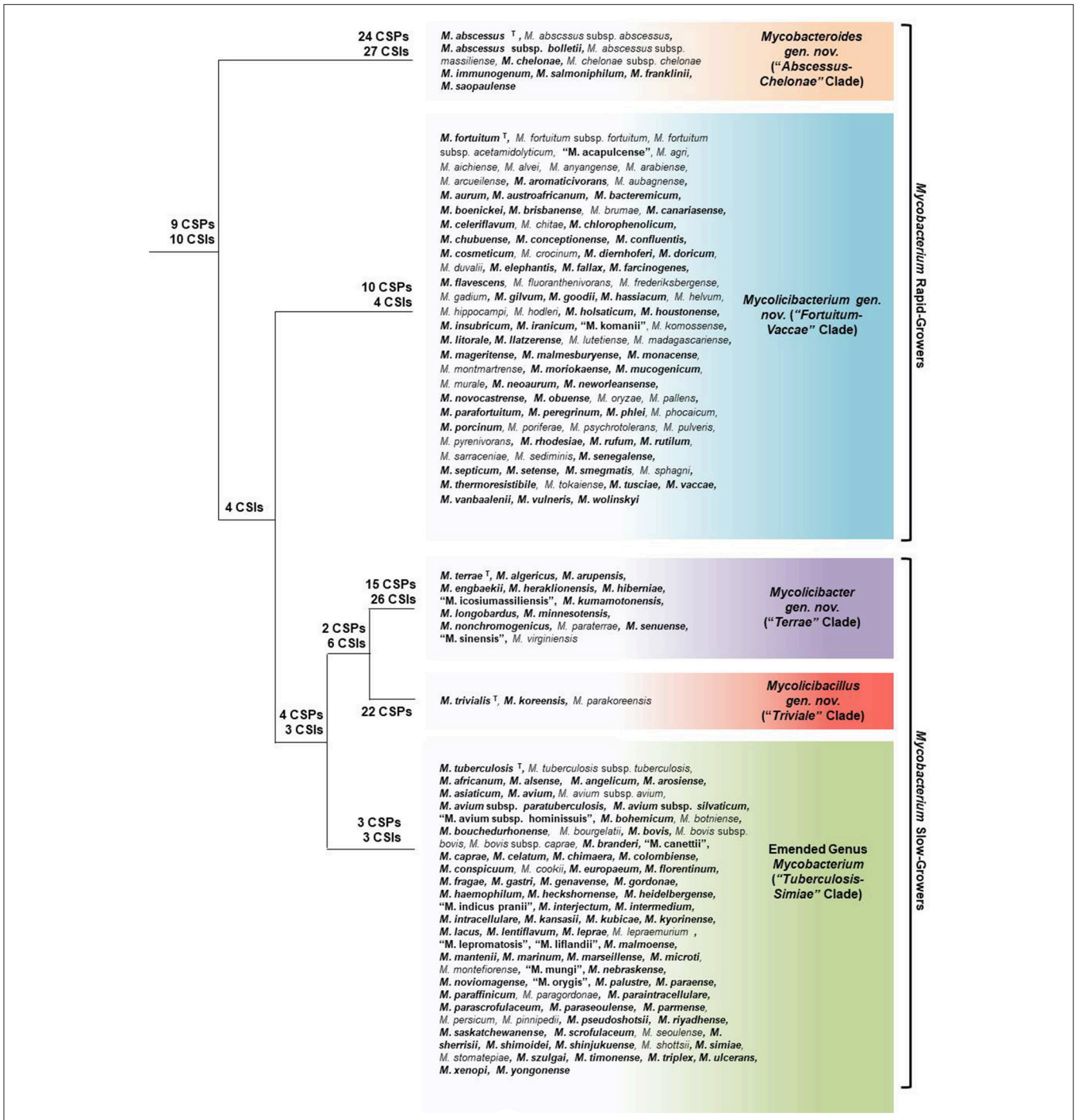


FIGURE 10 | A summary diagram depicting the overall relationships among the major groups of mycobacterial species. The numbers of identified CSIs and CSPs, which are specific for different clades are marked on the nodes. The names of the five main clades of mycobacterial species identified in this work, viz. "*Tuberculosis-Simiae,*" "*Terrae,*" "*Triviale,*" "*Fortuitum-Vaccae,*" and "*Abscessus-Chelonae*", along with their proposed or emended names and the species which are part of these clades are marked. Species which have had their genomes analyzed in this study are bolded. The superscript letter T beside a species indicates that it is the type species of the genus. The placements of other mycobacterial species, whose genomes have not been sequenced into these clades are based on their branching in the 16S rRNA tree (Supplmentary Figure 2). The species whose names are not italicized and are placed within quotation marks have not yet been validly published.

Palmer, 1993; Gupta, 1998, 2016b; Rokas and Holland, 2000; Dutilh et al., 2008; Chandra and Chater, 2014). Extensive earlier work on these markers show that they are highly reliable characteristics of different groups of organisms and species as relationships based on them are generally not affected by factors such as differences in evolutionary rates or lateral gene transfers (Bhandari et al., 2012; Gupta, 2014, 2016a,b). Further, each of these CSIs or CSPs, which are present in different genes/proteins, provide independent evidence supporting the monophyletic nature of the different identified clades, as well as providing novel and reliable means for the demarcation as well as diagnostics of species from these clades of bacteria (Ahmod et al., 2011; Wong et al., 2014). Extensive earlier work on CSIs/CSPs provides evidence that both large as well as small CSIs (even a one amino acid insert/deletion in protein sequence results from an in frame three nucleotides insertion/deletion within a conserved region) and CSPs provide reliable molecular markers for taxonomic and diagnostic studies, and they also exhibit a high degree of predictive ability to be present in other members of the indicated groups for which sequence information is lacking at present (Gao and Gupta, 2012; Adeolu and Gupta, 2014; Naushad et al., 2014; Sawana et al., 2014; Adeolu et al., 2016; Gupta, 2016b; Alnajar and Gupta, 2017). As noted earlier, some of the CSIs and CSPs specific for the genus *Mycobacterium* were identified when the sequence information was available for a limited number of mycobacterial genomes (Gao and Gupta, 2005, 2012; Gao et al., 2006). However, despite the large increase in the number of mycobacterial genomes, many of these CSIs and CSPs are still found to be specific for this genus. In view of their demonstrated specificity and reliability for the indicated group of organisms, the CSIs and CSPs in recent years have been used extensively for important taxonomic changes to a number of prokaryotic groups at various phylogenetic levels ranging from description of new classes, orders, families and genera including division of the original *Burkholderia*, *Borrelia* and *Thermotoga* genera into two or more genera (Gao and Gupta, 2012; Bhandari et al., 2013; Gupta et al., 2013a,b, 2015; Adeolu and Gupta, 2014; Bhandari and Gupta, 2014; Sawana et al., 2014; Adeolu et al., 2016; Alnajar and Gupta, 2017; Barbour et al., 2017).

It should be noted that a 12–14 nucleotide insert in the 16S rRNA sequences (in helix 18 between positions 451 and 482 in the *E. coli* sequence) is often used as a marker to differentiate between rapid-growing and slow-growing mycobacteria (Pitulle et al., 1992; Hartmans et al., 2006; Tortoli, 2012, 2014; Fedrizzi et al., 2017). The presence and absence of this insert in different sequenced mycobacterial species has been examined by us and this information is presented in Supplementary Figure 85. This insert, due to its presence in a conserved region, *also represents a CSI*. However, in contrast to the large numbers of CSIs described in this work, which are of fixed lengths and highly-specific characteristics of the described clades, this insert is of variable length (9-14 aa insertion) and it is lacking in many members of the slow-growing mycobacteria or the "*Tuberculosis-Simiae*" clade (Hartmans et al., 2006; Tortoli, 2012, 2014). Thus, unlike the different CSIs identified in the present work, this insert in the 16S RNA is not a distinguishing characteristic of either all slow-growing *Mycobacterium* species (i.e., "*Tuberculosis-Simiae*" + "*Terrae*" + "*Triviale*" clades) or of the "*Tuberculosis-Simiae*" clade. However, all of the species belonging to the "*Terrae*" clade contain a 14 nucleotide insert in this position, which provides a signature CSI for this clade, similar to the large numbers of other CSIs and CSPs reported here (see **Figure 9**, **Tables 6**, **7**). In contrast to the molecular markers described here, which are discrete and highly specific characteristics of the different indicated clades of mycobacteria, other physical and chemotaxonomic characteristics described in literature for various groups of mycobacteria are not specific for the indicated groups (see Supplementary Table 3; Magee and Ward, 2012). The presence or absence of the described physical and chemotaxonomic characteristics is often based on subjective criteria and information for such characteristics is not available for large numbers of mycobacterial species (Magee and Ward, 2012). This makes it difficult to reliably ascertain the potential usefulness of such characteristics as reliable markers for any particular group of mycobacteria.

The results presented here also strongly indicate that the "*Abscessus-Chelonae*" clade comprises the earliest branching lineage within the genus *Mycobacterium*. Its early divergence within the genus *Mycobacterium* is strongly supported by phylogenetic studies and multiple identified CSIs which are commonly shared by all or most *Mycobacterium* species, but absent in this clade of species. The deeper branching of the "*Abscessus-Chelonae*" clade as well as the "*Fortuitum-Vaccae*" clade of fast-growing mycobacteria, in comparison to the clades of slow-growing mycobacteria, supports the inference from earlier work that the rapid-growing mycobacterial species are ancestral and the slow-growers have evolved from them (Pitulle et al., 1992; Hartmans et al., 2006; Magee and Ward, 2012; Tortoli, 2012, 2014; Fedrizzi et al., 2017). Another important inference from the present work is that while the two clades of slow-growing mycobacteria (i.e., "*Tuberculosis-Simiae*" and the larger "*Terrae* + *Triviale*" clade) group together in phylogenetic trees, the grouping together of the two clades of rapid-growing mycobacteria is not observed in any phylogenetic trees. Further, while in our work 3 CSIs and 4 CSPs were identified that are commonly shared by members of the "*Tuberculosis-Simiae*" clade plus the "*Terrae*" + "*Triviale*" clade, no molecular marker was identified that is uniquely shared by the "*Abscessus-Chelonae*" and "*Fortuitum-Vaccae*" clades. It should be noted that while the distribution of most *Mycobacterium* species into the clades of slow-growing and fast-growing bacteria is generally in concordance with their rate of growth (Hartmans et al., 2006; Magee and Ward, 2012; Fedrizzi et al., 2017), a few exceptions are observed in this regard. In particular, the species *M. doricum*, *M. vulneris* and *M. tusciae*, which are slow-growing mycobacterial species (Magee and Ward, 2012; Fedrizzi et al., 2017), consistently branch within the "*Fortuitum-Vaccae*" clade of fast-growing mycobacteria. These species are also found to share the molecular signatures specific for the "*Fortuitum-Vaccae*" clade, but they lack the signatures for the slow-growing clades of mycobacteria. The anomalous branching of *M. doricum* and *M. tusciae* with the rapid-growing mycobacteria has also been reported in earlier work (Magee and Ward, 2012; Fedrizzi et al., 2017). This observation in conjunction with our

results showing that both the slow-growing and fast-growing *Mycobacterium* species form at least two distinct clades, and that the rapidly-growing species do not form a monophyletic lineage, indicates that the differentiation of the *Mycobacterium* species based solely on their growth rate is of limited use for developing a coherent taxonomic framework that is consistent with genomic and phylogenetic characteristics.

Of the main clades of mycobacteria described here, the "*Terrae*" + "*Triviale*" and the "*Abscessus-Chelonae*" clades are recognized from earlier phylogenetic studies (Adékambi and Drancourt, 2004; Mignard and Flandrois, 2008; Tortoli, 2012, 2014; Fedrizzi et al., 2017; Wee et al., 2017). In the present work, distinctness of the "*Abscessus-Chelonae*" clade is established by 51 molecular markers (CSIs and CSPs) which are specific for this clade. Although our work has identified some molecular markers that are specific for the larger "*Terrae*" + "*Triviale*" clade, our results strongly indicate that the species from the "*Triviale*" clade are phylogenetically and molecularly distinct from those of the "*Terrae*" clade. The distinctness of these two clades is also strongly supported by larger numbers of molecular markers identified in our work that are uniquely shared by the members of either the "*Terrae*" clade or the "*Triviale*" clades. The "*Terrae*" clade is also distinguished from others by the presence of a 14 nucleotide insertion in the helix 18 of the 16S rRNA gene (Tortoli, 2012, 2014; Ngeow et al., 2015). The other two main clades of mycobacteria described here namely the "*Tuberculosis-Simiae*" clade and the "*Fortuitum-Vaccae*" clade, harbor >85% of the known *Mycobacterium* species and no molecular markers or other characteristics specific for these clades are known from earlier work. However, both these large clades of mycobacteria can now be reliably demarcated on the basis of multiple highly-specific molecular signatures. In addition to the five clades described here, a number of other smaller clades are observed in the phylogenetic trees (**Figure 1** and Supplementary Figure 1). However, the work on characterization of these smaller subclades could be undertaken in future studies.

The work presented here based on multiple lines of evidence provide compelling support that the species from the genus *Mycobacterium* are comprised of five phylogenetically coherent clades, which can now be robustly distinguished from each other based on their branching in phylogenomic trees and multiple highly specific molecular signatures (**Figure 10**). These results provide a strong phylogenetic and genomic framework for division of the existing genus *Mycobacterium* into five distinct genera, corresponding to the five main clades described here. On the basis of the presented results, we are proposing that the genus *Mycobacterium* should be emended to include only members of the "*Tuberculosis-Simiae*" clade, which includes *Mycobacterium tuberculosis*, the type species of the genus (Zopf, 1883; Lehmann and Neumann, 1896), (Approved Lists, 1980; Skerman et al., 1980). The species from the other four main clades "*Fortuitum-Vaccae*", "*Terrae*", "*Triviale*" and "*Abscessus-Chelonae*" are transferred to four new genera with the following proposed names, *Mycolicibacterium* gen. nov., *Mycolicibacter* gen. nov., *Mycolicibacillus* gen. nov. and *Mycobacteroides* gen. nov., respectively. In the proposed classification, all of the major human pathogens are retained within the emended genus *Mycobacterium*, whereas the genus *Mycolicibacterium* is primarily comprised of environmental species. Most members of the proposed genera *Mycolicibacter* and *Mycolicibacillus* are also non-pathogenic, except occasional association of some species with animal hosts or human patients (Tasler and Hartley, 1981; Smith et al., 2000; Tortoli, 2014). Some members from the proposed genus *Mycobacteroides* are known to be associated with lung, skin and soft tissue infections (Simmon et al., 2011; Magee and Ward, 2012; Tortoli, 2014), however, none of them are considered as major life-threatening pathogens (Magee and Ward, 2012; Tortoli, 2014). Nonetheless, all five of these genera will remain part of the family *Mycobacteriaceae* and their proposed names bear close similarity to the original genus name *Mycobacterium*. Thus, all of them can still be referred to as mycobacterial species or as *M.* (*species name*), causing minimum confusion with any other species.

The proposed division of the existing genus *Mycobacterium* into the five proposed genera will have many benefits in terms of understanding and clarifying the relationships among the known mycobacterial species. The proposed division clearly separates the major human and animal pathogenic species, which are now part of the emended genus *Mycobacterium*, from all other (i.e., a majority of) mycobacterial species, which are either non-pathogenic or are of lesser clinical significance. With the explicit division of the mycobacterial species into these groups, attention can now be focused on unique genetic and molecular characteristics that differentiate the members of these groups of microbes. For each of these proposed genera, multiple CSIs and CSPs that are specific for these groups have been identified. Based on these molecular markers, it should be possible to develop novel and more reliable diagnostic methods for the identification of members of these groups by either *in silico* analysis of genomic sequences (based on BLASTp searches examining the presence or absence of these molecular sequences) or by experimental means utilizing PCR-based assays (Ahmod et al., 2011; Wong et al., 2014). Further, although the cellular functions of most of the identified CSIs or CSPs are not known, earlier work on other CSIs/CSPs has shown that these molecular characteristics are essential or play important functional roles in the organisms where they are found (Singh and Gupta, 2009; Schoeffler et al., 2010; Chandra and Chater, 2014; Gupta, 2016c). For example, some of the CSPs which are specific for the slow-growing mycobacterial species belong to the PE or PPE family of proteins, which play a role in virulence determination (Mukhopadhyay and Balaji, 2011). Hence, further functional investigations on the identified CSIs/CSPs are expected to lead to discovery of novel biochemical and/or other properties that are specific for either the entire *Mycobacteriaceae* family or for members of different genera that are part of this family.

The descriptions of the emended family *Mycobacteriaceae,* the emended genus *Mycobacterium* and of the four newly proposed genera viz, *Mycolicibacter* gen. nov., *Mycobacteroides* gen. nov., *Mycolicibacillus* gen. nov. and *Mycolicibacterium* gen. nov. are given below. Brief descriptions of the new species names combinations as well as some new species names resulting from the proposed taxonomic changes are also given below.

Emended Description of the Family *Mycobacteriaceae* Chester 1897 (Approved Lists 1980) (Skerman et al., 1980)

Mycobacteriaceae (My.co.bac.te.ri.a.ce´ae. N.L. neut. n. *Mycobacterium* type genus of the family; suff. *-aceae* ending to denote a family; N.L. fem. pl. n. *Mycobacteriaceae* the *Mycobacterium* family).

The family *Mycobacteriaceae* contains the type genus *Mycobacterium* as well as the genera *Mycolicibacter* gen. nov., *Mycolicibacterium* gen. nov., *Mycolicibacillus* gen nov., and *Mycobacteroides* gen. nov. Additionally, the genus *Amycolicoccus* is also indicated to be a part of this family (Wang et al., 2010; Parte, 2014). However, the sole type species of this genus, *Amycolicoccus subflavus,* is now reclassified as *Hoyosella subflava* (Hamada et al., 2016). The general characteristics of the family *Mycobacteriaceae* are as described by Magee and Ward (2012) for the genus *Mycobacterium*. The members of this family are aerobic to microaerophilic, slightly curved or straight rods (0.2–0.6 × 1.0–10 μm), which are acid–alcohol-fast at some stage of growth. Difficult to stain by Gram's-method, but are usually considered Gram-stain-positive. Some species may exhibit filamentous or mycelium-like growth. Cells are nonmotile and asporogenous. Colonies may be white- to cream-colored; some strains produce yellow- or orange-pigmented colonies with or without light stimulation. Whole-organism hydrolysates are rich in *meso*-diaminopimelic acid, arabinose, and galactose. The peptidoglycan is of the A1g type. Muramic acid moieties are *N*-glycolated. Cells and cell walls are rich in lipids. These include waxes which have characteristic, chloroform-soluble, mycolic acids with long (60–90 carbon atoms) branched chains. The fatty acid esters released on pyrolysis MS of mycolic acid esters have 22–26 carbon atoms. Cells contain diphosphatidylglycerol, phosphatidyl-ethanolamine, phosphatidylinositol, and phospatidylinositol mannosides as predominant polar lipids, straight-chain saturated, unsaturated, and 10-methyloctadecanoic (tuberculostearic) fatty acids as major fatty acid components, and dihydrogenated menaquinones with nine isoprene units as the predominant isoprenolog. The family includes obligate parasites, saprophytes, and opportunistic forms. The G+C content of genome-sequenced species varies from 57 to 71 (mol %) and genome size ranges from 3.1 to 10.5 Mbp. The members of the family *Mycobacteriaceae* form a distinct clade in the 16S rRNA tree and they are distinguished from all other members of the order *Corynebacteriales* by their unique shared presence of conserved signature indels described in this work (**Table 1**) in the following 10 proteins (viz. serine hydrolase, precorrin-4 C(11)-methyltransferase, NAD(P)H-quinone dehydrogenase, orotidine 5′-phosphate decarboxylase, deoxyribonuclease IV, peptidase C69, SGNH/GDSL hydrolase family protein, succinate dehydrogenase, N-dimethylarginine dimethylaminohydrolase, ergothioneine biosynthesis protein EgtB). Additionally, the homologs of the following nine proteins (accession numbers are in parenthesis) are also uniquely found in members of the family *Mycobacteriaceae* viz. hypothetical protein (WP_011723520.1), hypothetical protein (WP_011723901.1), MAV_11221(WP_011723955.1), membrane protein (WP_011724283.1), PE-PPE domain-containing protein (WP_011724324.1), DUF2561 domain-containing protein (WP_011724709.1), Membrane protein (WP_009976570.1), hypothetical protein (WP_003876314.1) and hypothetical protein (WP_003874755.1) (see **Table 2** in this work).

Emended Description of the Genus *Mycobacterium* Lehmann and Neuman 1896 (Approved Lists 1980) (Skerman et al., 1980)

Mycobacterium (My.co.bac.te´ri.um. Gr. n. *mykes* a fungus; N.L. neut. n. *bacterium*, a small rod; N.L. neut. n. *Mycobacterium*, a fungus rodlet).

The type species is *Mycobacterium tuberculosis* (Zopf 1883) Lehmann and Neumann 1896 (Approved Lists 1980) (Skerman et al., 1980).

Members of this genus whose are slow-growing bacteria requiring at least 7 days of incubation at optimal temperatures to form colonies. Several species are obligate parasites of human and animals and the genus harbors a number of important human (e.g., *Mycobacterium tuberculosis, M. leprae, M. ulcerans*) and animal (e.g., *Mycobacterium bovis*) pathogens. Other phenotypic and chemotaxonomic characteristics of this genus are similar to that for the family *Mycobacteriaceae*.

Some species from this clade contain a 9–12 nucleotide long insert in helix 18 of the 16S rRNA gene sequence (Supplementary Figure 85; Hartmans et al., 2006; Tortoli, 2014). Species are indicated to generally lack the *LivFGMH* operon and the *shaACDEFG* cluster of genes, which encodes respectively for proteins allowing the transportation of leucine, isoleucine and valine into the bacteria and a Na^+/H^+ antiporter that is important for the homeostasis of Na^+ and H^+ (Wee et al., 2017). Presence of the components of Type VII secretion system has been reported in members of this genus (Wee et al., 2017). The members of this genus form a monophyletic clade in phylogenetic trees constructed based on 16S rRNA gene sequences as well as multiple large datasets of protein sequences described in this work including a tree based on 1941 core mycobacterial proteins, a tree based on 136 core proteins for the phylum Actinobacteria, and a tree based on concatenated sequences for eight conserved housekeeping proteins (viz. RpoA, RpoB, RpoC, GyrA, GyrB, Hsp65, EF-Tu, and RecA). Members of the genus *Mycobacterium* can be clearly distinguished from other genera within the *Mycobacteriaceae* family based on conserved signature indels described in this study (**Table 4**) in the following three proteins, a hypothetical protein, aldehyde dehydrogenase family protein and 23S rRNA (guanosine(2251)-2′-O)-methyltransferase, that are uniquely shared by the members of this genus. In addition, the homologs of the following three proteins (accession numbers are in parenthesis): a histone-like protein HNS (NP_218369.1), a hypothetical protein Rv4010 (YP_004837050.1) and a membrane protein (NP_217322.1), are also unique characteristics of the members of this genus.

The G-C content and genome sizes of the member species ranges from 57.8–69.3 (mol %) to 3.2–7.3 Mbp, respectively.

Description of *Mycolicibacter* gen. nov.

Mycolicibacter (My.co.li.ci.bac´ter. N.L. n. *acidum mycolicum*, mycolic acid; N.L. masc. n. *bacter*, rod; N.L. masc. n. *Mycolicibacter*, a genus of mycolic acid containing rod-shaped bacteria).

The type species is *Mycolicibacter terrae*.

The members of the genus *Mycolicibacter* are commonly referred to as the *M. terrae* complex. This genus contains species that are slow-growing (more than 7 days) and nonchromogenic with some species that show intermediate growth duration (5–15 days) (Tortoli, 2014; Ngeow et al., 2015). In phylogenetic trees, the *Mycolicibacter* clade forms a sister clade to a clade comprising of the genus *Mycobacterium*, harboring other slow-growing mycobacteria. Most members of this genus are non-pathogenic, but some species have been isolated from animal hosts (Tasler and Hartley, 1981) and human patients (Smith et al., 2000). Multiple antibiotic resistance has been reported for many of the isolates (Milne et al., 2009; Zhang et al., 2013b).

The members of this genus form a monophyletic clade in phylogenetic trees based on 16S rRNA gene sequences as well as multiple datasets of gene/protein sequences including a tree based on 1941 core mycobacteria proteins and a tree based on 136 core proteins for the phylum Actinobacteria. The members of the genus *Mycolicibacter* exhibit a closer relationship to members of the genus *Mycolicibacillus* in phylogenetic trees, which is also supported by a number of CSIs listed (**Table 6**) in the proteins ATP-dependent helicase, PDZ domain-containing protein, Ferredoxin reductase, DUF2236 domain-containing protein and two hypothetical protein with the accession number WP_083040170 and DUF4185 domain-containing protein, as well as 2 CSPs (viz. accession numbers WP_013830140.1 and WP_013827845.1) that are commonly shared by the members from these two genera. All of the species from this genus contain a 14 nucleotide insertion in the helix 18 of the 16S rRNA gene (Supplementary Figure 85; Tortoli, 2014). Additionally, the members of this genus are distinguished from members of all other genera within the family *Mycobacteriaceae* due to their possession of 26 conserved signature indels described in this study (**Table 6**) present in the following proteins, non-ribosomal peptide synthetase, nucleoside hydrolase, three different indels in TetR family transcriptional regulator, carbon starvation protein A, error-prone DNA polymerase, amidohydrolase, carboxymunconolacton decarboxylase family protein, polyketide cyclase, spirocyclase AveC family protein, TobH protein, UDP-N-acetylmuramate–L-alanine ligase, DUF2236 domain-containing protein, cobaltochelatase subunit CobN, alpha/beta hydrolase, potassium transporter Kef, bifunctional tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit Type 1 TsaB/ribosomal protein alanine acetyltransferase RimI, a membrane protein, DUF222 domain-containing protein, MFS transporter, adenylate/guanylate cyclase domain-containing protein, DUF2029 domain-containing protein and the following hypothetical proteins with the accession numbers (WP_083037591, WP_083040170, WP_083036336 and WP_052618664), that are uniquely found in the members of this genus. In addition, the homologs of the 17 conserved signature proteins, whose accession numbers are as follows (viz. WP_013830140.1, WP_013827845.1, WP_013828100.1, WP_013830932.1, WP_013828443.1, WP_013828919.1, WP_013829267.1, WP_041317168.1, WP_013827978.1, WP_041318963.1, WP_013830185.1, WP_013828762.1, WP_013827315.1, WP_041318191.1, WP_013829648.1, WP_013829864.1, and WP_041317804.1) are also distinctive characteristics of either all or most members of this genus (**Table 7**).

The members of the genus *Mycolicibacter* are characterized by high G-C content (66.3–70.3 mol %) and they have relatively short genomes (range 3.87–5.11 Mbp).

The description of *Mycolicibacter terrae* comb. nov. as well as the descriptions of new name combinations for other species which are part of the genus *Mycolicibacter* are provided in **Table 8**.

In addition to the new name combinations for species which are part of the genus *Mycolicibacter*, we also provide below description of two new species that should also be placed in the genus *Mycolicibacter*.

Description of *Mycolicibacter icosiumassiliensis* sp. nov. (i.co.si.u.mas.si.li.en´sis; L. masc. n. *icosiumassiliensis*, from the combination of Icosium, the Latin name of Algiers where the strain was first isolated and Massilia, the Latin name of Marseille, where the strain was described).

The description of this taxon is as given by Djouadi et al. (2016) for "*Mycobacterium icosiumassilensis*". The type strain is 8WA6 (= CSUR P1561 = DSM 100711).

Description of *Mycolicibacter sinensis* sp. nov. (sin.en´sis. N.L. masc. adj. *sinensis* means "belonging to China," indicating the source of the type strain).

The description of this taxon is as given by Zhang et al. (2013b) for "*Mycobacterium sinense*". The type strain is JDM601.

Description of *Mycolicibacillus* gen. nov.

Mycolicibacillus (My.co.li.ci.ba.cil´lus. N.L. n. *acidum mycolicum*, mycolic acid; L. masc. n. *bacillus*, a small staff or rod; N.L. masc. n. *Mycolicibacillus*, a genus of mycolic acid containing rod-shaped bacteria).

The type species is *Mycolicibacillus trivialis*.

The genus *Mycolicibacillus* is comprised of slow-growing nonchromogenic bacteria requiring more than 7 days of incubation at optimal temperatures to form colonies. In phylogenetic trees, members of this genus form a deep-branching distinct clade that is most closely related to members of the genus *Mycolicibacter*. A close relationship of the species from the genera *Mycolicibacillus* and *Mycolicibacter* is also supported by a number of CSIs listed in **Table 6** in the proteins ATP-dependent helicase, PDZ domain-containing protein, ferredoxin reductase, DUF2236 domain-containing protein, non-ribosomal peptide synthetase, hypothetical protein with accession number WP_083040170 and DUF4185 domain-containing protein and CSPs listed in **Table 7** (viz. accession numbers WP_013830140.1 and WP_013827845.1) that are commonly shared by these two groups of bacteria. Unlike members of the genus *Mycolicibacter*, which contain a 14 nucleotide insertion in the helix 18 of the 16S rRNA gene, members of the genus *Mycolicibacillus* lack an insertion in this position (Tortoli, 2014) (Supplementary Figure 85). In addition, the homologs showing significant sequence similarity for the 22 proteins

TABLE 8 | Descriptions of new name combinations for species in the genus *Mycolicibacter*.

New name combinations	Description and type strain
Mycolicibacter terrae comb. nov. (ter´rae. L. gen. n. *terrae*, of the earth)	Basonym: *Mycobacterium terrae* Wayne 1966 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Wayne (1966). The type strain is ATCC 15755 = CCUG 27847 = CIP 104321 = DSM 43227 = JCM 12143 = LMG 10394.
Mycolicibacter algericus comb. nov. (al.ge´ri.cus. N.L. masc. adj. *algericus*, of or pertaining to Algeria, the country where the strain was first isolated)	Basonym: *Mycobacterium algericum* Sahraoui et al., 2011 The description of this taxon is as given by Sahraoui et al. (2011). The type strain is TBE 500028/10 = Bejaia = CIP 110121 = DSM 45454.
Mycolicibacter arupensis comb. nov. (a.rup.en´sis. N.L. masc. adj. *arupensis*, pertaining to the ARUP Institute for Clinical and Experimental Pathology, where the type strain was characterized)	Basonym: *Mycobacterium arupense* Cloud et al., 2006 The description of this taxon is as given by Cloud et al. (2006). The type strain is AR30097 = ATCC BAA-1242 = DSM 44942.
Mycolicibacter engbaekii comb. nov. (eng.bae´ki.i. N.L. gen. masc. n. *engbaekii*, of Engbaek, to honour of the Danish mycobacteriologist H. C. Engbaek)	Basonym: *Mycobacterium engbaekii* Tortoli et al., 2013 The description of this taxon is as given by Tortoli et al. (2013). The type strain is ATCC 27353 = DSM 45694.
Mycolicibacter heraklionensis comb. nov. (he.ra.kli.on.en´sis N.L. masc. adj. *heraklionensis* from Heraklion the city in Crete island where many such strains were isolated)	Basonym: *Mycobacterium heraklionense* Tortoli et al., 2013 The description of this taxon is as given by Tortoli et al. (2013). The type strain is GN-1 = CECT 7509 = LMG 24735 = NCTC 13432.
Mycolicibacter hiberniae comb. nov. (hi.ber´ni.ae. L. gen. n. *hiberniae*, of *Hibernia*, the Latin name for Ireland, the source of the strains)	Basonym: *Mycobacterium hiberniae* Kazda et al., 1993. The description of this taxon is as given by Kazda et al. (1993). The type strain is Hi 11 = ATCC 49874 = CIP 104537 = DSM 44241 = JCM 13571.
Mycolicibacter kumamotonensis comb. nov. (ku.ma.mo.to.nen´sis. N.L. masc. adj. *kumamotonensis*, of or pertaining to Kumamoto Prefecture in Japan, where the type strain was isolated)	Basonym: *Mycobacterium kumamotonense* Masaki et al., 2007 The description of this taxon is as given by Masaki et al. (2006, 2007). The type strain is CST 7247 = CCUG 51961 = JCM 13453.
Mycolicibacter longobardus comb. nov. (lon.go.bar´dus. N.L. masc. adj. *longobardus*, of or pertaining to Lombardy, the region where the strains were isolated)	Basonym: *Mycobacterium longobardum* Tortoli et al., 2013 The description of this taxon is as given by Tortoli et al. (2013). The type strain is FI-07034 = CCUG 58460 = DSM 45394.
Mycolicibacter minnesotensis comb. nov. (min.ne.so.ten´sis. N.L. masc. adj. *minnesotensis*, of or belonging to Minnesota)	Basonym: *Mycobacterium minnesotense* Hannigan et al., 2013 The description of this taxon is as given by Hannigan et al. (2013). The type strain is DL49 = DSM 45633 = JCM 17932 = NCCB 100399.
Mycolicibacter nonchromogenicus comb. nov. (non.chro.mo.ge´ni.cus. L. adv. *non*, not; Gr. n. *chroma*, color; Gr. v. *gennaio*, to produce; L. masc. suff. *-icus*, suffix used with the sense of pertaining to; N.L. masc. adj. *nonchromogenicus*, intended to mean not producing color)	Basonym: *Mycobacterium nonchromogenicum* Tsukamura 1965 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Tsukamura (1965a). The type strain is ATCC 19530 = CCUG 28009 = CIP 106811 = DSM 44164 = JCM 6364 = NCTC 10424.
Mycolicibacter paraterrae comb. nov. (pa.ra.ter´rae. Gr. prep. *para* beside; *terrae* of the earth; N.L. gen. n. *paraterrae* specific epithet of a *Mycobacterium* species; N.L. gen. n. *paraterrae* a species similar to members of the *Mycobacterium terrae* complex)	Basonym: *Mycobacterium paraterrae* Lee et al., 2016 The description of this taxon is as given by Lee et al. (2010, 2016). The type strain is 05-2522 = DSM 45127 = KCTC 19556.
Mycolicibacter senuensis comb. nov. (se.nu.en´sis. N.L. masc. adj. *senuensis*, arbitrary name formed from the initial letters of Seoul National University, the organization that carried out the taxonomic investigation of the type strain)	Basonym: *Mycobacterium senuense* Mun et al., 2008 The description of this taxon is as given by Mun et al. (2008). The type strain is 05-832 = DSM 44999 = JCM 16017 = KCTC 19147.
Mycolicibacter virginiensis comb. nov. (vir.gi.ni.en´sis. N.L. masc. adj. *virginiensis* referring to the geographic location of the first recognized case)	Basonym: *Mycobacterium virginiense* Vasireddy et al., 2017 The description of this taxon is as given by Vasireddy et al. (2016, 2017). The type strain is MO-233 = DSM 100883 = CIP 110918.

listed in **Table 6** with the accession numbers WP_069390591.1, WP_069390644.1, WP_069390667.1, WP_069390717.1, WP_069391089.1, WP_069391367.1, WP_069391463.1, WP_069391521.1, WP_069391698.1, WP_069391782.1, WP_069391793.1, WP_069392105.1, WP_069392126.1, WP_069392251.1, WP_069392420.1, WP_069392510.1, WP_069392884.1, WP_069392982.1, WP_069392983.1, WP_069393100.1, WP_069393493.1, and WP_069393844.1, are uniquely present in members of this genus. This genus presently contains only three species (*M. trivialis*, *M. koreensis* and *M. parakoreensis*) and their genome sizes (3.89–4.08 Mbp) are among the smallest within the family *Mycobacteriaceae*. The G+C content of the two sequenced species is 69.4 mol %. Although some members of this genus have been isolated from human patients with pulmonary dysfunction, it is unclear whether they exhibit pathogenicity.

The description of *Mycolicibacillus trivialis* comb. nov. as well as the descriptions of new name combinations for other species which are part of the genus *Mycolicibacillus* are provided in **Table 9**.

Description of *Mycobacteroides* gen. nov.

Mycobacteroides (My.co.bac.te.ro´i.des. N.L. neut. n. *Mycobacterium*, a bacterial genus; L. neut. suff. *-oides*, resembling; N.L. neut. n. *Mycobacteroides*, a genus resembling *Mycobacterium*).

TABLE 9 | Descriptions of new name combinations for species in the genus *Mycolicibacillus*.

New name combinations	Description and type strain
Mycolicibacillus trivialis comb. nov. (tri.vi.a´lis. L. masc. adj. *trivialis*, common, commonplace, vulgar, ordinary, of little importance)	Basonym: *Mycobacterium triviale* Kubica 1970 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Kubica et al. (1970). The type strain is ATCC 23292 = CCUG 42431 = DSM 44153.
Mycolicibacillus koreensis comb. nov. (ko.re.en´sis. N.L. masc. adj. *koreensis*, of or pertaining to the Republic of Korea, the geographical origin of the type strain)	Basonym: *Mycobacterium koreense* Kim et al., 2012 The description of this taxon is as given by Kim et al. (2012). The type strain is 01-305 = DSM 45576 = KCTC 19819.
Mycolicibacillus parakoreensis comb. nov. (pa.ra.ko.re.en´sis. Gr. prep. *para* beside, alongside of, near, like; N.L. masc. adj. *koreensis* of or belonging to Korea, and also a bacterial specific epithet; N.L. masc. adj. *parakoreensis* near (*Mycobacterium*) *koreensis* (*koreense*)	Basonym: *Mycobacterium parakoreense* Kim et al., 2013 The description of this taxon is as given by Kim et al. (2013). The type strain is 299 = DSM 45575 = KCTC 19818.

TABLE 10 | Descriptions of new name combinations for species in the genus *Mycobacteroides*.

New name combinations	Description and type strain
Mycobacteroides abscessus comb. nov. (abs.ces´sus. L. gen. n. *abscessus*, of an abscess, referring to the ability of the organism to form abscesses)	Basonym: *Mycobacterium abscessus* (Moore and Frerichs, 1953) and Kusunoki Ezaki 1992 The description of this taxon is as given by Kusunoki and Ezaki (1992), Tortoli et al. (2013). The type strain is Hauduroy L948 = TMC 1543 = ATCC 19977 = CCUG 20993 = CIP 104536 = DSM 44196 = JCM 13569 = NCTC 13031.
Mycobacteroides abscessus subsp. *abscessus* comb. nov. (abs.ces´sus. L. gen. n. *abscessus*, of an abscess)	Basonym: *Mycobacterium abscessus* subsp. *abscessus* (Moore and Frerichs, 1953) Leao et al., 2011. The description of this taxon is as given by Leao et al. (2011); Tortoli et al. (2013). The type strain is Hauduroy L948 = TMC 1543 = ATCC 19977= CCUG 20993 = CIP 104536 = DSM 44196 = JCM 13569 = NCTC 13031.
Mycobacteroides abscessus subsp. *bolletii* comb. nov. (bol.let´i.i. N.L. gen. masc. n. *bolletii* of Bollet, to honour Claude Bollet, a famous clinical microbiologist and taxonomist)	Basonym: *Mycobacterium abscessus* subsp. *bolletii* (Adékambi et al., 2006a) Leao et al., 2011. The description of this taxon is as given by Leao et al. (2011), Tortoli et al. (2013). The type strain is BD = CCUG 50184 = CIP 108541 = JCM 15297.
Mycobacteroides abscessus subsp. *massiliense* comb. nov. (mas.si.li.en´se. L. neut. adj. *massiliense*, of the French city of Massilia, now Marseilles, France)	Basonym: *Mycobacterium abscessus* subsp. *massiliense* Tortoli et al., 2016 The description of this taxon is as given by Tortoli et al. (2016). The type strain is CCUG 48898 = CIP 108297 = KCTC 19086 = DSM 45103.
Mycobacteroides chelonae comb. nov. (che.lo´nae. Gr. n. *khelone*, a tortoise; N.L. gen. n. *chelonae*, of a tortoise)	Basonym: *Mycobacterium chelonae* Bergey et al., 1923 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Bergey et al. (1923). The type strain is CM 6388 = ATCC 35752 = CCUG 47445 = CIP 104535 = DSM 43804 = JCM 6388 = NCTC 946.
Mycobacteroides immunogenum comb. nov. (im.mu.no.ge´num. N. L. neut. adj. *immunogenum* eliciting an immune response)	Basonym: *Mycobacterium immunogenum* Wilson et al., 2001 The description of this taxon is as given by Wilson et al. (2001). The type strain is BH29 = MC 779 = ATCC 700505 = DSM 45595.
Mycobacteroides salmoniphilum comb. nov. (sal.mo.ni´phi.lum. L. n. *salmo*, *-onis* a salmon; Gr. adj. philos loving; N.L. neut. adj. *salmoniphilum* salmon-loving)	Basonym: *Mycobacterium salmoniphilum* (*ex* Ross 1960) Wilson et al., 2001 The description of this taxon is as given by Whipps et al. (2007). The type strain is SC = ATCC 13758 = DSM 43276.
Mycobacteroides franklinii comb. nov. (frank.li´ni.i. N.L. masc. gen. n.*franklinii* of Franklin, pertaining to Benjamin Franklin, famous USA statesman and scientist from Pennsylvania where the first isolates originated)	Basonym: *Mycobacterium franklinii* Nogueira et al. 2015 The description of this taxon is as given by Nogueira et al. (2015a). The type strain is DSM 45524 = ATCC BAA-2149.
Mycobacteroides saopaulense comb. nov. (sa.o.paul.en´se. N.L. neut. adj. *saopaulense* of or pertaining to the Brazilian state of São Paulo, where the first strains were isolated)	Basonym: *Mycobacterium saopaulense* Nogueira et al. 2015 The description of this taxon is as given by Nogueira et al. (2015b). The type strain is EPM 10906 = CCUG 66554 = LMG 28586 = INCQS 0733.

The type species is *Mycobacteroides abscessus*. The genus *Mycobacteriodes* is comprised of bacteria that are commonly referred to as members of the *Abscessus-Chelonae* clade. This is another genus within the family *Mycobacteriaceae* of rapidly-growing bacterial species (besides *Mycolicibacterium*) which take <7 days to form colonies. Phenotypic characteristics of this genus include a positive 3-day arylsulfatase test, better growth at 30°C than at a 35°C, negative nitrate reductase, negative iron uptake and resistance to polymyxin B (Brown-Elliott and Wallace, 2002). The genome size for the species within this clade ranges from 4.5 to 5.6 Mbp and their G+C content ranges from 63.9 to 64.8 mol %. Phylogenetic studies show that members of the genus *Mycobacteriodes* form a deep branching monophyletic clade within the family *Mycobacteriaceae* that is distinct from all other genera within this family. Some members from this genus are known to be involved in causing lung, skin and soft

TABLE 11 | Descriptions of new name combinations for species in the genus *Mycolicibacterium*.

New name combinations	Description and type strain
Mycolicibacterium fortuitum comb. nov. (for.tu´i.tum. L. neut. adj. *fortuitum*, casual, accidental, fortuitous)	Basonym: *Mycobacterium fortuitum* da Costa Cruz 1938 (Approved Lists, 1980) (Skerman et al., 1980) The description of this taxon is as given by da Costa Cruz (1938). The type strain is ATCC 6841 = CCUG 20994 = CIP 104534 = DSM 46621 = IFO (now NBRC) 13159 = JCM 6387 = NCTC 10394.
Mycolicibacterium fortuitum subsp. *acetamidolyticum* comb. nov. (a.cet.a.mi.do.ly´ti.cum. N.L. neut. n. *acetamidum*, acetamide; N.L. neut. adj. *lyticum* (from Gr. neut. adj. *lytikon*), able to loosen, able to dissolve; N.L. neut. adj. *acetamidolyticum*, digesting acetamide)	Basonym: *Mycobacterium fortuitum* subsp. *acetamidolyticum* Tsukamura et al. 1986 The description of this taxon is as given by Tsukamura et al. (1986a,b). The type strain is NCH E11620 = ATCC 35931 = CIP 105423 = DSM 44220 = JCM 6368.
Mycolicibacterium fortuitum subsp. *fortuitum* comb. nov. (for.tu´i.tum. L. neut. adj. *fortuitum*, casual, accidental, fortuitous)	Basonym: *Mycobacterium fortuitum* subsp. *fortiutum* (da Costa Cruz, 1938) Tsukamura et al. 1986 The description of this taxon is as given by da Costa Cruz (1938). The type strain is ATCC 6841 = CCUG 20994 = CIP 104534 = DSM 46621 = IFO (now NBRC) 13159 = JCM 6387 = NCTC 10394.
Mycolicibacterium agri comb. nov. (ag´ri, L. gen. n. *agri*, of a field)	Basonym: *Mycobacterium agri* (*ex* Tsukamura 1972) Tsukamura 1981 The description of this taxon is as given by Tsukamura (1981). The type strain is 90012 = ATCC 27406 = CCUG 37673 A = CIP 105391 = DSM 44515 = JCM 6377.
Mycolicibacterium aichiense comb. nov. (ai.chi.en´se. N.L. neut. adj. *aichiense*, of or belonging to Aichi prefecture, Japan)	Basonym: *Mycobacterium aichiense* (*ex* Tsukamura et al. 1973) Tsukamura 1981 The description of this taxon is as given by Tsukamura et al. (1981). The type strain is 49005 (previously, strain 5545) = ATCC 27280 = CIP 106808 = DSM 44147 = JCM 6376 = LMG 19259 = NCTC 10820.
Mycolicibacterium alvei comb. nov. (al´ve.i. L. gen. n. *alvei*, of the bed of a river, referring to the place where this species was first isolated)	Basonym: *Mycobacterium alvei* Ausina et al. 1992 The description of this taxon is as given by Ausina et al. (1992). The type strain is CR-21 = ATCC 51304 = CIP 103464 = DSM 44176 = JCM 12272.
Mycolicibacterium anyangense comb. nov. (an.yang.en´se. N.L. neut. adj. *anyangense*, pertaining to Anyang, Republic of Korea, the geographical location of the agency isolating the type strain)	Basonym: *Mycobacterium anyangense* Kim et al. 2015 The description of this taxon is as given by Kim et al. (2015). The type strain is QIA-38 = JCM 30275 = KCTC 29443.
Mycolicibacterium arabiense comb. nov. (a.ra.bi.en´se. N.L. neut. adj. *arabiense*, of or belonging to Arabia, referring to the isolation of the type strain in Dubai, United Arab Emirates)	Basonym: *Mycobacterium arabiense* Zhang et al. 2013 The description of this taxon is as given by Zhang et al. (2013a). The type strain is YIM 121001T = DSM 45768T = JCM 18538.
Mycolicibacterium arcueilense comb. nov. (ar.cueil.en´se. N.L. neut. adj. *arcueilense*, of or belonging to Arcueil, pertaining to the town of Arcueil (south of Paris) where most isolates were recovered)	Basonym: *Mycobacterium arcueilense* Konjek et al. 2016 The description of this taxon is as given by Konjek et al. (2016). The type strain is 269 = ParisRGMnew_3 = CIP 110654 = DSM 46715.
Mycolicibacterium aromaticivorans comb. nov. [a.ro.ma.ti.ci.vo´rans. L. adj. *aromaticus*, aromatic, fragrant; L. pres. part. *vorans*, devouring; N.L. part. adj. *aromaticivorans*, devouring aromatic (compounds)]	Basonym: *Mycobacterium aromaticivorans* Hennessee et al. 2009 The description of this taxon is as given by Hennessee et al. (2009). The type strain is JS19b1 = ATCC BAA-1378 = CIP 109274 = JCM 16368.
Mycolicibacterium aubagnense comb. nov. (au.bag.nen´se. N.L. neut. adj. *aubagnense*, of or pertaining to Aubagne, the city from where the first patient originated)	Basonym: *Mycobacterium aubagnense* Adékambi et al. 2006 The description of this taxon is as given by Adékambi et al. (2006a). The type strain is U8 = CCUG 50186 = CIP 108543 = JCM 15296.
Mycolicibacterium aurum comb. nov. (au´rum. L. n. *aurum*, the gold, the color of gold, intended to mean gold-pigmented)	Basonym: *Mycobacterium aurum* Tsukamura 1966 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Tsukamura (1966). The type strain is ATCC 23366 = CCUG 37666 = CIP 104465 = DSM 43999 = HAMBI 2275 = JCM 6366 = LMG 19255 = NCTC 10437 = NRRL B-4037.
Mycolicibacterium austroafricanum comb. nov. (aus.tro.a.fri.ca´num. L. adj. *australis*, southern; L. adj. *africanus*, pertaining to Africa; N.L. neut. adj. *austroafricanum*, of or pertaining to South Africa, the source of the isolates)	Basonym: *Mycobacterium austroafricanum* Tsukamura et al. 1983 The description of this taxon is as given by Tsukamura et al. (1983c). The type strain is E9789-SA12441 = ATCC 33464 = CCUG 37667 = CIP 105395 = DSM 44191 = HAMBI 2271 = JCM 6369.
Mycolicibacterium bacteremicum comb. nov. (bac.ter.e´mi.cum. N.L. n. *bacteremia*, bacteremia; N.L. neut. suff. *-(i)cum*, suffix used with the sense of pertaining to; N.L. neut. adj. *bacteremicum*, pertaining to bacteremia, referring to the organism´s association with bloodstream infections)	Basonym: *Mycobacterium bacteremicum* (Brown-Elliott et al., 2012) The description of this taxon is as given by Brown-Elliott et al. (2010). The type strain is ATCC 25791 = DSM 45578.
Mycolicibacterium boenickei comb. nov. (boe.ni´cke.i. N.L. gen. masc. n. *boenickei*, of Bönicke, in honor of the contribution of Rudolf Bönicke, a German mycobacteriologist, who first recognized the heterogeneity within the *Mycobacterium fortuitum* complex)	Basonym: *Mycobacterium boenickei* Schinsky et al. 2004 The description of this taxon is as given by Schinsky et al. (2004). The type strain is W5998 = ATCC 49935 = DSM 44677 = JCM 15653.
Mycolicibacterium brisbanense comb. nov. (bris.ban.en´se. N.L. neut. adj. *brisbanense*, of or pertaining to Brisbane, Queensland, Australia, the source of the type strain)	Basonym: *Mycobacterium brisbanense* Schinsky et al. 2004 The description of this taxon is as given by Schinsky et al. (2004). The types strain is W6743 = ATCC 49938 = CCUG 47584 = DSM 44680 = JCM 15654.

(Continued)

TABLE 11 | Continued

New name combinations	Description and type strain
Mycolicibacterium brumae comb. nov. (bru´mae. L. gen. n. *brumae*, of winter, referring to the time of year at which the first strains were isolated)	Basonym: *Mycobacterium brumae* Luquin et al. 1993 The description of this taxon is as given by Luquin et al. (1993). The type strain is CR-270 = ATCC 51384 = CCUG 37586 = CIP 103465 = DSM 44177 = JCM 12273.
Mycolicibacterium canariasense comb. nov. (ca.na.ri.as.en´se. L. neut. adj. *canariasense*, of or belonging to the Canarias (the Spanish name of the Canary Islands), where all strains were isolated)	Basonym: *Mycobacterium canariasense* (Jiménez et al., 2004) The description of this taxon is as given by Jiménez et al. (2004). The type strain is 502329 = CCUG 47953 = CIP 107998 = JCM 15298.
Mycolicibacterium celeriflavum comb. nov. (ce.le.ri.fla´vum. L. adj. *celer*, rapid; L. neut. adj. *flavum*, yellow; N.L. neut. adj. *celeriflavum*, referring to rapid growth and yellow pigmentation features of the species)	Basonym: *Mycobacterium celeriflavum* Shahraki et al. 2015 The description of this taxon is as given by Shahraki et al. (2015). The type strain is AFPC-000207 = DSM 46765 = JCM 18439.
Mycolicibacterium chitae comb. nov. (chi´tae. N.L. gen. n. *chitae*, of Chita, a place in Japan)	Basonym: *Mycobacterium chitae* Tsukamura 1967 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Tsukamura (1967b). The type strain is ATCC 19627 = CCUG 39504 = CIP 105383 = DSM 44633 = JCM 12403 = NCTC 10485.
Mycolicibacterium chlorophenolicum comb. nov. (chlo.ro.phe.no´li.cum. N.L. n. *chlorophenol*, chlorophenol; L. neut. suffix *-icum*, suffix used with the sense of pertaining to; N.L. neut. adj. *chlorophenolicum*, related to chlorophenols)	Basonym: *Mycobacterium chlorophenolicum* (Apajalahti et al., 1986) Brigila et al. 1994 The description of this taxon is as given by Apajalahti et al. (1986). The type strain is PCP-I = ATCC 49826 = CIP 104189 = DSM 43826 = HAMBI 2278 = IEGM 559 = IFO (now NBRC) 15527 = JCM 7439 = NRRL B-16528.
Mycolicibacterium chubuense comb. nov. (chu.bu.en´se. N.L. neut. adj. *chubuense*, of or belonging to Chubu, coming from soil of Chubu hospital)	Basonym: *Mycobacterium chubuense* (*ex* Tsukamura et al. 1973) Tsukamura et al. 1981 The description of this taxon is as given by Tsukamura et al. (1981). The type strain is 48013 (previously, strain 5517) = ATCC 27278 = CCUG 37670 = CIP 106810 = DSM 44219 = JCM 6374 = JCM 16420 = NCTC 10819.
Mycolicibacterium conceptionense comb. nov. (con.cep.tio.nen´se. N.L. neut. adj. *conceptionense*, of or pertaining to Hôpital de la Conception, the hospital where the first strain was isolated)	Basonym: *Mycobacterium conceptionense* Adékambi et al. 2006 The description of this taxon is as given by Adékambi et al. (2006b,c). The type strain is D16 = CCUG 50187 = CIP 108544 = JCM 15299.
Mycolicibacterium confluentis comb. nov. (con.flu.en´tis. M.L. gen. n. *confluentis*, of *Confluentes*, now Koblenz, the source of the strain on which the species description is based)	Basonym: *Mycobacterium confluentis* Kirschner et al. 1992 The description of this taxon is as given by Kirschner et al. (1992). The type strain is 1389/90 = ATCC 49920 = CIP 105510 = DSM 44017 = JCM 13671.
Mycolicibacterium cosmeticum comb. nov. (cos.me´ti.cum. N.L. neut. adj. *cosmeticum* (from Gr. adj. *kosmetikos*), referring to cosmetics)	Basonym: *Mycobacterium cosmeticum* Cooksey et al. 2004 The description of this taxon is as given by Cooksey et al. (2004). The type strain is LTA-388 = ATCC BAA-878 = CIP 108170 = JCM 14739.
Mycolicibacterium crocinum comb. nov. (cro´ci.num. L. neut. adj. *crocinum*, saffron-colored, pertaining to the colony pigmentation of known strains)	Basonym: *Mycobacterium crocinum* Hennessee et al. 2009 The description of this taxon is as given by Hennessee et al. (2009). The type strain is czh-42 = ATCC BAA-1373 = CIP 109269 = JCM 16369.
Mycolicibacterium diernhoferi comb. nov. (diern.ho´fe.ri. N.L. gen. masc. n. *diernhoferi*, of Diernhofer, who originally isolated the organisms)	Basonym: *Mycobacterium diernhoferi* (*ex* Bönicke and Juhasz 1965) Tsukamura et al. 1983 The description of this taxon is as given by Tsukamura et al. (1983c). The type strain is 41001 = ATCC 19340 = CIP 105384 = DSM 43524 = HAMBI 2269 = IFO (now NBRC) 14756 = JCM 6371.
Mycolicibacterium doricum comb. nov. (do´ri.cum. L. neut. adj. *doricum*, of or belonging to *Dorica civitas*, the ancient name of the Italian city of Ancona, from where the organism was first isolated)	Basonym: *Mycobacterium doricum* Tortoli et al., 2001 The description of this taxon is as given by Tortoli et al. (2001). The type strain is FI-13295 = CCUG 46352 = CIP 106867 = DSM 44339 = JCM 12405.
Mycolicibacterium duvalii comb. nov. (du.va´li.i. N.L. gen. masc. n. *duvalii*, of Duval, named for Professor C.W. Duval who isolated two strains of the organism)	Basonym: *Mycobacterium duvalii* Stanford and Gunthorpe1971 The description of this taxon is as given by Stanford and Gunthorpe (1971). The type strain is ATCC 43910 = CCUG 41352 = CIP 104539 = DSM 44244 = JCM 6396 = NCTC 358.
Mycolicibacterium elephantis comb. nov. (e.le.phan´tis. L. gen. n. *elephantis*, of an elephant)	Basonym: *Mycobacterium elephantis* Shojaei et al. 2000 The description of this taxon is as given by Shojaei et al. (2000). The type strain is 484 = CIP 106831 = DSM 44368 = JCM 12406.
Mycolicibacterium fallax comb. nov. (fal´lax. L. neut. adj. *fallax*, deceptive, in the sense that the colonies resemble those of *Mycobacterium tuberculosis*)	Basonym: *Mycobacterium fallax* Lévy-Frébault et al. 1983 The description of this taxon is as given by Lévy-Frébault et al. (1983). The type strain is ATCC 35219 = CCUG 37584 = CIP 81.39 = DSM 44179 = JCM 6405.
Mycolicibacterium farcinogenes comb. nov. (far.ci.no´ge.nes. Fr. n. *farcin* (from L. n. *farciminum*, a disease in horses and other animals), farcy or glanders; Gr. v. *gennaio*, produce; N.L. part. adj. *farcinogenes*, producing farcy)	Basonym: *Mycobacterium farcinogenes* Chamoiseau 1973 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Chamoiseau (1973). The type strain is IEMVT 75 = ATCC 35753 = CCUG 21047 = DSM 43637 = JCM 15463 = NCTC 10955.

(Continued)

TABLE 11 | Continued

New name combinations	Description and type strain
Mycolicibacterium flavescens comb. nov. (fla.ves´cens. L. v. *flavesco*, to become golden yellow; L. part. adj. *flavescens*, becoming yellow)	Basonym: *Mycobacterium flavescens* Bojalil et al. 1962 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Bojalil et al. (1962). The type strain is ATCC 14474 = CCUG 29041 = CIP 104533 = DSM 43991 = JCM 12274 = NCTC 10271 = NRRL B-4038.
Mycolicibacterium fluoranthenivorans comb. nov. (flu.or.an.the.ni.vo´rans. N.L. n. *fluoranthenum*, fluoranthene; L. pres. part. *vorans*, devouring; N.L. part. adj. *fluoranthenivorans*, digesting fluoranthene)	Basonym: *Mycobacterium fluoranthenivorans* Hormisch et al., 2006 The description of this taxon is as given by Hormisch et al. (2004, 2006). The type strain is FA4 = DSM 44556 = CIP 108203 = JCM 14741.
Mycolicibacterium frederiksbergense comb. nov. (fre.de.riks.ber.gen´se. N.L. neut. adj. *frederiksbergense*, of or belonging to Frederiksberg, Denmark, referring to the place of isolation)	Basonym: *Mycobacterium frederiksbergense* Willumsen et al. 2001 The description of this taxon is as given by Willumsen et al. (2001). The type strain is FAn9 = CIP 107205 = DSM 44346 = NRRL B-24126.
Mycolicibacterium gadium comb. nov. [ga´di.um. L. gen. pl. n. *gadium*, of *Gades*, the modern Cadiz (a town on the Atlantic coast of Spain)]	Basonym: *Mycobacterium gadium* Casal and Calero 1974 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Casal and Calero (1974). The type strain is ATCC 27726 = CCUG 37515 = CIP 105388 = DSM 44077 = HAMBI 2274 = JCM 12688 = NCTC 10942.
Mycolicibacterium gilvum comb. nov. (gil´vum. L. neut. adj. *gilvum*, pale yellow)	Basonym: *Mycobacterium gilvum* Stanford and Gunthorpe 1971 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Stanford and Gunthorpe (1971). The type strain is ATCC 43909 = CIP 106743 = JCM 15464 = NCTC 10742.
Mycolicibacterium goodii comb. nov. (good´i.i. N.L. gen. masc. n. *goodii*, of Good, named for Robert Good who made significant contributions to the study of mycobacteria)	Basonym: *Mycobacterium goodii* Brown et al. 1999 The description of this taxon is as given by Brown et al. (1999). The type strain is MO69 = ATCC 700504= CIP 106349 = DSM 44492 = JCM 12689.
Mycolicibacterium hassiacum comb. nov. (has.si.a´cum. M.L. neut. adj. *hassiacum*, of or belonging to *Hassia*, the German province of Hesse, where the organism was first isolated)	Basonym: *Mycobacterium hassiacum* Schröder et al. 1997 The description of this taxon is as given by Schröder et al. (1997). The type strain is 3849 = CCUG 37519 = CIP 105218 = DSM 44199 = JCM 12690.
Mycolicibacterium helvum comb. nov. (hel´vum. L. neut. adj. *helvum*, pale yellow, intended to mean pale yellow-pigmented)	Basonym: *Mycobacterium helvum* Tran and Dahl 2016 The description of this taxon is as given by Tran and Dahl (2016). The type strain is DL739 = JCM 30396 = NCCB 100520.
Mycolicibacterium hippocampi comb. nov. (hip.po.cam´pi. L. gen. n. *hippocampi*, of the seahorse)	Basonym: *Mycobacterium hippocampi* Balcázar et al. 2014 The description of this taxon is as given by Balcázar et al. (2014a,b). The type strain is BFLP-6 = DSM 45391 = LMG 25372.
Mycolicibacterium hodleri comb. nov. (hod´le.ri. N.L. gen. masc. n. *hodleri*, of Hodler, named after Christian Hodler, director of the Ministry of Science and Culture of the State of Lower Saxony, Germany, a strong supporter of natural sciences)	Basonym: *Mycobacterium hodleri* Kleespies et al. 1996 The description of this taxon is as given by Kleespies et al. (1996). The type strain is EMI2 = CIP 104909 = DSM 44183 = JCM 12141 = LMG 19253.
Mycolicibacterium holsaticum comb. nov. (hol.sa´ti.cum. M.L. neut. adj. *holsaticum*, of or belonging to *Holsatia*, the German region of Holstein, the location of the institute in which the strains were first analyzed)	Basonym: *Mycobacterium holsaticum* Richter et al. 2002 The description of this taxon is as given by Richter et al. (2002). The type strain is 1406 = CCUG 46266 = DSM 44478 = JCM 12374.
Mycolicibacterium houstonense comb. nov. (hous.ton.en´se. N.L. neut. adj. *houstonense*, of or pertaining to Houston, TX, USA, where the first isolate of the *Mycobacterium fortuitum* third biovariant (sorbitol positive) was identified)	Basonym: *Mycobacterium houstonense* Schinsky et al. 2004 The description of this taxon is as given by Schinsky et al. (2004). The type strain is W5198 = ATCC 49403 = DSM 44676 = JCM 15656.
Mycolicibacterium insubricum comb. nov. (in.su´bri.cum. L. neut. adj. *insubricum*, pertaining to *Insubria*, the Latin name of part of the Lombardy region of Italy that includes the cities in which four of the first five strains were isolated, including the type strain)	Basonym: *Mycobacterium insubricum* Tortoli et al., 2009 The description of this taxon is as given by Tortoli et al. (2009). The type strain is FI-06250 = CIP 109609 = DSM 45132 = JCM 16366.
Mycolicibacterium iranicum comb. nov. (i.ra´ni.cum. N.L. neut. adj. *iranicum*, of or belonging to Iran, isolated in Iran)	Basonym: *Mycobacterium iranicum* Shojaei et al. 2013 The description of this taxon is as given by Shojaei et al. (2013). The type strain is M05 = DSM 45541 = CCUG 62053 = JCM 17461.
Mycolicibacterium komossense comb. nov. (ko.mos.sen´se. N.L. neut. adj. *komossense*, of or belonging to Komosse sphagnum bog in south Sweden)	Basonym: *Mycobacterium komossense* Kazda and Müller 1979 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Kazda and Müller (1979). The type strain is Ko 2 = ATCC 33013 = CIP 105293 = DSM 44078 = HAMBI 2279 = HAMBI 2280 = JCM 12408.
Mycolicibacterium litorale comb. nov. (li.to.ra´le. L. neut. adj. *litorale*, of or belonging to the seashore)	Basonym: *Mycobacterium litorale* Zhang et al. 2012 The description of this taxon is as given by Zhang et al. (2012). The type strain is F4 = CGMCC 4.5724 = JCM 17423.
Mycolicibacterium llatzerense comb. nov. (llat.ze.ren´se. N.L. neut. adj. *llatzerense*, pertaining to Hospital Son Llàtzer, the hospital where the strains were isolated)	Basonym: *Mycobacterium llatzerense* Gomila et al. 2008 The description of this taxon is as given by Gomila et al. (2008). The type strain is MG13 = CCUG 54744 = CECT 7273 = JCM 16229.

(Continued)

TABLE 11 | Continued

New name combinations	Description and type strain
Mycolicibacterium lutetiense comb. nov. (lu.te.ti.en´se. N.L. neut. adj. *lutetiense*, of or belonging to Lutetia, now Paris, pertaining to the widespread distribution of this species within the Paris water distribution system)	Basonym: *Mycobacterium lutetiense* Konjek et al. 2016 The description of this taxon is as given by Konjek et al. (2016). The type strain is 071 = ParisRGMnew_1 = CIP 110656 = DSM 46713.
Mycolicibacterium madagascariense comb. nov. (ma.da.gas.car.i.en´se. N.L. neut. adj. *madagascariense*, of or belonging to the island of Madagascar, the source of the strains)	Basonym: *Mycobacterium madagascariense* Kazda et al., 1992 The description of this taxon is as given by Kazda et al. (1992). The type strain is P2 = ATCC 49865= CIP 104538 = JCM 13574.
Mycolicibacterium mageritense comb. nov. (ma.ge.ri.ten´se. N.L. neut. adj. *mageritense*, of or pertaining to *Magerit*, old (first) Arabic name of Madrid, the source of most of the isolates)	Basonym: *Mycobacterium mageritense* Domenech et al. 1997 The description of this taxon is as given by Domenech et al. (1997). The type strain is 938 = ATCC 700351 = CCUG 37984 = CIP 104973 = DSM 44476 = JCM 12375.
Mycolicibacterium malmesburyense comb. nov. (mal.mes.bu.ry.en´se. N.L. neut. adj. *malmesburyense* pertaining to Malmesbury, after a town (Malmesbury) in South Africa, where one of the isolates (the type strain) of this species originated from)	Basonym: *Mycobacterium malmesburyense* Gcebe et al. 2017 The description of this taxon is as given by Gcebe et al. (2017). The type strain is WCM 7299 = ATCC BAA-2759 = CIP 110822.
Mycolicibacterium monacense comb. nov. (mo.na.cen´se. M.L. neut. adj. *monacense*, of or belonging to *Monacum*, the Latin name of the German city Munich where the first strain was isolated)	Basonym: *Mycobacterium monacense* Reischl et al. 2006 The description of this taxon is as given by Reischl et al. (2006). The type strain is B9-21-178 = CIP 109237 = DSM 44395 = JCM 15658.
Mycolicibacterium montmartrense comb. nov. (mont.mar.tren´se. N.L. neut. adj. *montmartrense*, pertaining to the Parisian quartier of Montmartre where most isolates were recovered 'Mycobacterium sp. NL-JvlW-016' (van Ingen *et al.*, 2010))	Basonym: *Mycobacterium montmartrense* Konjek et al. 2016 The description of this taxon is as given by Konjek et al. (2016). The type strain is 196 = ParisRGMnew_2 = CIP 110655 = DSM 46714. A RGM isolate putatively belonging to this species on the basis of partial *rpoB* sequence (99% identity across 636 bp) has been reported in the Netherlands under the designation
Mycolicibacterium moriokaense comb. nov. (mo.ri.o.ka.en´se. N.L. neut. adj. *moriokaense*, of or belonging to Morioka, the locality where the species was first isolated)	Basonym: *Mycobacterium moriokaense* Tsukamura et al. 1986c The description of this taxon is as given by Tsukamura et al. (1986c). The type strain is NCH E11715 = ATCC 43059 = CCUG 37671 = CIP 105393 = DSM 44221 = JCM 6375 = VKM Ac-1183.
Mycolicibacterium mucogenicum comb. nov. (mu.co.ge´ni.cum. L. n. *mucus*, mucus, Gr. v. *gennaio*, to produce; L. neut. suff. *-icum*, suffix used with the sense of pertaining to; N.L. neut. adj. *mucogenicum*, intended to mean producing mucus, referring to the highly mucoid character of most strains on solid agar)	Basonym: *Mycobacterium mucogenicum* Springer et al. 1995 The description of this taxon is as given by Springer et al. (1995). The type strain is MO76 = ATCC 49650.
Mycolicibacterium murale comb. nov. (mu.ra´le. L. neut. adj. *murale*, of or belonging to a wall)	Basonym: *Mycobacterium murale* Vuorio et al. 1999 The description of this taxon is as given by Vuorio et al. (1999). The type strain is MA112/96 = CCUG 39728 = CIP 105980 = DSM 44340 = HAMBI 2320 = JCM 13392.
Mycolicibacterium neoaurum comb. nov. (ne.o.au´rum. Gr. adj. *neos*, new; L. n. *aurum*, gold; N.L. n. *neoaurum*, a new gold, intended to mean a new gold-pigmented organism)	Basonym: *Mycobacterium neoaurum* Tsukamura 1972 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Tsukamura (1972). The type strain is ATCC 25795 = CCUG 37665 = CIP 105387 = DSM 44074 = HAMBI 2273 = JCM 6365 = NCTC 10818.
Mycolicibacterium neworleansense comb. nov. (new.or.le.ans.en´se. N.L. neut. adj. *neworleansense*, of or pertaining to New Orleans, LA, USA, the source of the type strain)	Basonym: *Mycobacterium neworleansense* Schinsky et al. 2004 The description of this taxon is as given by Schinsky et al. (2004). The type strain is W6705 = ATCC 49404 = DSM 44679 = JCM 15659.
Mycolicibacterium novocastrense comb. nov. (no.vo.cas.tren´se. L. adj. *novus*, new; L. n. *castrum*, castle; N.L. neut. adj. *novocastrense*, of or pertaining to Newcastle, a city in the northeast of England)	Basonym: *Mycobacterium novocastrense* Shojaei et al. 1997 The description of this taxon is as given by Shojaei et al. (1997). The type strain is 73 = CIP 105546 = DSM 44203 = JCM 18114.
Mycolicibacterium obuense comb. nov. (o.bu.en´se. N.L. neut. adj. *obuense*, of or belonging to Obu, Japan)	Basonym: *Mycobacterium obuense* (*ex* Tsukamura and Mizuno 1971) Tsukamura and Mizuno 1981 The description of this taxon is as given by Tsukamura et al. (1981). The type strain is 47001 (previously, strain 4388) = ATCC 27023 = CCUG 37669 = CIP 106803 = DSM 44075 = HAMBI 2272 = JCM 6372 = NCTC 10778.
Mycolicibacterium oryzae comb. nov. (o.ry´zae. L. gen. n. *oryzae*, of rice, the origin of the type strain).	Basonym: *Mycobacterium oryzae* Ramaprasad et al. 2016 The description of this taxon is as given by Ramaprasad et al. (2016). The type strain is JC290 = KCTC 39560 = LMG 28809.
Mycolicibacterium pallens comb. nov. (pal´lens. L. neut. adj. *pallens*, pale yellow, pertaining to the colony pigmentation of the type strain)	Basonym: *Mycobacterium pallens* Hennessee et al. 2009 The description of this taxon is as given by Hennessee et al. (2009). The type strain is czh-8 = ATCC BAA-1372 = CIP 109268 = JCM 16370.
Mycolicibacterium parafortuitum comb. nov. (pa.ra.for.tu´i.tum. Gr. prep. *para*, alongside of or near; L. neut. adj. *fortuitum*, casual, accidental, and also a specific epithet; N.L. neut. adj. *parafortuitum*, alongside of (*Mycobacterium*) *fortuitum*)	Basonym: *Mycobacterium parafortuitum* Tsukamura, 1965 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Tsukamura (1965b). The type strain is ATCC 19686 = CCUG 20999 = CIP 106802 = DSM 43528 = JCM 6367 = NCTC 10411 = NRRL B-4035.

(Continued)

TABLE 11 | Continued

New name combinations	Description and type strain
Mycolicibacterium peregrinum comb. nov. (pe.re.gri´num. L. neut. adj. *peregrinum*, strange, foreign)	Basonym: *Mycobacterium peregrinum* (*ex* Bojalil et al. 1962) Kusunoki and Ezaki 1992 The description of this taxon is as given by Kusunoki and Ezaki (1992). The type strain is ATCC 14467 = CCUG 27976 = CIP 105382 = DSM 43271 = JCM 12142 = NCTC 10264.
Mycolicibacterium phlei comb. nov. (phle´i. N.L. neut. n. *Phleum*, a genus of grass, timothy; N.L. gen. n. *phlei*, of *Phleum*, of timothy)	Basonym: *Mycobacterium phlei* Lehmann and Neumann 1899 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Lehmann and Neumann (1899). The type strain is ATCC 11758 = CCUG 21000 = CIP 105389 = DSM 43239 = JCM 5865 = JCM 6385 = NCTC 8151 = NRRL B-14615 = VKM Ac-1291.
Mycolicibacterium phocaicum comb. nov. (pho.ca´i.cum. L. neut. adj. *phocaicum*, Phocœan, referred to Phocaea, a maritime town of Ionia, a colony of the Athenians, whose inhabitants fled, to escape from Persian domination, and founded Massilia (Marseille), which was the source of the type strain)	Basonym: *Mycobacterium phocaicum* Adékambi et al. 2006 The description of this taxon is as given by Adékambi et al. (2006a). The type strain is N4 = CCUG 50185 = CIP 108542 = JCM 15301.
Mycolicibacterium porcinum comb. nov. (por.ci.num. L. neut. adj. *porcinum*, pertaining to swine)	Basonym: *Mycobacterium porcinum* Tsukamura et al. 1983 The description of this taxon is as given by Tsukamura et al. (1983b). The type strain is E10241-1 = ATCC 33776 = CCUG 37674 = CIP 105392 = DSM 44242 = JCM 6378.
Mycolicibacterium poriferae comb. nov. (po.ri´fe.rae. N.L. gen. *poriferae* of the *Porifera*, the phylum of sponges)	Basonym: *Mycobacterium poriferae* Padgitt and Moshier 1987 The description of this taxon is as given by Padgitt and Moshier (1987). The type strain is 47 = ATCC 35087 = CIP 105394 = JCM 12603.
Mycolicibacterium psychrotolerans comb. nov. (psy.chro.to´le.rans. Gr. adj. *psychros*, cold; L. pres. part. *tolerans*, tolerating; N.L. part. adj. *psychrotolerans*, cold-tolerating)	Basonym: *Mycobacterium psychrotolerans* Trujillo et al. 2004 The description of this taxon is as given by Trujillo et al. (2004). The type strain is WA101 = DSM 44697 = JCM 13323 = LMG 21953.
Mycolicibacterium pulveris comb. nov. (pul´ve.ris. L. gen. n. *pulveris*, of dust, referring to the source, house dust)	Basonym: *Mycobacterium pulveris* Tsukamura et al. 1983 The description of this taxon is as given by Tsukamura et al. (1983a). The type strain is NCH 33505 = ATCC 35154 = CCUG 37668 = CIP 106804 = DSM 44222 = JCM 6370.
Mycolicibacterium pyrenivorans comb. nov. (py.re.ni.vo´rans. N.L. n. *pyrenum*, pyrene; L. pres. part. *vorans*, devouring, destroying; N.L. part. adj. *pyrenivorans*, destroying pyrene)	Basonym: *Mycobacterium pyrenivorans* Derz et al. 2004 The description of this taxon is as given by Derz et al. (2004). The type strain is 17A3 = DSM 44605 = JCM 15927 = NRRL B-24349.
Mycolicibacterium rhodesiae comb. nov. (rho.de.si´ae. N.L. gen. n. *rhodesiae*, of/from Rhodesia)	Basonym: *Mycobacterium rhodesiae* (*ex* Tsukamura et al. 1971) Tsukamura et al. 1981 The description of this taxon is as given by Tsukamura et al. (1981). The types strain is 02002 (previously, strain 5295) = ATCC 27024 = CIP 106806 = DSM 44223 = JCM 6363 = NCTC 10779.
Mycolicibacterium rufum comb. nov. (ru´fum. L. neut. adj. *rufum* ruddy or red, pertaining to the colony pigmentation of the type strain)	Basonym: *Mycobacterium rufum* Hennessee et al. 2009 The description of this taxon is as given by Hennessee et al. (2009). The type strain is JS14 = ATCC BAA-1377 = CIP 109273 = JCM 16372.
Mycolicibacterium rutilum comb. nov. (ru´ti.lum. L. neut. adj. *rutilum*, rust-colored, pertaining to the colony pigmentation of known strains)	Basonym: *Mycobacterium rutilum* Hennessee et al. 2009 The description of this taxon is as given by Hennessee et al. (2009). The type strain is czh-117 = ATCC BAA-1375 = CIP 109271 = JCM 16371.
Mycolicibacterium sarraceniae comb. nov. (sar.ra.ce´ni.ae. N.L. fem. gen. n. *sarraceniae*, of *Sarracenia*, for the pitcher plant from where the species was isolated)	Basonym: *Mycobacterium sarraceniae* Tran and Dahl 2016 The description of this taxon is as given by Tran and Dahl (2016). The type strain is DL734 = JCM 30395 = NCCB 100519.
Mycolicibacterium sediminis comb. nov. (se.di´mi.nis. L. gen. n. *sediminis*, of a sediment)	Basonym: *Mycobacterium sediminis* Zhang et al. 2013 The description of this taxon is as given by Zhang et al. (2013a). The type strain is YIM M13028 = DSM 45643 = KCTC 19999.
Mycolicibacterium senegalense comb. nov. (se.ne.gal.en´se. N.L. neut. adj. *senegalense*, of or belonging to the West African Republic of Senegal)	Basonym: *Mycobacterium senegalense* (Chamoiseau, 1973) Chamoiseau 1979 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Chamoiseau (1973, 1979). The type strain is IEMVT 378 = ATCC 35796 = CCUG 21001 = CIP 104941 = DSM 43656= JCM 15467 = NCTC 10956.
Mycolicibacterium septicum comb. nov. (sep´ti.cum. L. neut. adj. *septicum*, producing a putrefaction, putrefying, septic, referring to the isolation of the organism from blood)	Basonym: *Mycobacterium septicum* Schinsky et al. 2000 The description of this taxon is as given by Schinsky et al. (2000). The type strain is W4964 = ATCC 700731 = CCUG 43574 = CIP 106642 = DSM 44393 = JCM 14743.
Mycolicibacterium setense comb. nov. (se.ten´se. N.L. neut. adj. *setense*, pertaining to Sète (France), the city from which the infected patient originated)	Basonym: *Mycobacterium setense* Lamy et al. 2008 The description of this taxon is as given by Lamy et al. (2008). The type strain is ABO-M06 = CIP 109395 = DSM 45070 = JCM 15660.

(Continued)

TABLE 11 | Continued

New name combinations	Description and type strain
Mycolicibacterium smegmatis comb. nov. (smeg.ma´tis. L. n. *smegma -atis*, an unguent (for making the skin smooth), a detergent, a cleansing medicine, and in biology the sebaceous humor; L. gen. n. *smegmatis*, of smegma)	Basonym: *Mycobacterium smegmatis* (Trevisan, 1889) Lehmann and Neumann, 1899 (Approved Lists, 1980) (Lehmann and Neumann, 1899) The description of this taxon is as given by Trevisan (1889); Lehmann and Neumann (1899). The type strain is ATCC 19420 = CCUG 21002 = CCUG 21815 = CIP 104444 = DSM 43756 = JCM 5866 = JCM 6386 = NCTC 8159 = NRRL B-14616 = VKM Ac-1239.
Mycolicibacterium sphagni comb. nov. (sphag´ni. N.L. n. *Sphagnum*, generic name of the moss of sphagnum bogs, the habitat of these strains; N.L. gen. n. *sphagni*, of *Sphagnum*)	Basonym: *Mycobacterium sphagni* Kazda 1980 The description of this taxon is as given by Kazda (1980). The type strain is Sph 38 = ATCC 33027 = DSM 44076.
Mycolicibacterium thermoresistibile comb. nov. (ther.mo.re.sis.ti´bi.le. Gr. n. *thermê*, heat; L. v. *resisto*, to stand back, remain standing, endure; L. neut. suff. *-ile*, suffix denoting an active quality, able to; N.L. neut. adj. *thermoresistibile*, able to resist to high temperature)	Basonym: *Mycobacterium thermoresistibile* Tsukamura 1966 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Tsukamura (1966). The type strain is ATCC 19527 = CCUG 28008 = CCUG 41353 = CIP 105390 = DSM 44167 = JCM 6362 = NCTC 10409.
Mycolicibacterium tokaiense comb. nov. (to.kai.en´se. N.L. neut. adj. *tokaiense*, of or belonging to Tokai district of Japan)	Basonym: *Mycobacterium tokaiense* (*ex* Tsukamura et al. 1973) Tsukamura 1981 The description of this taxon is as given by Tsukamura et al. (1981). The type strain is 47503 (previously, strain 5553) = ATCC 27282 = CIP 106807 = DSM 44635 = JCM 6373 = NCTC 10821.
Mycolicibacterium tusciae comb. nov. (tus´ci.ae. L. gen. n. *tusciae*, of Tuscia (the country of the Tuscans), now Tuscany, the Italian region where all the organisms were isolated)	Basonym: *Mycobacterium tusciae* Tortoli et al., 1999 The description of this taxon is as given by Tortoli et al. (1999). The type strain is FI-25796 = CCUG 50996 = CIP 106367 = DSM 44338 = JCM 12692.
Mycolicibacterium vaccae comb. nov. (vac´cae. L. gen. n. *vaccae*, of a cow)	Basonym: *Mycobacterium vaccae* Bönicke and Juhasz 1964 (Approved Lists 1980) (Skerman et al., 1980) The description of this taxon is as given by Bonicke and Juhasz (1964). The type strain is ATCC 15483 = CCUG 21003 = CIP 105934 = DSM 43292 = HAMBI 2276 = IFO (now NBRC) 14118 = JCM 6389 = NCTC 10916.
Mycolicibacterium vanbaalenii comb. nov. (van.baa.len´i.i. N.L. gen. masc. n. *vanbaalenii*, of Van Baalen, in memory of Chase Van Baalen, late Professor at The University of Texas Marine Science Institute, Port Aransas Marine Laboratory, Port Aransas, Texas, USA)	Basonym: *Mycobacterium vanbaalenii* Khan et al. 2002 The description of this taxon is as given by Khan et al. (2002). The type strain is PYR-1 = DSM 7251 = JCM 13017 = NRRL B-24157.
Mycolicibacterium vulneris comb. nov. (vul´ne.ris. L. gen. n. *vulneris*, of a wound, from which the type strain was isolated)	Basonym: *Mycobacterium vulneris* van Ingen et al. 2009 The description of this taxon is as given by van Ingen et al. (2009). The type strain is NLA000700772 = CIP 109859 = DSM 45247 = JCM 18115.
Mycolicibacterium wolinskyi comb. nov. (wo.lins´ky.i. N.L. masc. gen. n. *wolinskyi*, of Wolinsky, named for Emanuel Wolinsky for his significant contributions to the study of non-tuberculous mycobacteria)	Basonym: *Mycobacterium wolinskyi* Brown et al. 1999 The description of this taxon is as given by Brown et al. (1999). The type strain is MO739 = ATCC 700010 = CCUG 47168 = CIP 106348 = DSM 44493 = JCM 13393.

tissue infections (Magee and Ward, 2012; Tortoli, 2014) and some exhibit resistance to multiple antimicrobial drugs (Nessar et al., 2012).

The members of the genus *Mycobacteriodes* can be reliably distinguished from all other *Mycobacteriaceae* species as well as other bacteria based upon unique shared presence of 27 CSIs in different proteins listed in **Table 3** (viz. uracil phosphoribosyltransferase, L-histidine N(alpha)-methyltransferase, DUF58 domain-containing protein, NADH-quinone oxidoreducatase subunit G, ATP-dependent helicase, tRNA (cytidine(34)-2′-O)-methyltransferase, glutamine -fructose-6-phosphate transaminase (isomerizing), error-prone DNA polymerase, 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase, DEAD/DEAH box helicase, anion transporter, a membrane protein, nicotinate-nucleotide adenylyltransferase, CoA ester lyase, bifunctional ADP-dependent (S)-NAD(P)H-hydrate dehydratase/NAD(P)H-hydrate epimerase, pyridoxal phosphate-dependent aminotransferase, carotenoid oxygenase, SAM-dependent methyltransferase, phosphoribosylamine-glycine ligase, and hypothetical proteins) and the presence of 24 conserved signature proteins listed in **Table 2**, (viz. MAB_0188c, MAB_0375, MAB_0601, MAB_2852c, MAB_3058, MAB_3079c, MAB_1107c, MAB_1519, MAB_1642, MAB_0008, MAB_0245c, MAB_2487, MAB_3020c, MAB_1440c, MAB_0014, MAB_0015, MAB_0345, MAB_0448c, MAB_0456, MAB_0460, MAB_2549, MAB_1765, MAB_1767, and MAB_1806) that are also specifically found in these bacteria.

The description of *Mycobacteroides abscessus* comb. nov. as well as the descriptions of new name combinations for other species which are part of the genus *Mycobacteroides* are provided in **Table 10**.

Description of *Mycolicibacterium* gen. nov.

Mycolicibacterium (My.co.li.ci.bac.te´ri.um. N.L. n. *acidum mycolicum*, mycolic acid; N.L. neut. n. *bacterium*, a small rod; N.L. neut. n. *Mycolicibacterium*, a genus of mycolic acid containing rod-shaped bacteria).

The type species *Mycolicibacterium fortuitum*.

The genus is comprised of rapidly-growing bacterial species, which take <7 days to form colonies upon primary isolation (Parte, 2014). Some other phenotypic characteristics generally common to the members of this genus include absence of pigmentation, positive 3-day arylsulfatase activity (Brown-Elliott and Wallace, 2002), positive for nitrate reductase and iron uptake (Magee and Ward, 2012). Most species are saprophytic and considered non-pathogenic to humans, however some cases of infections and diseases by members of this group have been reported (Stahl and Urbance, 1990; Brown-Elliott and Wallace, 2002; Ripoll et al., 2009). The members of this genus form a monophyletic clade in phylogenetic trees based on concatenated sequences of multiple large datasets of conserved proteins including a tree based on 1941 core proteins from mycobacterial genomes, a tree based on 136 core proteins for the phylum Actinobacteria, and another tree based on concatenated sequences for 8 conserved proteins described in the present study.

The members of the genus *Mycolicibacterium* can be distinguished from other genera within the family *Mycobacteriaceae* as well as other bacteria based upon conserved signature indels in the following four proteins viz. LacI family transcriptional regulator, Cyclase, CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase and CDP-diacylglycerol–serine O-phosphatidyltransferase (**Table 4**) that are uniquely shared by the members of this genus. Additionally, the homologs of the 10 conserved signature proteins, whose accession numbers are as follows (WP_048630777.1, WP_048632025.1, WP_048632497.1, WP_048634851.1, WP_048633467.1, WP_048633322.1 WP_048631132.1, WP_048634509.1, WP_048630657.1, and WP_048632441.1) are also uniquely found in the members of this genus (**Table 5**). The genome size for the members of this genus ranges from 3.95 to 8.0 Mbp and their G+C content ranges from 65.4 to 70.3 mol %.

The description of *Mycolicibacterium fortuitum* comb. nov. as well as the descriptions of new name combinations for other species which are part of the genus *Mycolicibacterium* are provided in **Table 11**.

In addition to the new name combinations for species which are part of this genus, we also provide below description of two new species that should also be placed in the genus *Mycolicibacterium*.

Description of *Mycolicibacterium acapulense* sp. nov. (a.ce.pul.cen´se. N.L. neut. adj. *acapulcense* from Acapulco, a town on the Pacific coast of México).

The description of this taxon is as given by Bojalil et al. (1962) for *"Mycobacterium acapulensis"*. The type strain is AC-103 (= ATCC 14473 = JCM 6402).

Description of *Mycolicibacterium komanii* sp. nov. (ko.ma´ni.i. N.L. gen. n. *komanii* named after a town in South Africa where one of the isolates originated from, Komani is the Xhosa name for Queenstown (South Africa)).

The description of this taxon is as given by Gcebe (2015) and Gcebe et al. (2016) for *"Mycobacterium komanii"*. The type strain is GPK 1020.

AUTHOR CONTRIBUTIONS

RG was responsible for conceiving the idea of this study, carried out phylogenomic and other analyses reported here, supervised and directed the entire project and obtained funds for carrying out these studies. Involved in the writing and finalizing of the manuscript and all presented data. BL and JS were responsible for analysis and organization of the comparative genomic data on identification of described molecular signatures, under the direction of RG. They also helped in the preparation of a draft version of the manuscript.

ACKNOWLEDGMENTS

This work was supported by the research grant No. 249924 from the Natural Science and Engineering Research Council of Canada awarded to RG. We thank T. Vijaykumar for carrying out preliminary work in this regard. Lastly, we express our sincere thanks and the deepest appreciation to Professor Aharon Oren for his valuable input/suggestions regarding the correct etymology and protologues for the names of newly proposed taxa and the new name combinations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00067/full#supplementary-material

REFERENCES

Adékambi, T., and Drancourt, M. (2004). Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, hsp65, sodA, recA and rpoB gene sequencing. *Int. J. Syst. Evol. Microbiol.* 54, 2095–2105. doi: 10.1099/ijs.0.63094-0

Adékambi, T., Berger, P., Raoult, D., and Drancourt, M. (2006a). rpoB gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. *Int. J. Syst. Evol. Microbiol.* 56, 133–143. doi: 10.1099/ijs.0.63969-0

Adékambi, T., Stein, A., Carvajal, J., Raoult, D., and Drancourt, M. (2006b). Description of *Mycobacterium conceptionense* sp. nov., a *Mycobacterium fortuitum* group organism isolated from a posttraumatic osteitis inflammation. *J. Clin. Microbiol.* 44, 1268–1273. doi: 10.1128/JCM.44.4.1268-1273.2006

Adékambi, T., Stein, A., Carvajal, J., Raoult, D., and Drancourt, M. (2006c). *Mycobacterium conceptionense* sp. nov. in: list of new names and new combinations previously effectively, but not Validly, Published. List no. 111. *Int. J. Syst. Evol. Microbiol.* 56, 2025–2027. doi: 10.1099/ijs.0.64643-0

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, S. (2016). Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* 66, 5575–5599. doi: 10.1099/ijsem.0.001485

Adeolu, M., and Gupta, R. S. (2014). A phylogenomic and molecular marker based proposal for the division of the genus *Borrelia* into two genera: the emended

genus *Borrelia* containing only the members of the relapsing fever *Borrelia*, and the genus *Borreliella* gen. nov. containing the members of the Lyme disease Borrelia (Borrelia burgdorferi sensu lato complex). *Antonie van Leeuwenhoek* 105, 1049–1072. doi: 10.1007/s10482-014-0164-x

Ahmod, N. Z., Gupta, R. S., and Shah, H. N. (2011). Identification of a *Bacillus anthracis* specific indel in the *yeaC* gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J. Microbiol. Methods* 87, 278–285. doi: 10.1016/j.mimet.2011.08.015

Alnajar, S., and Gupta, R. S. (2017). Phylogenomics and comparative genomic studies delineate six main clades within the family *Enterobacteriaceae* and support the reclassification of several polyphyletic members of the family. *Infect. Genet. Evol.* 54, 108–127. doi: 10.1016/j.meegid.2017.06.024

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and 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

Apajalahti, J. H. A., Karpanoja, P., and Salkinoja-Salonen, M. S. (1986). *Rhodococcus chlorophenolicus* sp. nov., a chlorophenol-mineralizing actinomycete. *Int. J. Syst. Evol. Microbiol.* 36, 246–251. doi: 10.1099/00207713-36-2-246

Ausina, V., Luquin, M., García-Barceló, M., Laneelle, M. A., Lévy-Frébault, V. V., Belda, F., and et al. (1992). *Mycobacterium alvei* sp. nov. *Int. J. Syst. Bacteriol.* 42, 529–535. doi: 10.1099/00207713-42-4-529

Balcázar, J. L., Planas, M., and Pintado, J. (2014a). *Mycobacterium hippocampi* sp. nov. in Oren A., Garrity G.M. (editors). List of new names and new combinations previously effectively, but not validly, published, list no. 160. *Int. J. Syst. Evol. Microbiol. 64*, 3603–3606. doi: 10.1099/ijsem.0.001173

Balcázar, J. L., Planas, M., and Pintado, J. (2014b). *Mycobacterium hippocampi* sp. nov., a rapidly growing scotochromogenic species isolated from a seahorse with tail rot. *Curr. Microbiol.* 69, 329–333. doi: 10.1007/s00284-014-0588-6

Baldauf, S. L., and Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11558–11562. doi: 10.1073/pnas.90.24.11558

Barbour, A. G., Adeolu, M., and Gupta, R. S. (2017). Division of the genus *Borrelia* into two genera (corresponding to Lyme disease and relapsing fever groups) reflects their genetic and phenotypic distinctiveness and will lead to a better understanding of these two groups of microbes. *Int. J. Syst. Evol. Microbiol.* 67, 2058–2067. doi: 10.1099/ijsem.0.001815

Bergey, D. H., Harrison, F. C., Breed, R. S., Hammer, B. W., and Huntoon, F. M. (1923). *Bergey's Manual of Determinative Bacteriology.* Baltimore, MD: The Williams and Wilkins Co.

Bhandari, V., Ahmod, N. Z., Shah, H. N., and Gupta, R. S. (2013). Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. *Int. J. Syst. Evol. Microbiol.* 63, 2712–2726. doi: 10.1099/ijs.0.048488-0

Bhandari, V., and Gupta, R. S. (2014). Molecular signatures for the phylum (class) *Thermotogae* and a proposal for its division into three orders (*Thermotogales*, *Kosmotogales* ord. *nov.* and *Petrotogales* ord. nov.) containing four families *(Thermotogaceae, Fervidobacteriaceae* fam. nov., *Kosmotogaceae* fam. nov. and *Petrotogaceae* fam. nov.) and a new genus *Pseudothermotoga gen.* nov. with five new combinations. *Antonie van Leeuwenhoek* 105, 143–168. doi: 10.1007/s10482-013-0062-7

Bhandari, V., Naushad, H. S., and Gupta, R. S. (2012). Protein based molecular markers provide reliable means to understand prokaryotic phylogeny and support Darwinian mode of evolution. *Front. Cell Infect. Microbiol.* 2:98. doi: 10.3389/fcimb.2012.00098

Bojalil, L. F., Cerbon, J., and Trujillo, A. (1962). Adansonian classification of mycobacteria. *J. Gen. Microbiol.* 28, 333–346. doi: 10.1099/00221287-28-2-333

Bonicke, R., and Juhasz, E. (1964). Description of the new species *Mycobacterium vaccae* n. sp. Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. 1. Abt. Medizinisch-hygienische Bakteriologie, Virusforschung und Parasitologie. *Originale* 192, 133–135.

Brown, B. A., Springer, B., Steingrube, V. A., Wilson, R. W., Pfyffer, G. E., Garcia, M. J., et al. (1999). *Mycobacterium wolinskyi* sp. nov. and *Mycobacterium goodii* sp. nov., two new rapidly growing species related to *Mycobacterium smegmatis* and associated with human wound infections: a cooperative study from the International Working Group on Mycobacterial Taxonomy. *Int. J. Syst. Bacteriol.* 49, 1493–1511. doi: 10.1099/00207713-49-4-1493

Brown-Elliott, B. A., and Wallace, R. J. (2002). Clinical and taxonomic status of pathogenic nonpigmented or late-pigmenting rapidly growing mycobacteria. *Clin. Microbiol. Rev.* 15, 716–746. doi: 10.1128/CMR.15.4.716-746.2002

Brown-Elliott, B. A., Wallace, R. J. Jr., Petti, C. A., Mann, L. B., McGlasson, M., Chihara, S., and et al. (2010). *Mycobacterium neoaurum* and *Mycobacterium bacteremicum* sp. *nov.* as causes of mycobacteremia. *J. Clin. Microbiol.* 48, 4377–4385. doi: 10.1128/JCM.00853-10

Brown-Elliott, B. A., Wallace, R. J. Jr., Petti, C. A., Mann, L. B., McGlasson, M., Chihara, S., et al. (2012). *Mycobacterium bacteremicum* sp. nov. in: list of new names and new combinations previously effectively, but not validly, published, list no. 148. *Int. J. Syst. Evol. Microbiol.* 62, 2549–2554. doi: 10.1099/ijs.0.048033-0

Brzostek, A., Pawelczyk, J., Rumijowska-Galewicz, A., Dziadek, B., and Dziadek, J. (2009). *Mycobacterium tuberculosis* is able to accumulate and utilize cholesterol. *J. Bacteriol.* 191, 6584–6591. doi: 10.1128/JB.00488-09

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Casal, M., and Calero, J. R. (1974). *Mycobacterium gadium* sp. nov. a new species of rapid-growing scotochromogenic mycobacteria. *Tubercle* 55, 299–308. doi: 10.1016/0041-3879(74)90039-7

Chamoiseau, G. (1973). *Mycobacterium farcinogenes*, agent causal du farcin du boeuf en Afrique. *Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux* 27, 61–65. doi: 10.19182/remvt.7975

Chamoiseau, G. (1979). Etiology of farcy in African bovines. *Int. J. Syst. Evol. Microbiol.* 29, 407–410.

Chandra, G., and Chater, K. F. (2014). Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* 38, 345–379. doi: 10.1111/1574-6976.12047

Cloud, J. L., Meyer, J. J., Pounder, J. I., Jost, K. C. Jr., Sweeney, A., Carroll, K. C., et al. (2006). *Mycobacterium arupense* sp. nov., a non-chromogenic bacterium isolated from clinical specimens. *Int. J. Syst. Evol. Microbiol.* 56, 1413–1418. doi: 10.1099/ijs.0.64194-0

Cooksey, R. C., de Waard, J. H., Yakrus, M. A., Rivera, I., Chopite, M., Toney, S. R., et al. (2004). *Mycobacterium cosmeticum* sp. nov., a novel rapidly growing species isolated from a cosmetic infection and from a nail salon. *Int. J. Syst. Evol. Microbiol.* 54, 2385–2391. doi: 10.1099/ijs.0.63238-0

da Costa Cruz, J. C. (1938). *Mycobacterium fortuitum*, un novo bacilo acido-resistente pathogenico para o human. *Acta Med.* 1, 297–301.

Derz, K., Klinner, U., Schuphan, I., Stackebrandt, E., and Kroppenstedt, R. M. (2004). *Mycobacterium pyrenivorans* sp. nov., a novel polycyclic-aromatic-hydrocarbon-degrading species. *Int. J. Syst. Evol. Microbiol.* 54, 2313–2317. doi: 10.1099/ijs.0.03003-0

Devulder, G., Perouse de Montclos, M., and Flandrois, J. P. (2005). A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int. J. Syst. Evol. Microbiol.* 55, 293–302. doi: 10.1099/ijs.0.63222-0

Djouadi, L. N., Levasseur, A., Khalil, J. B., Blanc-Taileur, C., Asmar, S., Ghiloubi, W., et al. (2016). *Mycobacterium icosiumassiliensis* sp. nov., a new member in the Mycobacterium terrae complex isolated from surface water in Algeria. *Curr. Microbiol.* 73, 255–264. doi: 10.1007/s00284-016-1062-4

Domenech, P., Jimenez, M. S., Menendez, M. C., Bull, T. J., Samper, S., Manrique, A., et al. (1997). *Mycobacterium mageritense* sp. nov. *Int. J. Syst. Bacteriol.* 47, 535–540. doi: 10.1099/00207713-47-2-535

Dutilh, B. E., Snel, B., Ettema, T. J., and Huynen, M. A. (2008). Signature genes as a phylogenomic tool. *Mol. Biol. Evol.* 25, 1659–1667. doi: 10.1093/molbev/msn115

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Fahey, R. C. (2001). Novel thiols of prokaryotes. *Annu. Rev. Microbiol.* 55, 333–356. doi: 10.1146/annurev.micro.55.1.333

Falkinham, J. O. (2009). Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. *J. Appl. Microbiol.* 107, 356–367. doi: 10.1111/j.1365-2672.2009.04161.x

Fedrizzi, T., Meehan, C. J., Grottola, A., Giacobazzi, E., Fregni, S. G., Tagliazucchi, S., et al. (2017). Genomic characterization of Nontuberculous Mycobacteria. *Sci. Rep.* 7:45258. doi: 10.1038/srep45258

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gao, B., and Gupta, R. S. (2005). Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int. J. Syst. Evol. Microbiol*. 55, 2401–2412. doi: 10.1099/ijs.0.63785-0

Gao, B., and Gupta, R. S. (2012). Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol. Mol. Biol. Rev*. 76, 66–112. doi: 10.1128/MMBR.05011-11

Gao, B., Paramanathan, R., and Gupta, R. S. (2006). Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. *Antonie van Leeuwenhoek* 90, 69–91. doi: 10.1007/s10482-006-9061-2

Gcebe, N. (2015). *The Occurrence and Molecular Characterization of Non-tuberculous Mycobacteria in Cattle, African Buffalo (Syncerus caffer) and Their Environments in South Africa and Genomic Characterization and Proteomic Comparison with Mycobacterium bovis*. Ph.D. thesis, University of Pretoria, Pretoria, South Africa.

Gcebe, N., Michel, A., Gey van Pittius, N. C., and Rutten, V. (2016). Comparative genomics and proteomic analysis of four non-tuberculous *Mycobacterium* species and *Mycobacterium tuberculosis* complex: occurrence of shared immunogenic proteins. *Front. Microbiol*. 7:795. doi: 10.3389/fmicb.2016.00795

Gcebe, N., Rutten, V., Pittius, N. G. V., Naicker, B., and Michel, A. (2017). *Mycobacterium malmesburyense* sp. nov., a non-tuberculous species of the genus *Mycobacterium* revealed by multiple gene sequence characterization. *Int. J. Syst. Evol. Microbiol*. 67, 832–838. doi: 10.1099/ijsem.0.001678

Gomila, M., Ramirez, A., Gasco, J., and Lalucat, J. (2008). *Mycobacterium llatzerense* sp. nov., a facultatively autotrophic, hydrogen-oxidizing bacterium isolated from haemodialysis water. *Int. J. Syst. Evol. Microbiol*. 58, 2769–2773. doi: 10.1099/ijs.0.65857-0

Goodfellow, M., and Magee, G. M. (1998). "Taxonomy of Mycobacteria," in *Mycobacteria,* eds P. A. Gangadharam and P. A. Jenkins (New York, NY: Chapman and Hall), 1–71.

Guillemin, I., Cambau, E., and Jarlier, V. (1995). Sequences of conserved region in the A subunit of DNA gyrase from nine species of the genus *Mycobacterium*: phylogenetic analysis and implication for intrinsic susceptibility to quinolones. *Antimicro. Agents Chemother*. 39, 2145–2149. doi: 10.1128/AAC.39.9.2145

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol*. 59, 307–321. doi: 10.1093/sysbio/syq010

Gupta, R. S. (1998). Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among Archaebacteria, Eubacteria, and Eukaryotes. *Microbiol. Mol. Biol. Rev*. 62, 1435–1491.

Gupta, R. S. (2014). "Identification of conserved indels that are useful for classification and evolutionary studies," in *Methods in Microbiology: New Approaches to Prokaryotic Systematics,* eds M. Goodfellow, I. Sutcliffe, and J. Chun (Oxford: Academic Press), 153–182.

Gupta, R. S. (2016a). Editorial: Applications of genome sequences for discovering characteristics that are unique to different groups of organisms and provide insights into evolutionary relationships. *Front. Genet*. 7:27. doi: 10.3389/fgene.2016.00027

Gupta, R. S. (2016b). Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. *FEMS Microbiol. Rev*. 40, 520–553. doi: 10.1093/femsre/fuw011

Gupta, R. S. (2016c). Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates. *Mol. Phylogenet. Evol*. 94, 383–391. doi: 10.1016/j.ympev.2015.09.019

Gupta, R. S., Chander, P., and George, S. (2013a). Phylogenetic framework and molecular signatures for the class *Chloroflexi* and its different clades; proposal for division of the class *Chloroflexia* class. nov. [corrected] into the suborder *Chloroflexineae* subord. nov., consisting of the emended family Oscillochloridaceae and the family *Chloroflexaceae* fam. nov., and the suborder *Roseiflexineae* subord. nov., containing the family *Roseiflexaceae* fam. nov. *Antonie van Leeuwenhoek* 103, 99–119. doi: 10.1007/s10482-012-9790-3

Gupta, R. S., Chen, W. J., Adeolu, M., and Chai, Y. (2013b). Molecular signatures for the class *Coriobacteriia* and its different clades; Proposal for division of the class *Coriobacteriia* into the emended order *Coriobacteriales*, containing the emended family *Coriobacteriaceae* and *Atopobiaceae* fam. nov., and *Eggerthellales* ord. nov., containing the family *Eggerthellaceae* fam. nov. *Int. J. Syst. Evol. Microbiol*. 63, 3379–3397. doi: 10.1099/ijs.0.048371-0

Gupta, R. S., Naushad, S., and Baker, S. (2015). Phylogenomic analyses and molecular signatures for the class *Halobacteria* and its two major clades: a proposal for division of the class *Halobacteria* into an emended order *Halobacteriales* and two new orders, Haloferacales ord. *nov*. and *Natrialbales* ord. nov., containing the novel families *Haloferacaceae* fam. nov. and *Natrialbaceae* fam. nov. *Int. J. Syst. Evol. Microbiol*. 65, 1050–1069. doi: 10.1099/ijs.0.070136-0

Hamada, M., Shibata, C., Sakurai, K., Hosoyama, A., Oji, S., Teramoto, K., et al. (2016). Reclassification of *Amycolicicoccus subflavus* as *Hoyosella subflava* comb. nov. and emended descriptions of the genus *Hoyosella* and *Hoyosella altamirensis. Int. J. Syst. Evol. Microbiol*. 66, 4711–4715. doi: 10.1099/ijsem.0.001415

Hannigan, G. D., Krivogorsky, B., Fordice, D., Welch, J. B., and Dahl, J. L. (2013). *Mycobacterium minnesotense* sp. nov., a photochromogenic bacterium isolated from sphagnum peat bogs. *Int. J. Syst. Evol. Microbiol*. 63, 124–128. doi: 10.1099/ijs.0.037291-0

Hartmans, S., de Bont, J. A. M., and Stackebrandt, E. (2006). "Chapter 1.1.18 - The genus *Mycobacterium*—nonmedical," in *The Prokaryotes: Volume 3: Archaea. Bacteria: Firmicutes, Actinomycetes,* eds M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (New York, NY: Springer New York), 889–918.

Hennessee, C. T., Seo, J. S., Alvarez, A. M., and Li, Q. X. (2009). Polycyclic aromatic hydrocarbon-degrading species isolated from Hawaiian soils: *Mycobacterium crocinum* sp. nov., *Mycobacterium pallens* sp. nov., *Mycobacterium rutilum* sp. nov., *Mycobacterium rufum* sp. nov. and *Mycobacterium aromaticivorans* sp. nov. *Int. J. Syst. Evol. Microbiol*. 59, 378–387. doi: 10.1099/ijs.0.65827-0

Hormisch, D., Brost, I., Kohring, G. W., Giffhorn, F., Kroppenstedt, R. M., Stackebrandt, E., et al. (2004). *Mycobacterium fluoranthenivorans* sp. nov., a fluoranthene and aflatoxin B1 degrading bacterium from contaminated soil of a former coal gas plant. *Syst. Appl. Microbiol*. 27, 653–660. doi: 10.1078/0723202042369866

Hormisch, D., Brost, I., Kohring, G. W., Giffhorn, F., Kroppenstedt, R. M., Stackebrandt, E., et al. (2006). *Mycobacterium fluoranthenivorans* sp. nov," in: list of new names and new combinations previously effectively, but not validly, published, list no. 110. *Int. J. Syst. Evol. Microbiol*. 56, 1459–1460. doi: 10.1099/ijs.0.64507-0

Jiménez, M. S., Campos-Herrero, M. I., Garcia, D., Luquin, M., Herrera, L., and Garcia, M. J. (2004). *Mycobacterium canariasense* sp. nov. *Int. J. Syst. Evol. Microbiol*. 54, 1729–1734. doi: 10.1099/ijs.0.02999-0

Kasai, H., Ezaki, T., and Harayama, S. (2000). Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences. *J. Clin. Microbiol*. 38, 301–308.

Kazda, J. (1980). *Mycobacterium sphagni* sp. nov. *Int. J. Syst. Evol. Microbiol*. 30, 77–81. doi: 10.1099/00207713-30-1-77

Kazda, J., and Müller, K. (1979). *Mycobacterium komossense* sp. nov. *Int. J. Syst. Evol. Microbiol*. 29, 361–365. doi: 10.1099/00207713-29-4-361

Kazda, J., Cooney, R., Monaghan, M., Quinn, P. J., Stackebrandt, E., Dorsch, M., et al. (1993). *Mycobacterium hiberniae* sp. nov. *Int. J. Syst. Bacteriol*. 43, 352–357. doi: 10.1099/00207713-43-2-352

Kazda, J., Müller, H. J., Stackebrandt, E., Daffé, M., Müller, K., and Pitulle, C. (1992). *Mycobacterium madagascariense* sp. nov. *Int. J. Syst. Evol. Microbiol*. 42, 524–528. doi: 10.1099/00207713-42-4-524

Khan, A. A., Kim, S. J., Paine, D. D., and Cerniglia, C. E. (2002). Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as *Mycobacterium vanbaalenii* sp. nov. *Int. J. Syst. Evol. Microbiol*. 52, 1997–2002. doi: 10.1099/00207713-52-6-1997

Kim, B. J., Hong, S. H., Yu, H. K., Park, Y. G., Jeong, J., Lee, S. H., et al. (2013). *Mycobacterium parakoreense* sp. nov., a slowly growing non-chromogenic species related to Mycobacterium koreense, isolated from a human clinical specimen. *Int. J. Syst. Evol. Microbiol*. 63, 2301–2308. doi: 10.1099/ijs.0.045070-0

Kim, B. J., Jeong, J., Lee, S. H., Kim, S. R., Yu, H. K., Park, Y. G., et al. (2012). *Mycobacterium koreense* sp. nov., a slowly growing non-chromogenic species closely related to *Mycobacterium triviale. Int. J. Syst. Evol. Microbiol*. 62, 1289–1295. doi: 10.1099/ijs.0.033274-0

Kim, B. J., Kim, J. M., Kim, B. R., Lee, S. Y., Kim, G., Jang, Y. H., et al. (2015). *Mycobacterium anyangense* sp. nov., a rapidly growing species isolated from blood of Korean native cattle, Hanwoo (*Bos taurus coreanae*). *Int. J. Syst. Evol. Microbiol.* 65, 2277–2285. doi: 10.1099/ijs.0.000255

Kim, H., Kim, S. H., Shim, T. S., Kim, M. N., Bai, G. H., Park, Y. G., et al. (2005). Differentiation of *Mycobacterium* species by analysis of the heat-shock protein 65 gene (*hsp65*). *Int. J. Syst. Evol. Microbiol.* 55, 1649–1656. doi: 10.1099/ijs.0.63553-0

Kim, S. H., and Shin, J. H. (2017). Identification of nontuberculous mycobacteria using multilocous sequence analysis of 16S rRNA, *hsp65*, and *rpoB*. *J. Clin. Lab. Anal.* 32:e22184. doi: 10.1002/jcla.22184

Kirschner, P., Teske, A., Schroder, K. H., Kroppenstedt, R. M., Wolters, J., and Bottger, E. C. (1992). *Mycobacterium confluentis* sp. nov. *Int. J. Syst. Bacteriol.* 42, 257–262. doi: 10.1099/00207713-42-2-257

Kleespies, M., Kroppenstedt, R. M., Rainey, F. A., Webb, L. E., and Stackebrandt, E. (1996). *Mycobacterium hodleri* sp. nov., a new member of the fast-growing mycobacteria capable of degrading polycyclic aromatic hydrocarbons. *Int. J. Syst. Bacteriol.* 46, 683–687. doi: 10.1099/00207713-46-3-683

Klenk, H. P., and Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Syst. Appl. Microbiol.* 33, 175–182. doi: 10.1016/j.syapm.2010.03.003

Konjek, J., Souded, S., Guerardel, Y., Trivelli, X., Bernut, A., Kremer, L., et al. (2016). *Mycobacterium lutetiense* sp. nov., *Mycobacterium montmartrense* sp. nov. and *Mycobacterium arcueilense* sp. nov., members of a novel group of non-pigmented rapidly growing mycobacteria recovered from a water distribution system. *Int. J. Syst. Evol. Microbiol.* 66, 3694–3702. doi: 10.1099/ijsem.0.001253

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504–509. doi: 10.1016/j.mib.2007.08.006

Kubica, G. P., Silcox, V. A., Kilburn, J. O., Smithwick, R. W., Beam, R. E., Jones, W. D., et al. (1970). Differential identification of mycobacteria VI. *Mycobacterium triviale Kubica* sp. nov. *Int. J. Syst. Evol. Microbiol.* 20, 161–174. doi: 10.1099/00207713-20-2-161

Kusunoki, S., and Ezaki, T. (1992). Proposal of *Mycobacterium peregrinum* sp. nov., nom. rev., and elevation of *Mycobacterium chelonae* subsp. abscessus (Kubica et al.) to species status: *Mycobacterium abscessus* comb. nov. *Int. J. Syst. Bacteriol.* 42, 240–245. doi: 10.1099/00207713-42-2-240

Lamy, B., Marchandin, H., Hamitouche, K., and Laurent, F. (2008). *Mycobacterium setense* sp. nov., a *Mycobacterium fortuitum*-group organism isolated from a patient with soft tissue infection and osteitis. *Int. J. Syst. Evol. Microbiol.* 58, 486–490. doi: 10.1099/ijs.0.65222-0

Leao, S. C., Tortoli, E., Euzeby, J. P., and Garcia, M. J. (2011). Proposal that *Mycobacterium massiliense* and *Mycobacterium bolletii* be united and reclassified as *Mycobacterium abscessus* subsp. *bolletii* comb. nov., designation of *Mycobacterium abscessus* subsp. *abscessus* subsp. nov. and emended description of *Mycobacterium abscessus*. *Int. J. Syst. Evol. Microbiol.* 61, 2311–2313. doi: 10.1099/ijs.0.023770-0

Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi: 10.1093/molbev/msn067

Lee, H., Lee, S. A., Lee, I. K., Yu, H. K., Park, Y. G., Jeong, J., et al. (2010). *Mycobacterium paraterrae* sp. nov. recovered from a clinical specimen: novel chromogenic slow growing mycobacteria related to *Mycobacterium terrae* complex. *Microbiol. Immunol.* 54, 46–53. doi: 10.1111/j.1348-0421.2009.00184.x

Lee, H., Lee, S. A., Lee, I. K., Yu, H. K., Park, Y. G., Jeong, J., et al. (2016). *Mycobacterium parraterrae* sp. nov. in: Oren A., Garrity G.M. (editors). List of new names and new combinations previously effectively, but not validly, published, list no. 172. Int. J. Syst. Evol. Microbiol. 66, 4299-4305. doi: 10.1099/ijsem.0.001585

Lehmann, K. B., and Neumann, R. O. (1896). *Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik*. München: J. F. Lehmann. Available online at: http://www.worldcat.org/title/atlas-und-grundriss-der-bakteriologie-und-lehrbuch-der-speziellen-bakteriologischen-diagnostik/oclc/715501858

Lehmann, K. B., and Neumann, R. O. (1899). *Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik*. München: J. F. Lehmann. Available online at: http://www.worldcat.org/title/atlas-und-grundriss-der-bakteriologie-und-lehrbuch-der-speziellen-bakteriologischen-diagnostik/oclc/715501858

Lévy-Frébault, V. V., Rafidinarivo, A., Prome, J. C., Grandry, J., Boisvert, H., and David, H. L. (1983). *Mycobacterium fallax* sp. nov. *Int. J. Syst. Evol. Microbiol.* 33, 336–343. doi: 10.1099/00207713-33-2-336

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

Lory, S. (2014). "The Family Mycobacteriaceae," in *The Prokaryotes-Actinobacteria,* eds E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (New York, NY: Springer), 571–575.

Ludwig, W., and Klenk, H.-P. (2005). "Overview: A phylogenetic backbone and taxonomic framework for prokaryotic systamatics," in *Bergey's Manual of Systematic Bacteriology*, eds D. J.Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (Berlin: Springer-Verlag), 49–65.

Luquin, M., Ausina, V., Vincent-Lévy-Frébault, V., Lanéelle, M. A., Belda, F., García-Barceló, M., et al. (1993). *Mycobacterium brumae* sp. nov., a rapidly growing, nonphotochromogenic *Mycobacterium*. *Int. J. Syst. Evol. Microbiol.* 43, 405–413. doi: 10.1099/00207713-43-3-405

Magee, G. M., and Ward, A. C. (2012). "Genus I. *Mycobacterium* Lehmann and Neumann 1896, 363^{AL}," in *Bergey's Manual of Systematic Bacteriology*, *Vol. 5, Actinobacteria*, eds M. Goodfellow, P. Kampfer, H. J. Busse, M. E. Trujillo, K. Suzuki, W. Ludwig, and W. Whitman (New York, NY: Springer), 312–375.

Masaki, T., Ohkusu, K., Hata, H., Fujiwara, N., Iihara, H., Yamada-Noda, M., et al. (2006). *Mycobacterium kumamotonense* sp. nov. recovered from clinical specimen and the first isolation report of *Mycobacterium arupense* in Japan: novel slowly growing, nonchromogenic clinical isolates related to *Mycobacterium terrae* complex. *Microbiol. Immunol.* 50, 889–897. doi: 10.1111/j.1348-0421.2006.tb03865.x

Masaki, T., Ohkusu, K., Hata, H., Fujiwara, N., Iihara, H., Yamada-Noda, M., et al. (2007). *Mycobacterium kumamotonense* sp. nov. in: *list of new names and new combinations previously effectively, but not validly, published,* list no. 114. *Int. J. Syst. Evol. Microbiol.* 57, 433–434. doi: 10.1099/ijs.0.65052-0

Medjahed, H., Gaillard, J. L., and Reyrat, J. M. (2010). *Mycobacterium abscessus*: a new player in the mycobacterial field. *Trends Microbiol.* 18, 117–123. doi: 10.1016/j.tim.2009.12.007

Mignard, S., and Flandrois, J. P. (2008). A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. *Int. J. Syst. Evol. Microbiol.* 58, 1432–1441. doi: 10.1099/ijs.0.65658-0

Milne, B. W., Arnold, M. H., Hudson, B., and Coolican, M. R. (2009). Infectious arthritis of the knee caused by *Mycobacterium terrae*: a case report. *J. Orthop. Surg. (Hong. Kong.)* 17, 103–108. doi: 10.1177/230949900901700123

Moore, M., and Frerichs, J. B. (1953). An unusual acid-fast infection of the knee with subcutaneous, abscess-like lesions of the gluteal region; report of a case with a study of the organism, *Mycobacterium abscessus*, n. sp. *J. Invest. Dermatol.* 20, 133–169. doi: 10.1038/jid.1953.18

Mukhopadhyay, S., and Balaji, K. N. (2011). The PE and PPE proteins of *Mycobacterium tuberculosis*. *Tuberculosis. (Edinb.)* 91, 441–447. doi: 10.1016/j.tube.2011.04.004

Mun, H. S., Park, J. H., Kim, H., Yu, H. K., Park, Y. G., Cha, C. Y., et al. (2008). *Mycobacterium senuense* sp. nov., a slowly growing, non-chromogenic species closely related to the *Mycobacterium terrae* complex. *Int. J. Syst. Evol. Microbiol.* 58, 641–646. doi: 10.1099/ijs.0.65374-0

Naushad, H. S., Lee, B., and Gupta, R. S. (2014). Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya, Pectobacterium* and *Brenneria*. *Int. J. Syst. Evol. Microbiol.* 64, 366–383. doi: 10.1099/ijs.0.054213-0

Nessar, R., Cambau, E., Reyrat, J. M., Murray, A., and Gicquel, B. (2012). *Mycobacterium abscessus*: a new antibiotic nightmare. *J. Antimicrob. Chemother.* 67, 810–818. doi: 10.1093/jac/dkr578

Ngeow, Y. F., Wong, Y. L., Tan, J. L., Hong, K. W., Ng, H. F., Ong, B. L., et al. (2015). Identification of new genomospecies in the *Mycobacterium terrae* complex. *PLoS ONE* 10:e0120789. doi: 10.1371/journal.pone.0120789

Nogueira, C. L., Simmon, K. E., Chimara, E., Cnockaert, M., Palomino, J. C., et al. (2015a). *Mycobacterium franklinii* sp. nov., a species closely related to the members of the *Mycobacterium chelonae-Mycobacterium abscessus* group. *Int. J. Syst. Evol. Microbiol.* 65, 2148–2153. doi: 10.1099/ijs.0.000234

Nogueira, C. L., Whipps, C. M., Matsumoto, C. K., Chimara, E., Droz, S., Tortoli, E., et al. (2015b). *Mycobacterium saopaulense* sp. nov., a rapidly growing *Mycobacterium* closely related to members of the *Mycobacterium chelonae–Mycobacterium abscessus* group. *Int. J. Syst. Evol. Microbiol.* 65, 4403–4409. doi: 10.1099/ijsem.0.000590

Padgitt, P. J., and Moshier, S. E. (1987). *Mycobacterium poriferae* sp. nov., a scotochromogenic, rapidly growing species isolated from a marine sponge. *Int. J. Syst. Evol. Microbiol.* 37, 186–191. doi: 10.1099/00207713-37-3-186

Parte, A. C. (2014). LPSN-list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.* 42, D613–D616. doi: 10.1093/nar/gkt1111

Pitulle, C., Dorsch, M., Kazda, J., Wolters, J., and Stackebrandt, E. (1992). Phylogeny of rapidly growing members of the genus *Mycobacterium*. *Int. J. Syst. Bacteriol.* 42, 337–343. doi: 10.1099/00207713-42-3-337

Prasanna, A. N., and Mehra, S. (2013). Comparative phylogenomics of pathogenic and non-pathogenic *Mycobacterium*. *PLoS ONE* 8:e71248. doi: 10.1371/journal.pone.0071248

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

Qin, Q. L., Xie, B. B., Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196, 2210–2215. doi: 10.1128/JB.01688-14

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

Ramaprasad, E. V. V., Rizvi, A., Banerjee, S., Sasikala, C., and Ramana, C. (2016). *Mycobacterium oryzae* sp. nov., a scotochromogenic, rapidly growing species is able to infect human macrophage cell line. *Int. J. Syst. Evol. Microbiol.* 66, 4530–4536. doi: 10.1099/ijsem.0.001386

Reischl, U., Melzl, H., Kroppenstedt, R. M., Miethke, T., Naumann, L., Mariottini, A., et al. (2006). *Mycobacterium monacense* sp. nov. *Int. J. Syst. Evol. Microbiol.* 56, 2575–2578. doi: 10.1099/ijs.0.64527-0

Richter, E., Niemann, S., Gloeckner, F. O., Pfyffer, G. E., and Rusch-Gerdes, S. (2002). *Mycobacterium holsaticum* sp. nov. *Int. J. Syst. Evol. Microbiol.* 52, 1991–1996. doi: 10.1099/00207713-52-6-1991

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Ripoll, F., Pasek, S., Schenowitz, C., Dossat, C., Barbe, V., Rottman, M., et al. (2009). Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. *PLoS ONE* 4:e5660. doi: 10.1371/journal.pone.0005660

Rogall, T., Wolters, J., Flohr, T., and Bottger, E. C. (1990). Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *Int. J. Syst. Bacteriol.* 40, 323–330. doi: 10.1099/00207713-40-4-323

Rokas, A., and Holland, P. W. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15, 454–459. doi: 10.1016/S0169-5347(00)01967-4

Roth, A., Fischer, M., Hamid, M. E., Michalke, S., Ludwig, W., and Mauch, H. (1998). Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. *J. Clin. Microbiol.* 36, 139–147.

Runyon, E. H. (1965). Typical mycobacteria: their classification. *Am. Rev. Respir. Dis.* 91, 288–289.

Sahraoui, N., Ballif, M., Zelleg, S., Yousfi, N., Ritter, C., Friedel, U., et al. (2011). *Mycobacterium algericum* sp. nov., a novel rapidly growing species related to the *Mycobacterium terrae* complex and associated with goat lung lesions. *Int. J. Syst. Evol. Microbiol.* 61, 1870–1874. doi: 10.1099/ijs.0.024851-0

Sawana, A., Adeolu, M., and Gupta, R. S. (2014). Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front. Genet.* 5:429. doi: 10.3389/fgene.2014.00429

Schinsky, M. F., McNeil, M. M., Whitney, A. M., Steigerwalt, A. G., Lasker, B. A., Floyd, M. M., et al. (2000). *Mycobacterium septicum* sp. nov., a new rapidly growing species associated with catheter-related bacteraemia. *Int. J. Syst. Evol. Microbiol.* 50, 575–581. doi: 10.1099/00207713-50-2-575

Schinsky, M. F., Morey, R. E., Steigerwalt, A. G., Douglas, M. P., Wilson, R. W., Floyd, M. M., et al. (2004). Taxonomic variation in the *Mycobacterium fortuitum* third biovariant complex: description of *Mycobacterium boenickei* sp. nov., *Mycobacterium houstonense* sp. nov., *Mycobacterium neworleansense* sp. nov. and *Mycobacterium brisbanense* sp. nov. and recognition of *Mycobacterium porcinum* from human clinical isolates. *Int. J. Syst. Evol. Microbiol.* 54, 1653–1667. doi: 10.1099/ijs.0.02743-0

Schoeffler, A. J., May, A. P., and Berger, J. M. (2010). A domain insertion in *Escherichia coli* GyrB adopts a novel fold that plays a critical role in gyrase function. *Nucleic Acids Res.* 38, 7830–7844. doi: 10.1093/nar/gkq665

Schröder, K. H., Naumann, L., Kroppenstedt, R. M., and Reischl, U. (1997). *Mycobacterium hassiacum* sp. nov., a new rapidly growing thermophilic *Mycobacterium*. *Int. J. Syst. Bacteriol.* 47, 86–91. doi: 10.1099/00207713-47-1-86

Segata, N., Bornigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4:2304. doi: 10.1038/ncomms3304

Shahraki, A. H., Cavusoglu, C., Borroni, E., Heidarieh, P., Koksalan, O. K., Cabibbe, A. M., et al. (2015). *Mycobacterium celeriflavum* sp. nov., a rapidly growing scotochromogenic bacterium isolated from clinical specimens. *Int. J. Syst. Evol. Microbiol.* 65, 510–515. doi: 10.1099/ijs.0.064832-0

Shojaei, H., Daley, C., Gitti, Z., Hashemi, A., Heidarieh, P., Moore, E. R., et al. (2013). *Mycobacterium iranicum* sp. nov., a rapidly growing scotochromogenic species isolated from clinical specimens on three different continents. *Int. J. Syst. Evol. Microbiol.* 63, 1383–1389. doi: 10.1099/ijs.0.043562-0

Shojaei, H., Goodfellow, M., Magee, J. G., Freeman, R., Gould, F. K., and Brignall, C. G. (1997). *Mycobacterium novocastrense* sp. nov., a rapidly growing photochromogenic *Mycobacterium*. *Int. J. Syst. Bacteriol.* 47, 1205–1207. doi: 10.1099/00207713-47-4-1205

Shojaei, H., Magee, J. G., Freeman, R., Yates, M., Horadagoda, N. U., and Goodfellow, M. (2000). *Mycobacterium elephantis* sp. nov., a rapidly growing non-chromogenic Mycobacterium isolated from an elephant. *Int. J. Syst. Evol. Microbiol.* 50, 1817–1820. doi: 10.1099/00207713-50-5-1817

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

Simmon, K. E., Brown-Elliott, B. A., Ridge, P. G., Durtschi, J. D., Mann, L. B., Slechta, E. S., et al. (2011). *Mycobacterium chelonae-abscessus* complex associated with sinopulmonary disease, Northeastern USA. *Emerg. Infect. Dis.* 17, 1692–1700. doi: 10.3201/eid1709.101667

Singh, B., and Gupta, R. S. (2009). Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genomics* 281, 361–373. doi: 10.1007/s00438-008-0417-3

Skerman, V. B. D., McGowan, V., and Sneath, P. H. A. (1980). Approved lists of bacteria names. *Int. J. Syst. Evol. Microbiol.* 30, 225–420. doi: 10.1099/00207713-30-1-225

Smith, D. S., Lindholm-Levy, P., Huitt, G. A., Heifets, L. B., and Cook, J. L. (2000). *Mycobacterium terrae*: case reports, literature review, and *in vitro* antibiotic susceptibility testing. *Clin. Infect. Dis.* 30, 444–453. doi: 10.1086/313693

Springer, B., Bottger, E. C., Kirschner, P., and Wallace, R. J. Jr. (1995). Phylogeny of the *Mycobacterium chelonae*-like organism based on partial sequencing of the 16S rRNA gene and proposal of *Mycobacterium mucogenicum* sp. *nov. Int. J. Syst. Bacteriol.* 45, 262–267. doi: 10.1099/00207713-45-2-262

Stackebrandt, E. (1992). "Unifying phylogeny and phenotypic diversity," in *The Prokaryotes*, eds A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K. H.Schleifer (New York, NY: Springer-Verlag), 19–47.

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

Stahl, D. A., and Urbance, J. W. (1990). The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *J. Bacteriol.* 172, 116–124. doi: 10.1128/jb.172.1.116-124.1990

Stanford, J. L., and Gunthorpe, W. J. (1971). A study of some fast-growing scotochromogenic mycobacteria including species descriptions of *Mycobacterium gilvum* (new species) and *Mycobacterium duvalii* (new species). *Br. J. Exp. Pathol.* 52, 627–637.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tasler, G. R. W., and Hartley, W. J. (1981). Foal abortion associated with *Mycobacterium terrae* infection. *Vet. Pathol.* 18, 122–125. doi: 10.1177/030098588101800115

Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., and Thompson, F. L. (2013). Microbial genomic taxonomy. *BMC genomics* 14:913. doi: 10.1186/1471-2164-14-913

Tortoli, E. (2012). Phylogeny of the genus *Mycobacterium*: many doubts, few certainties. *Infect. Genet. Evol.* 12, 827–831. doi: 10.1016/j.meegid.2011.05.025

Tortoli, E. (2014). Microbiological features and clinical relevance of new species of the genus *Mycobacterium*. *Clin. Microbiol. Rev.* 27, 727–752. doi: 10.1128/CMR.00035-14

Tortoli, E., Baruzzo, S., Heijdra, Y., Klenk, H. P., Lauria, S., Mariottini, A., et al. (2009). *Mycobacterium insubricum* sp. nov. *Int. J. Syst. Evol. Microbiol.* 59, 1518–1523. doi: 10.1099/ijs.0.003459-0

Tortoli, E., Fedrizzi, T., Pecorari, M., Giacobazzi, E., De Sanctis, V., Bertorelli, R., et al. (2015). The new phylogenesis of the genus *Mycobacterium*. *Int. J. Mycobacteriol.* 4, 77. doi: 10.1016/j.ijmyco.2014.10.017

Tortoli, E., Gitti, Z., Klenk, H. P., Lauria, S., Mannino, R., Mantegani, P., et al. (2013). Survey of 150 strains belonging to the *Mycobacterium terrae* complex and description of *Mycobacterium engbaekii* sp. *nov., Mycobacterium heraklionense* sp. nov. and *Mycobacterium longobardum* sp. nov. *Int. J. Syst. Evol. Microbiol.* 63, 401–411. doi: 10.1099/ijs.0.038737-0

Tortoli, E., Kohl, T. A., Brown-Elliott, B. A., Trovato, A., Leao, S. C., Garcia, M. J., et al. (2016). Emended description of *Mycobacterium abscessus*, *Mycobacterium abscessus* subsp. *abscessus* and *Mycobacterium abscessus* subsp. *bolletii* and designation of *Mycobacterium abscessus* subsp. *massiliense* comb. nov. *Int. J. Syst. Evol. Microbiol.* 66, 4471–4479. doi: 10.1099/ijsem.0.001376

Tortoli, E., Kroppenstedt, R. M., Bartoloni, A., Caroli, G., Jan, I., Pawlowski, J., et al. (1999). *Mycobacterium tusciae* sp. nov. *Int. J. Syst. Bacteriol.* 49, 1839–1844. doi: 10.1099/00207713-49-4-1839

Tortoli, E., Piersimoni, C., Kroppenstedt, R. M., Montoya-Burgos, J. I., Reischl, U., Giacometti, A., et al. (2001). *Mycobacterium doricum* sp. nov. *Int. J. Syst. Evol. Microbiol.* 51, 2007–2012. doi: 10.1099/00207713-51-6-2007

Tran, P. M., and Dahl, J. L. (2016). *Mycobacterium sarraceniae* sp. nov. and *Mycobacterium helvum sp.* nov., isolated from the pitcher plant *Sarracenia purpurea. Int. J. Syst. Evol. Microbiol.* 66, 4480–4485. doi: 10.1099/ijsem.0.001377

Trevisan, V. (1889). *I Generi e le Specie delle Batteriacee.* Milano: Zanaboni, L. and Gabuzzi, E.

Trujillo, M. E., Velazquez, E., Kroppenstedt, R. M., Schumann, P., Rivas, R., Mateos, P. F., et al. (2004). *Mycobacterium psychrotolerans* sp. nov., isolated from pond water near a uranium mine. *Int. J. Syst. Evol. Microbiol.* 54, 1459–1463. doi: 10.1099/ijs.0.02938-0

Tsukamura, M. (1965a). A group of mycobacteria from soil resources resembling nonphotochromogens (Group 3). *Med. Biol.* 71, 110–113.

Tsukamura, M. (1965b). *Mycobacterium parafortuitum*: a new species. *J. Gen. Microbiol.* 42, 7–12. doi: 10.1099/00221287-42-1-7

Tsukamura, M. (1966). Adansonian classification of mycobacteria. *J. Gen. Microbiol.* 45, 253–273. doi: 10.1099/00221287-45-2-253

Tsukamura, M. (1967a). Identification of mycobacteria. *Tubercle* 48, 311–338. doi: 10.1016/S0041-3879(67)80040-0

Tsukamura, M. (1967b). *Mycobacterium chitae*: a new species. *Microbiol. Immunol.* 11, 43–47. doi: 10.1111/j.1348-0421.1967.tb00319.x

Tsukamura, M. (1972). *Mycobacterium neoaurum* Tsukamura sp. nov. *Med. Biol.* 85, 229–233.

Tsukamura, M. (1981). Numerical analysis of rapidly growing, nonphotochromogenic mycobacteria, including *Mycobacterium agri* (Tsukamura 1972) Tsukamura sp. nov., nom. rev. *Int. J. Syst. Evol. Microbiol.* 31, 247–258. doi: 10.1099/00207713-31-3-247

Tsukamura, M., Mizuno, S., and Toyama, H. (1983a). *Mycobacterium pulveris* sp. nov., a nonphotochromogenic *Mycobacterium* with an intermediate growth rate. *Int. J. Syst. Evol. Microbiol.* 33, 811–815. doi: 10.1099/00207713-33-4-811

Tsukamura, M., Mizuno, S., and Tsukamura, S. (1981). Numerical analysis of rapidly growing, scotochromogenic mycobacteria, including *Mycobacterium obuense* sp. *nov.,* nom. rev., *Mycobacterium rhodesiae* sp. nov., nom. rev., *Mycobacterium aichiense* sp. nov., nom. rev., *Mycobacterium chubuense* sp. nov., nom. rev., and *Mycobacterium tokaiense* sp. nov., nom. rev. *Int. J. Syst. Evol. Microbiol.* 31, 263–275. doi: 10.1099/00207713-31-3-263

Tsukamura, M., Nemoto, H., and Yugi, H. (1983b). *Mycobacterium porcinum* sp. nov., a porcine pathogen. *Int. J. Syst. Evol. Microbiol.* 33, 162–165. doi: 10.1099/00207713-33-2-162

Tsukamura, M., Van Der Meulen, H. J., and Grabow, W. O. K. (1983c). Numerical taxonomy of rapidly growing, scotochromogenic mycobacteria of the *Mycobacterium parafortuitum* complex: *Mycobacterium austroafricanum* sp. nov. and *Mycobacterium diernhoferi* sp. nov., nom. rev. *Int. J. Syst. Bacteriol.* 33, 460–469. doi: 10.1099/00207713-33-3-460

Tsukamura, M., Yano, I., and Imaeda, T. (1986a). "*Mycobacterium fortuitum* subsp. *acetamidolyticum* sp. Nov," in Validation of the Publication of New Names and New Combinations Previously Effectively Published Outside the IJSB, List no. 21. Int. J. Syst. Evol. *Microbiol.* 36, 489.

Tsukamura, M., Yano, I., and Imaeda, T. (1986b). *Mycobacterium fortuitum* subspecies *acetamidolyticum*, a new subspecies of *Mycobacterium fortuitum*. *Microbiol. Immunol.* 30, 97–110. doi: 10.1111/j.1348-0421.1986.tb00925.x

Tsukamura, M., Yano, I., and Imaeda, T. (1986c). *Mycobacterium moriokaense* sp. nov., a rapidly growing, nonphotochromogenic *Mycobacterium. Int. J. Syst. Evol. Microbiol.* 36, 333–338. doi: 10.1099/00207713-36-2-333

van Ingen, J., Boeree, M. J., Kosters, K., Wieland, A., Tortoli, E., Dekhuijzen, P. N., et al. (2009). Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. *Int. J. Syst. Evol. Microbiol.* 59, 2277–2282. doi: 10.1099/ijs.0.008854-0

Vasireddy, R., Vasireddy, S., Brown-Elliott, B. A., Wengenack, N. L., Eke, U. A., Benwill, J. L., et al. (2016). *Mycobacterium arupense, Mycobacterium heraklionense*, and a newly proposed species, "*Mycobacterium virginiense*" sp. nov., but not *Mycobacterium* nonchromogenicum, as species of the *Mycobacterium terrae* complex causing tenosynovitis and osteomyelitis. *J. Clin. Microbiol.* 54, 1340–1351. doi: 10.1128/JCM.00198-16

Vasireddy, R., Vasireddy, S., Brown-Elliott, B. A., Wengenack, N. L., Eke, U. A., Benwill, J. L., et al. (2017). *Mycobacterium virginiense* sp. nov. In: Oren A., Garrity G.M. (editors). list of new names and new combinations previously effectively, but not validly, published, list no. 175. *Int. J. Syst. Evol. Microbiol.* 67, 1–3. doi: 10.1099/ijsem.0.001733

Vuorio, R., Andersson, M. A., Rainey, F. A., Kroppenstedt, R. M., Kampfer, P., Busse, H. J., et al. (1999). A new rapidly growing mycobacterial species, *Mycobacterium murale* sp. nov., isolated from the indoor walls of a children's day care centre. *Int. J. Syst. Bacteriol.* 49, 25–35. doi: 10.1099/00207713-49-1-25

Wang, J., McIntosh, F., Radomski, N., Dewar, K., Simeone, R., Enninga, J., et al. (2015). Insights on the emergence of *Mycobacterium tuberculosis* from the analysis of *Mycobacterium kansasii*. *Genome Biol. Evol.* 7, 856–870. doi: 10.1093/gbe/evv035

Wang, Y. N., Chi, C. Q., Cai, M., Lou, Z. Y., Tang, Y. Q., Zhi, X. Y., et al. (2010). *Amycolicicoccus subflavus* gen. nov., sp. nov., an actinomycete isolated from a saline soil contaminated by crude oil. *Int. J. Syst. Evol. Microbiol.* 60, 638–643. doi: 10.1099/ijs.0.010546-0

Wang, Z., and Wu, M. (2013). A phylum-level bacterial phylogenetic marker database. *Mol. Biol. Evol.* 30, 1258–1262. doi: 10.1093/molbev/mst059

Wayne, L. G. (1966). Classification and identification of mycobacteria. 3. Species within group 3. *Am. Rev. Respir. Dis.* 93, 919–928.

Wayne, L. G., and Kubica, G. P. (1986). "Genus *Mycobacterium*," in *Bergey's Manual of Systematic Bacteriology,* eds P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt (Baltimore, MD: Williams and Wilkins), 1436–1457.

Wee, W. Y., Dutta, A., and Choo, S. W. (2017). Comparative genome analyses of mycobacteria give better insights into their evolution. *PLoS ONE* 12:e0172831. doi: 10.1371/journal.pone.0172831

Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using

a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851

Whipps, C. M., Butler, W. R., Pourahmad, F., Watral, V. G., and Kent, M. L. (2007). Molecular systematics support the revival of *Mycobacterium salmoniphilum* (ex Ross 1960) sp. nov., nom. rev., a species closely related to *Mycobacterium chelonae*. *Int. J. Syst. Evol. Microbiol.* 57, 2525–2531. doi: 10.1099/ijs.0.64841-0

Willumsen, P., Karlson, U., Stackebrandt, E., and Kroppenstedt, R. M. (2001). *Mycobacterium frederiksbergense* sp. nov., a novel polycyclic aromatic hydrocarbon-degrading *Mycobacterium species*. *Int. J. Syst. Evol. Microbiol.* 51, 1715–1722. doi: 10.1099/00207713-51-5-1715

Wilson, R. W., Steingrube, V. A., Bottger, E. C., Springer, B., Brown-Elliott, B. A., Vincent, V., et al. (2001). *Mycobacterium immunogenum* sp. nov., a novel species related to *Mycobacterium abscessus* and associated with clinical disease, pseudo-outbreaks and contaminated metalworking fluids: an international cooperative study on mycobacterial taxonomy. *Int. J. Syst. Evol. Microbiol.* 51, 1751–1764. doi: 10.1099/00207713-51-5-1751

Wong, S. Y., Paschos, A., Gupta, R. S., and Schellhorn, H. E. (2014). Insertion/deletion-based approach for the detection of *Escherichia coli* O157:H7 in freshwater environments. *Environ. Sci. Technol.* 48, 11462–11470. doi: 10.1021/es502794h

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462, 1056–1060. doi: 10.1038/nature08656

Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K. H., et al. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol.* 31, 241–250. doi: 10.1016/j.syapm.2008.07.001

Yarza, P., Yilmaz, P., Pruesse, E., Gloeckner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences 1. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330

Zhang, D. F., Chen, X., Zhang, X. M., Zhi, X. Y., Yao, J. C., Jiang, Y., et al. (2013a). *Mycobacterium sediminis* sp. nov. and *Mycobacterium arabiense* sp. nov., two rapidly growing members of the genus *Mycobacterium*. *Int. J. Syst. Evol. Microbiol.* 63, 4081–4086. doi: 10.1099/ijs.0.050567-0

Zhang, Y., Zhang, J., Fang, C., Pang, H., and Fan, J. (2012). *Mycobacterium litorale* sp. nov., a rapidly growing *Mycobacterium* from soil. *Int. J. Syst. Evol. Microbiol.* 62, 1204–1207. doi: 10.1099/ijs.0.033449-0

Zhang, Z. Y., Sun, Z. Q., Wang, Z. L., Hu, H. R., Wen, Z. L., Song, Y. Z., et al. (2013b). Identification and pathogenicity analysis of a novel non-tuberculous *Mycobacterium* clinical isolate with nine-antibiotic resistance. *Clin. Microbiol. Infect.* 19, 91–96. doi: 10.1111/j.1469-0691.2012.03818.x

Zhang, Z. Y., Sun, Z. Q., Wang, Z. L., Wen, Z. L., Sun, Q. W., Zhu, Z. Q., et al. (2011). Complete genome sequence of a novel clinical isolate, the nontuberculous *Mycobacterium* strain JDM601. *J. Bacteriol.* 193, 4300–4301. doi: 10.1128/JB.05291-11

Zopf, W. (1883). *Die Spaltpilze: Nachdem neuesten Standpunkte*. Breslau: Eduard Trewendt. Available online at: http://www.worldcat.org/title/spaltpilze-nachdem-neuesten-standpunkte/oclc/7445289&referer=brief_results

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 Gupta, Lo and Son. 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

CORRECTION
published: 09 April 2019
doi: 10.3389/fmicb.2019.00714

Corrigendum: Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera

Radhey S. Gupta, Brian Lo and Jeen Son*

Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, CA, Canada

Keywords: *Mycobacterium* classification, slow-growing and fast-growing mycobacteria, conserved signature indels and signature proteins, phylogenomic analysis, fortuitum-vaccae clade, abscessus-chelonae clade, terrae clade, triviale clade

OPEN ACCESS

Edited and reviewed by:
Ludmila Chistoserdova, University of Washington, United States

****Correspondence:***
Radhey S. Gupta
gupta@mcmaster.ca

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

Received: *23 November 2018*
Accepted: *21 March 2019*
Published: *09 April 2019*

Citation:
Gupta RS, Lo B and Son J (2019) Corrigendum: Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus Mycobacterium into an Emended Genus Mycobacterium and Four Novel Genera. Front. Microbiol. 10:714. doi: 10.3389/fmicb.2019.00714

A Corrigendum on

Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera
by Gupta, R. S., Lo, B., and Son, J. (2018). Front. Microbiol. 9:67. doi: 10.3389/fmicb.2018.00067

In the original article, there was an error. Based on the branching position of *Mycobacterium vulneris* (van Ingen et al., 2009) in different phylogenomic trees and on multiple identified molecular signatures that this species shared with a clade of rapid growing mycobacteria, we proposed a reclassification of *M. vulneris*, into a new genus, *Mycolicibacterium*, corresponding to a clade of rapid-growing mycobacteria. However, it was noted in our article that the branching of *M. vulneris*, which is a slow-growing species with rapid-growing mycobacteria, was anomalous.

In a Frontiers commentary, Tortoli (2018) indicated that the genome sequence of *M. vulneris*, originally available in the NCBI genome database (accession CCBG00000000; Croce et al., 2014), was mislabeled and very likely corresponded to *Mycobacterium porcinum* (a rapid grower). Tortoli (2018) also reported the sequencing of the type strain of *M. vulneris*, DSM 45247^T and this genome sequence (accession NCXM01000000) showed the branching of *M. vulneris* within the slow-growing group of mycobacteria, belonging to the genus *Mycobacterium*.

Our own analysis with this new genome sequence also confirms the branching of *M. vulneris* within the delimited genus *Mycobacterium*, encompassing different slow-growing mycobacteria. As a result, the transfer of *M. vulneris* into the genus *Mycolicibacterium* as proposed in Table 11 of our article was incorrect as a direct result of the mislabeling of the available genome sequence for this species. To correct this error, we propose that the species *Mycolicibacterium vulneris* (Gupta et al., 2018) should be reinstated to its previous basonym *Mycobacterium vulneris* (van Ingen et al., 2009) and as part of the genus *Mycobacterium* (Gupta et al., 2018).

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way.

REFERENCES

Croce, O., Robert, C., Raoult, D., and Drancourt, M. (2014). Draft genome sequence of *Mycobacterium vulneris* DSM 45247T. *Genome Announc.* 2:e00370-14. doi: 10.1128/genomeA.00370-14

Gupta, R. S., Lo, B., and Son, J. (2018). Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. *Front. Microbiol.* 9:67. doi: 10.3389/fmicb.2018.00067

Tortoli, E. (2018). Commentary: phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. *Front. Microbiol.* 9:2065. doi: 10.3389/fmicb.2018.02065

van Ingen, J., Boeree, M. J., Kosters, K., Wieland, A., Tortoli, E., Dekhuijzen, P. N., et al. (2009). Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. *Int. J. Syst. Evol. Microbiol.* 59, 2277–2282. doi: 10.1099/ijs.0.008854-0

Copyright © 2019 Gupta, Lo and Son. 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: 04 September 2018
doi: 10.3389/fmicb.2018.02077

Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos

Alba Pérez-Cataluña[1], Nuria Salas-Massó[1], Ana L. Diéguez[2], Sabela Balboa[2], Alberto Lema[2], Jesús L. Romalde[2] and Maria J. Figueras[1]**

[1] *Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina, Institut d'Investigació Sanitària Pere Virgili, Universitat Rovira i Virgili, Reus, Spain,* [2] *Departamento de Microbiología y Parasitología, CIBUS-Facultad de Biología, Universidade de Santiago de Compostela, Santiago de Compostela, Spain*

OPEN ACCESS

Edited by:
Martha E. Trujillo, Universidad de Salamanca, Spain

Reviewed by:
John Phillip Bowman, University of Tasmania, Australia
Javier Pascual, Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Germany

***Correspondence:**
Jesus L. Romalde
jesus.romalde@usc.es
María J. Figueras
mariajose.figueras@urv.cat

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

Received: *15 February 2018*
Accepted: *14 August 2018*
Published: *04 September 2018*

Citation:
Pérez-Cataluña A, Salas-Massó N, Diéguez AL, Balboa S, Lema A, Romalde JL and Figueras MJ (2018) Revisiting the Taxonomy of the Genus Arcobacter: Getting Order From the Chaos. Front. Microbiol. 9:2077. doi: 10.3389/fmicb.2018.02077

Since the description of the genus *Arcobacter* in 1991, a total of 27 species have been described, although some species have shown 16S rRNA similarities below 95%, which is the cut-off that usually separates species that belong to different genera. The objective of the present study was to reassess the taxonomy of the genus *Arcobacter* using information derived from the core genome (286 genes), a Multilocus Sequence Analysis (MLSA) with 13 housekeeping genes, as well as different genomic indexes like Average Nucleotide Identity (ANI), *in silico* DNA–DNA hybridization (*is*DDH), Average Amino-acid Identity (AAI), Percentage of Conserved Proteins (POCPs), and Relative Synonymous Codon Usage (RSCU). The study included a total of 39 strains that represent all the 27 species included in the genus *Arcobacter* together with 13 strains that are potentially new species, and the analysis of 57 genomes. The different phylogenetic analyses showed that the *Arcobacter* species grouped into four clusters. In addition, *A. lekithochrous* and the candidatus species '*A. aquaticus*' appeared, as did *A. nitrofigilis,* the type species of the genus, in separate branches. Furthermore, the genomic indices ANI and *is*DDH not only confirmed that all the species were well-defined, but also the coherence of the clusters. The AAI and POCP values showed intra-cluster ranges above the respective cut-off values of 60% and 50% described for species belonging to the same genus. Phenotypic analysis showed that certain test combinations could allow the differentiation of the four clusters and the three orphan species established by the phylogenetic and genomic analyses. The origin of the strains showed that each of the clusters embraced species recovered from a common or related environment. The results obtained enable the division of the current genus *Arcobacter* in at least seven different genera, for which the names *Arcobacter*, *Aliiarcobacter* gen. nov., *Pseudoarcobacter* gen. nov., *Haloarcobacter* gen. nov., *Malacobacter* gen. nov., *Poseidonibacter* gen. nov., and Candidate '*Arcomarinus*' gen. nov. are proposed.

Keywords: *Arcobacter*, *Aliiarcobacter* gen. nov., *Pseudoarcobacter* gen. nov., *Haloarcobacter* gen. nov., *Malacobacter* gen. nov., *Poseidonibacter* gen. nov., taxonomic criteria

INTRODUCTION

The genus *Arcobacter* was created by Vandamme et al. (1991) to accommodate Gram-negative, curved-shaped bacteria belonging to two species *Campylobacter cryaerophila* (now *Arcobacter cryaerophilus*) and *Campylobacter nitrofigilis* (now *A. nitrofigilis*), considered atypical campylobacters due to their ability to grow at lower temperatures (15°C–30°C) and without microaerophilic conditions (Vandamme et al., 1991). The latter species was selected as the type species for the new genus (Vandamme et al., 1991). One year later the genus was enlarged with the addition of two new species, *A. skirrowii* with an animal origin being isolated from aborted ovine, porcine and bovine fetuses, and from lambs with diarrhea, and *A. butzleri,* which was recovered from cases of human and animal diarrhea (Vandamme et al., 1992). Another two new species were incorporated into the genus in 2005. *A. halophilus* was isolated from water from a hypersaline lagoon in Hawaii (Donachie et al., 2005), and *A. cibarius* was isolated from broiled carcasses in Belgium (Houf et al., 2005). These species were assigned to the genus *Arcobacter* on the basis of the 16S rRNA gene similarity (94% and 95% for *A. nitrofigilis* with *A. halophilus* and *A. cibarius,* respectively). However, these values are equal, or even below, the cut-off of 95% for genus definition (Rosselló-Mora and Amann, 2001; Yarza et al., 2008, 2014; Tindall et al., 2010).

From 2009 onward, new species were being described year-by-year, reaching a total number of 27 in 2017. In some of these descriptions, the similarity of the 16S rRNA gene was the decisive character for taxonomic assignation at genus level, although phylogeny based on housekeeping genes (*rpo*B first and then *gyr*B and *hsp*60) was also included as additional, more discriminatory tools for the species (Collado et al., 2009a, 2011; De Smet et al., 2011). Using this approach, *A. molluscorum*, *A. ellisii*, *A. defluvii*, or *A. bivalviorum* were defined, among others (Collado et al., 2009a, 2011; Figueras et al., 2011a,b; Levican et al., 2012), which showed 16S rRNA similarities ranging from 91.1 to 94.7%, not supporting their common affiliation. On the other hand, the most closely related species, which showed a similarity of 99.1% were *A. ellisii* and *A. defluvii* (Collado et al., 2011), giving evidence for the first time of the poor resolution of the 16S rRNA gene for separating closely related species in the genus *Arcobacter*. However, the phylogenetic analysis based on the concatenated sequences of *gyr*B, *rpo*B, and *cpn*60 genes, together with the DNA–DNA hybridization results, clearly supported the existence of these two differentiated taxa (Figueras et al., 2011a). Also in 2011, *A. trophiarum* was discovered from the intestinal tract of healthy fattening pigs, which interestingly showed the closest similarities (≥97.4%) with the other species also recovered from humans or animals, i.e., *A. cryaerophilus*, *A. thereius*, *A. cibarius,* or *A. skirrowii* (De Smet et al., 2011; Figueras et al., 2014; Van den Abeele et al., 2014).

In 2013, the species *A. cloacae* and *A. suis* were described, using a Multilocus Sequence Analysis (MLSA) approach including five housekeeping genes (Levican et al., 2013) for the first time. Simultaneously, and due to the highest 16S rRNA gene similarity with *A. marinus* (95.5%), the species *A. anaerophilus* was incorporated to the genus (Sasi-Jyothsna et al., 2013). However, this species showed atypical characteristics, including lack of motility and obligate anaerobic metabolism, which led to the original description of the genus *Arcobacter* being emended (Sasi-Jyothsna et al., 2013). The most recently described species from shellfish are *A. lekithochrous*, *A. haliotis*, and *A. canalis* (Diéguez et al., 2017; Tanaka et al., 2017; Pérez-Cataluña et al., 2018a). The first one included several isolates recovered from scallop larvae and from tank seawater of a Norwegian hatchery (Diéguez et al., 2017), the second species came from an abalone of Japan (Tanaka et al., 2017) and the third from oysters submerged in a water channel contaminated with wastewater (Pérez-Cataluña et al., 2018a). However, Diéguez et al. (2018) evidenced that the species *A. haliotis* is a later heterotypic synonym of *A. lekithochrous*. Additionally, the low 16S rRNA gene similarity of *A. lekithochrous* with the known *Arcobacter* species (91.0–94.8%) found in the *A. lekithochrous* description made Diéguez et al. (2017) suggest that certain species might belong to other genera and recommend that a profound revision of the genus might clarify the taxonomy.

On the other hand, adding 2.5% NaCl to the enrichment medium and subculturing on marine agar, Salas-Massó et al. (2016) recognized seven potential new species from water and shellfish (mussels and/or oysters), and recovered new isolates of *A. halophilus* and *A. marinus* of which only the type strains had been known. In addition, during the characterization of the most recently described species *A. canalis* (Pérez-Cataluña et al., 2018a) and when trying to define the seven mentioned new species, we observed that the *Arcobacter* species formed several different clusters distant enough to suspect they might correspond to different genera, in agreement with Diéguez et al. (2017).

There are clear criteria for describing new bacterial species (Tindall et al., 2010; Figueras et al., 2011a,b). However, the description of a genus is usually based on a cut-off of <95% similarity in the 16S rRNA gene sequence, and a G+C (% mol) content differing by more than 10% (Rosselló-Mora and Amann, 2001; Yarza et al., 2008; Tindall et al., 2010; Yarza et al., 2014). Nowadays, genomic data like the Average Nucleotide Identity (ANI) and the *in silico* DNA–DNA hybridization (*is*DDH) are used to define bacterial species, although have not yet been fully explored for delineating genera (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Qin et al., 2014; Chun et al., 2018).

A percentage of Average Amino-acid Identity (AAI) ranging from 60 to 80% between the compared genomes of species or strains and a Percentage of Conserved Proteins (POCPs) above 50% has been proposed if they are to belong to the same genus (Konstantinidis and Tiedje, 2005; Qin et al., 2014). Finally, the Relative Synonymous Codon Usage (RSCU) has also been used by some authors to infer evolutionary and ecological links among bacterial species (Ma et al., 2015; Farooqi et al., 2016).

Very recently, Waite et al. (2017) carried out a comparative genomic analysis of the class *Epsilonproteobacteria*. Using 16S and 23S rRNA, 120 single-copy marker proteins and AAI analysis they proposed its reclassification as the new phylum Epsilonbacteraeota. In that study, Waite et al. (2017) also proposed a reclassification of the genus *Arcobacter* as a new

Family *Arcobacteraceae*, within the class *Campylobacteria*, order *Campylobacterales*. One weakness of this study, specifically regarding the genus *Arcobacter*, is that only seven validated species were included in the analysis. The new family therefore comprised only the genus *Arcobacter*. However, these findings also support the need for a clarification of the taxonomy of the current genus *Arcobacter*.

The rise of genome sequencing has dramatically changed the landscape of systematics of prokaryotes, improving different aspects such as the identification of species, the functional characterization for resolving taxonomic groups, and the resolution of the phylogeny of higher taxa (Whitman, 2015). It seems clear that the incorporation of genomics into the taxonomy will boost its credibility providing reproducible, reliable, highly informative means to infer phylogenetic relationships among prokaryotes, and avoiding unreliable methods and subjective difficult-to-replicate data (Chun and Rainey, 2014; Chun et al., 2018).

Within this modern taxonomy context, the objective of the present study was to reassess the taxonomy of the known and newly recognized *Arcobacter* species by using a MLSA of 13 housekeeping genes, the whole genome sequences and the derived genomic analysis. The latter analysis included ANI, *is*DDH, AAI, POCP, and RSCU of all *Arcobacter* type strains. In addition, phylogenies based on 16S and 23S rRNA gene sequences were also performed with comparative purposes. The new taxonomic criteria were stable when including whole genome sequences of a second strain of each species or of unassigned sequences obtained from the public databases.

MATERIALS AND METHODS

Bacterial Strains

All 27 valid species included in the genus *Arcobacter* have been studied. They are represented by 39 strains, and 13 strains that are potentially new species (**Table 1**). Furthermore, 50 genomes of *Arcobacter* strains identified at species level were investigated, 39 of which were obtained in our laboratory (27 from known species and 13 from potentially new species) and the others from the public databases[1,2]. Five genomes that had been deposited as *Arcobacter* sp. in the databases were also included in the study. If there was more than one strain of a known *Arcobacter* species, two representative genomes for each species were included in the analysis. The only exceptions were: *A. acticola* (Park et al., 2016) and *A. pacificus* (Zhang et al., 2015), whose taxonomic positions were only inferred by the phylogenetic analysis of the 16S rRNA gene sequences published in their species descriptions, together with a MLSA of three housekeeping genes (*atp*A, *gyr*B, and *rpo*B) for *A. pacificus* (Zhang et al., 2015; Park et al., 2016). The strains considered potentially new species, and named hereafter as 'candidate species,' had been recognized with an MLSA analysis of five housekeeping genes (*atp*A, *gyr*A, *gyr*B, *hsp*60, and *rpo*B) (data not shown).

Culturing for genome sequencing was carried out either on blood agar (DIFCO, Madrid, Spain) or marine agar (Scharlau, Sentmenat, Spain) at 30°C in aerobiosis for 24–72 h, depending on the requirements. DNA was extracted using Easy-DNA™ gDNA Purification kit (Invitrogen, Madrid, Spain) following the manufacturer's instructions. The integrity of the DNA was evaluated by electrophoresis of 10 μl of the sample in a 1.5% agarose gel. The total amount of DNA was quantified using Qubit™ with the dsDNA Broad Range Assay kit (Invitrogen). Paired-end libraries were constructed with 50 ng of DNA using Nextera DNA Library Preparation Kit (Illumina, Lisbon, Portugal) and sequenced with MiSeq platform (Illumina). Sequencing generated 2 × 300 bp paired-end reads. Clean reads were assembled with SPAdes (Nurk et al., 2013) and the CGE assembler (Larsen et al., 2012) in order to select the better assembly. Before depositing the genomes in the NCBI database, FASTA files were screened for eukaryotic and prokaryotic sequences using BLASTn, and for adaptors with VecScreen standalone software[3]. The five housekeeping genes used in the first MLSA analysis (*atp*A, *gyr*A, *gyr*B, *hsp*60, and *rpo*B) were extracted from each genome and compared with the Sanger sequences of these genes obtained originally for the identification of the strain. The existence of a single and identical copy of these genes confirmed that the genomes were not contaminated and belonged to the correct strain. Finally, contigs were deleted if they had less than 200 bp. The genomes were deposited in the GenBank database and **Table 1** lists the accession numbers.

The 55 genomes were annotated with a local installation of Prokka v1.2 (Seemann, 2014) using an e-value of 1e-06. The annotation was performed with Prokka, with the prediction tools Prodigal v2.6 (Hyatt et al., 2010) and ARAGORN v1.2 (Laslett and Canback, 2004). The prediction tool Barrnap v0.6[4] included in Prokka v1.2 was used for the annotation of rRNA genes. Coding sequences (CDS) were annotated, combining the Rapid Annotation Subsystems Technology (RAST) (Overbeek et al., 2014) using the *classic RAST* scheme and the Annotation Tools of PATRIC server (Wattam et al., 2017). The characteristics of each genome (i.e., N50, number of contigs, number of CDS, G+C content) were obtained from NCBI annotations.

Analysis of Housekeeping Genes, Ribosomal Genes, and Core Genome

Thirteen housekeeping genes (*atp*A, *atp*D, *dna*A, *dna*J, *dna*K, *fts*Z, *gyr*A, *hsp*60, *rad*A, *rec*A, *rpo*B, *rpo*D, and *tsf*) were obtained from the genomes using BLASTn search. Sequence similarities of housekeeping genes were determined using the MegAlign program (DNASTAR®, Madison, WI, United States). Genes were aligned using ClustalW (Larkin et al., 2007) and phylogenies based on individual genes and on the concatenated sequences was constructed with MEGA version 6.0 (Tamura et al., 2013) using the Neighbor-Joining (NJ) and Maximum-Likelihood (ML) algorithms.

[1] https://www.ncbi.nlm.nih.gov/genome/

[2] https://gold.jgi.doe.gov/

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

[4] http://www.vicbioinformatics.com/software.barrnap.shtml

The phylogenetic analysis of the core genome was assessed with the Roary software (Page et al., 2015) using 80% as cut-off for the BLASTp search. The core genome alignment was extracted with the latter software and the phylogeny was inferred using SplitsTree version 4.14.2 as described in Sawabe et al. (2007) using SplitsTree version 4.14.2, with a neighbor net drawing and Jukes-Cantor correction (Bandelt and Dress, 1992; Huson and Bryant, 2005).

Furthermore, the 16S and 23S rRNA genes of each genome were obtained using RNammer (Lagesen et al., 2007). In some cases, 16S rRNA gene sequences were obtained in our laboratories by Sanger sequencing or from the GenBank. The similarity of the 16S rRNA genes was calculated using MegAlign version 7.0.0 (DNASTAR®, Madison, WI, United States). Phylogenetic trees were reconstructed with MEGA version 6.0 (Tamura et al., 2013) also using the NJ and ML algorithms. Alignments obtained for both genes were visually analyzed in order to localize signature sequences for strains or groups of strains.

Genomic Indices

In order to ensure the correct assignation at species level of each analyzed genome, the ANI and the *is*DDH were calculated between all the genomes (Konstantinidis and Tiedje, 2005; Richter and Rosselló-Móra, 2009; Qin et al., 2014). The ANIb was calculated using JSpeciesWS (Richter et al., 2016), the resulting matrix was clustered and visualized using ggplot2 2.2.1 package (Wickham, 2009) and the *is*DDH was calculated with the GGDC software using results obtained with the formula 2 (Meier-Kolthoff et al., 2013). Two other indices (AAI and POCP) described for genus classification (Konstantinidis and Tiedje, 2005; Luo et al., 2014; Qin et al., 2014) were calculated among the genomes that corresponded to the type strains of the accepted species and the reference strains of the candidate species. The AAI was calculated with the Lycoming College Newman Lab AAIr Calculator[5] using the Sequence-Based Comparison Tools output file from RAST (Overbeek et al., 2014). The POCP was determined as described by Qin et al. (2014) using the following parameters to consider a peptide as a conserved protein: an e-value lower than 1e-5 and an identity percentage higher than 40% from an aligned region higher than 50%.

Finally, the RSCU was computed using the Codon Adaptation Index (CAI) developed by Sharp and Li (1987) through the CAIcal web-server (Puigbò et al., 2008). Statistical differences in the RSCU were assessed by a multinomial regression approach using the R software environment (R Core Team, 2015). The principal component analysis (PCA) was performed by the R software environment (R Core Team, 2015, and visualized using ggplot2 2.2.1 and ggfortify 0.4.4 (Wickham, 2009; Horikoshi and Tang, 2015; Tang et al., 2016) or pca3d 0.10 (Weiner, 2017) packages.

Phenotypic Analysis and Metabolic Inference

Phenotypic characterization of each described species was obtained from this study, from the original descriptions or from the summary published by On et al. (2017). For the potentially new *Arcobacter* species, the phenotype was characterized following the recommended minimal standards described for new taxa of the family *Campylobacteraceae* (Ursing et al., 1994; On et al., 2017) and with complementary tests used in the description of other *Arcobacter* species (Levican et al., 2013).

Inference of the metabolic routes from the genome sequences was performed with the software package Traitar (Microbial Trait Analyzer) (Weimann et al., 2016), using the protein coding genes files obtained with Prokka v1.2 (Seemann, 2014). Traitar software is based on phenotypic data extracted from the Global Infectious Disease and Epidemiology Online Network (GIDEON) and Bergey's Systematic Bacteriology. The software uses two prediction models: the phypat classifier, which predicts the presence/absence of proteins found in the phenotype of 234 bacterial species; and the phypat+PGL classifier, which uses the same information as the phypat combined with the information of the acquisition and loss of protein families and phenotypes during evolutive events. A total of 67 traits available within the software, related to oxygen requirement, enzymatic activities, proteolysis, antibiotic resistance, morphology and motility and the use of different carbon sources, were tested and the combined results of the two predictors were analyzed using a heat map.

RESULTS AND DISCUSSION

Strains and Genomes

All the 27 species currently included in the genus *Arcobacter* and 13 candidate species have been investigated in the present study, which has analyzed 55 genomes, 16 of them from the public databases and 39 sequenced in this study (**Tables 1**, **2**). It was not possible to analyze the genomes from *A. acticola* and *A. pacificus* because we were unable to get the type strains of the species. The contigs obtained and the N50 values complied with the recently proposed minimal standards for the use of genomes in taxonomic studies (Chun et al., 2018). The genome size ranged from 1.81 Mb for *A. skirrowii* F28 to 3.60 Mb for *A. lekithochrous* CECT 8942^T (**Table 2**). The G+C content ranged from 26.1% in *A. molluscorum* CECT 7696^T to 34.9% in '*A. aquaticus*' W112-28. The G+C values agree with the range from 24.6% (which corresponded to the type strain of *A. anaerophilus*) to 31% indicated for the genus *Arcobacter* in the recent emended description by Sasi-Jyothsna et al. (2013). Interestingly, 26 genomes (47.3%) showed the presence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and CRISPR-associated genes, related with the immune response of the bacteria.

Taxonomic and Phylogenetic Analysis

Similarities in the 16S rRNA gene sequences among type and representative strains of the different *Arcobacter* species (all the 27 species currently included in the genus and the 13 new candidate species) showed a wide range of values (**Supplementary Tables S1, S4**). They ranged from 90.8% (observed between *A. anaerophilus* and *A. faecis*) to 99.9% (between *A. butzleri and* '*A. lacus*'). The lower range of

[5]http://lycofs01.lycoming.edu/~newman/AAI/

TABLE 1 | Strains used in this study, source of isolation and accession numbers of the available genomes.

Species	Strain	Source	Acc. No. Genome	Species	Strain	Source	Acc. No. Genome
A. acticola	KCTC 52212^T	Seawater	NA[a]	*A. mytili*	T234	Seawater	PDJW00[b]
A. anaerophilus	DSM 24636^T	Estuarine sediment	PDKO00[b]	*A. nitrofigilis*	DSM7299^T	Marshland plant	NC014166[c]
	IR-1	Utsira aquifer	NZ_JXXG00[c]	*A. pacificus*	DSM 25018^T	Seawater	NA[a]
A. aquimarinus	CECT 8442^T	Mediterranean Sea	NXIJ00[b]	*A. skirrowii*	LMG 6621^T	Diarrheic lamb	NXIC00[b]
A. bivalviorum	CECT 7835^T	Mussels	PDKM00[b]		F28	Wild pig	PDJT00[b]
	F118-4	Mussels	PDKL00[b]	*A. suis*	CECT 7833^T	Pork meat	NREO00[b]
A. butzleri	RM4018^T	Human (Clinical)	NC_009850[c]	*A. thereius*	LMG 24486^T	Aborted pig foetus	LLKQ01[c]
	ED1	Microbial fuel cell	NC_017187[c]		DU22	Duck cloaca	LCUJ01[c]
A. canalis	F138-33	Oyster PNC[e]	NWVW01[b]	*A. trophiarum*	LMG 25534^T	Piglet feces	PDKD00[b]
	SH-4D_Col1	Unknown	FUYO00[c]		CECT 7650	Chicken cloacal swab	PDJS00[b]
A. cibarius	LMG 21996^T	Broiler, skin	NZ_JABW00[c]	*A. venerupis*	CECT 7836^T	Clams	NREP00[b]
A. cloacae	CECT 7834^T	Sewage	NXII00[b]	*Arcobacter* sp.	L	Microbial fuel cell	NC_017192[c]
	F26	Mussels	PDJZ00[b]		AF1028	Human feces	JART01[c]
A. cryaerophilus	LMG 24291^T	Aborted bovine foetus	NXGK00[b]		CAB	Marine	Go0012496[d]
A. defluvii	CECT 7697^T	Sewage	NXIH00[b]		LA11	Marine	BDIR01[c]
A. ebronensis	CECT 8441^T	Mussels	PDKK00[b]		LPB0137	Environmental	CP019070[c]
	CECT 8993	Seawater	PDKJ00[b]				
A. ellisii	CECT 7837^T	Mussels	NXIG00[b]	'*A. aquaticus*'	W112-28	Freshwater PNC[e]	PDKN00[b]
A. faecis	LMG 28519^T	Human septic tank	NZ_JARS00[c]	'*A. caeni*'	RW17-10	Recycled wastewater	MUXE00[b]
A. halophillus	DSM 18005^T	Hypersaline lagoon	PDJY00[b]	'*A. hispanicus*'	FW-54	Wastewater	PDKI00[b]
	F166-45	Oyster PNC[e]	PDJY00[b]	'*A. lacus*'	RW43-9	Recycled wastewater	MUXF00[b]
A. lanthieri	LMG 28516^T	Pig manure	JARU01[c]	'*A. mediterraneus*'	F156-34	Mussels Alfacs Bay	NXIE00[b]
	LMG 28517	Dairy cattle manure	JARV01[c]	'*A. miroungae*'	9Ant[f]	Cloaca elephant seal	PDKH00[b]
A. lekithochrous	CECT 8942^T	Great scallop larvae	NZ_MKCO00[b]	'*A. neptunis*'	F146-38	Mussels Alfacs Bay	PDKG00[b]
	LMG 28652	Abalon	PZYW00[c]	'*A. porcinus*'	LMG 24487^T	Aborted pig foetus	LCUH01[c]
A. marinus	CECT 7727^T	Seawater	NXAO01[b]	'*A. ponticus*'	F161-33	Cockle Alfacs Bay	PDKF00[b]
	F140-37	Clams Alfacs Bay	NWVX01[b]	'*A. salis*'	F155-33	Oyster PNC[e]	PDKE00[b]
A. molluscorum	CECT 7696^T	Mussels	NZ_NXFY00[b]	'*A. viscosus*'	F142-34[g]	Mussels PNC[e]	PDKC00[b]
	F91	Mussels	PDJX00[b]	'*A. vitoriensis*'	FW59[g]	Wastewater	PDKB00[b]
A. mytili	CECT 7386^T	Mussels	NXID00[b]	*Arcobacter* sp.	F2176	Mussels	PDJV00[b]

[a]*Genome not available;* [b]*Genome sequenced in this study;* [c]*Genome obtained from NCBI database;* [d]*Genome obtained from JGI Gold atabase;* [e]*PNC means PobleNou Channel, which is a freshwater channel heavily (geometric mean of E. coli counts 4.1 × 10^4 c.f.u./100ml) contaminated with wastewater where shellfish were exposed for 72h (Salas-Massó et al., 2016, 2018).* [f]*This strain was obtained from F.J. García from the Laboratorio Central de Veterinaria de Algete, MAGRAMA, Madrid, Spain;* [g]*These strains were recovered at the Faculty of Pharmacy, University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain, by R. Alonso, I. Martinez-Malaxetxebarria and A. Fernández-Astorga.*

similarity (90.8%) is due to the fact that those species, as occurred with others, were assigned within the genus based on the premise that 16S rRNA gene similarity was higher with any type strain of *Arcobacter* than with other taxa. However, in some cases being below the 95% cut-off value for genus delimitation (Rosselló-Mora and Amann, 2001; Yarza et al., 2008; Tindall et al., 2010; Figueras et al., 2011a,b). It is interesting to point out that 16S rRNA gene sequence similarities among *A. nitrofigilis*, the type species of the genus, and the other described species ranged from 93.2% (with *A. thereius*) to 95.9% (with *A. venerupis*). Furthermore, *A. nitrofigilis* showed higher similarities than the threshold value of 95% with only seven species (*A. acticola*, '*A. caeni*,' *A. cloacae*, *A. defluvii, A. ellisii*, *A. suis*, and *A. venerupis*) out of the 27 accepted species. In any case, from the analysis of the similarities in the 16S rRNA gene sequences among the *Arcobacter* species it is clear that this gene has limited value and that other approaches available in the genomic era of taxonomy are needed for their study.

Phylogenetic analysis based on the core genome made up of 286 genes (**Figure 1** and **Supplementary Table S5**) and also on the concatenated sequences of 13 housekeeping genes of the representative *Arcobacter* strains (**Figure 2**) revealed that the *Arcobacter* species could be grouped into 4 major monophyletic clusters. Cluster 1, comprised seven validated species: *A. butzleri*, *A. cibarius*, *A. cryaerophilus*, *A. lanthieri*, *A. skirrowii*, *A. thereius*, and *A. trophiarum*, together with *A. faecis* (species described but not validated yet) and five candidate taxa '*A. hispanicus,*' '*A. lacus,*' '*A. miroungae,*' '*A. porcinus,*' and '*A. vitoriensis*' (**Figure 1**). Cluster 2 embraced the species *A. aquimarinus*, *A. cloacae*, *A. defluvii*, *A. ellisii*, *A. suis*, and *A. venerupis*, as well as the non-validated *A. acticola* and the candidatus '*A. caeni*.' Cluster 3 included five species, *A. canalis*, *A. halophilus*, *A. marinus*, *A. molluscorum*,

TABLE 2 | Genome characteristics and annotation results. Source of whole genome sequences as indicated in **Table 1**.

Species	No. Contigs	N50 (Kb)	CDS (Total)	CDS (Coding)	RNA Genes	tRNAs	ncRNAs	CRISPR Arrays	G+C (%)	Size (Mb)
A. anaerophilus DSM 24636^T	40	186	2,938	2,922	45	40	2	1	29.9	2.98
A. anaerophilus IR1	7	1,179	3,360	3,024	61	47	2	3	30.2	3.25
'A. aquaticus' W112-28^T	20	370	2,500	2,487	55	45	3	0	34.9	2.53
A. aquimarinus CECT 8442^T	68	75	2,473	2,463	46	42	2	0	26.6	2.46
A. bivalviorum CECT 7835^T	179	461	2,786	2,728	50	41	3	0	28.2	2.75
A. bivalviorum F118-4	26	209	2,652	2,652	47	38	3	0	28.1	2.71
A. butzleri RM4018^T	1	–	2,261	2,256	71	54	2	0	27.0	2.34
A. butzleri ED1	1	–	2,151	2,145	71	54	2	0	27.1	2.26
'A. caeni RW17-10^T	59	123	2,357	2,337	58	51	3	0	27.1	2.42
A. canalis CECT8984^T	50	166	2,733	2,720	53	48	2	1	27.3	2.78
A. canalis SH-4D_Col1	69	72	2,716	2,663	63	52	2	1	27.1	2.82
A. cibarius LMG 21996^T	44	119	2,156	2,110	68	46	2	0	27.1	2.20
A. cloacae CECT 7834^T	135	135	2,826	2,795	58	51	2	3	26.8	2.78
A. cloacae F26	40	218	2,470	2,459	53	44	2	1	26.9	2.51
A. cryaerophilus LMG 24291^T	91	54	2,092	2,081	49	40	3	0	27.2	2.06
A. defluvii CECT 7697^T	80	166	2,921	2,894	57	49	2	2	26.3	2.94
A. ebronensis CECT 8441^T	103	188	3,089	3,072	47	39	3	1	29.2	3.15
A. ebronensis W129-34	126	217	3,206	3,171	46	40	3	2	29.2	3.23
A. ellisii CECT 7837^T	135	177	2,875	2,840	64	52	2	1	26.9	2.80
A. faecis LMG 28519^T	55	127	2,429	2,376	76	53	2	1	27.2	2.50
A. halophilus DSM 18005^T	111	56	2,677	2,660	54	46	3	2	27.4	2.75
A. halophilus F166-45	90	56	2,879	2,864	59	51	2	2	27.0	2.96
'A. hispanicus' FW54^T	76	148	2,228	2,207	46	40	3	1	26.4	2.21
'A. lacus' RW43-9^T	24	295	2,194	2,182	47	40	2	0	26.8	2.22
A. lanthieri LMG 28516^T	29	466	2,223	2,190	73	52	3	1	26.7	2.29
A. lanthieri AF1581	24	353	2,199	2,186	88	57	3	0	26.8	2.26
A. lekithochrous CECT 8942^T	436	343	3,628	3,316	88	75	3	0	28.6	3.61
A. lekithochrous LMG 28652	82	343	3,499	3,330	61	55	3	0	28.2	3.50
A. marinus CECT 7727^T	162	54	2,809	2,781	55	50	2	0	27.0	2.87
A. marinus F140-37	76	67	2,725	2,652	59	48	2	0	27.0	2.78
'A. mediterraneus' F156-34^T	29	689	2,769	2,750	47	41	3	1	27.3	2.83
'A. miroungae' 9Ant^T	35	363	1,868	1,847	46	41	2	1	28.1	1.84
A. molluscorum CECT 7696^T	117	121	2,746	2,736	58	49	3	6	26.1	2.76

(Continued)

TABLE 2 | Continued

Species	No. Contigs	N50 (Kb)	CDS (Total)	CDS (Coding)	RNA Genes	tRNAs	ncRNAs	CRISPR Arrays	G+C (%)	Size (Mb)
A. molluscorum F91	240	150	2,951	2,889	71	58	3	2	26.3	2.89
A. mytili CECT 7386^T	126	70	2,950	2,934	58	48	3	1	26.3	2.97
A. mytili T234	145	37	2,735	2,723	54	48	3	0	26.4	2.77
'A. neptunis' F146-38^T	36	267	2,627	2,614	57	45	3	0	27.1	2.65
A. nitrofigilis DSM 7299^T	1	–	3,101	3,086	69	55	2	1	28.4	3.19
'A. ponticus' F161-33	24	597	2,632	2,621	46	36	3	0	28.1	2.74
'A. porcinus' LMG 24487^T	70	123	2,186	2,112	47	41	2	0	27.0	2.14
'A. salis' F155-33^T	153	169	2,932	2,904	50	43	3	0	29.0	2.93
A. skirrowii LMG 6621^T	62	306	2,029	2,006	48	42	2	2	27.7	1.97
A. skirrowii F28	110	40	1,911	1,897	46	41	2	0	27.8	1.81
A. suis CECT 7833^T	122	142	2,646	2,613	57	52	2	0	27.3	2.62
A. thereius LMG 24486^T	2	1,039	1,896	1,883	57	46	2	3	27.0	1.91
A. thereius DU22	19	252	2,006	1,983	47	42	2	1	26.8	2.01
A. trophiarum CECT 7650	37	152	1,911	1,894	48	37	3	0	28.0	1.90
A. trophiarum LMG 25534^T	266	86	2,167	2,071	49	41	3	0	29.4	2.00
A. venerupis CECT 7836^T	234	182	3,319	3,267	64	52	2	0	28.0	3.28
'A. viscosus' F142-34^T	82	65	2,772	2,756	55	48	3	1	26.6	2.79
'A. vitoriensis' FW59^T	144	179	2,617	2,570	53	46	2	0	27.4	2.58
Arcobacter sp. CAB	367	20	3,596	3,392	NA	31	NA	NA	28.2	3.48
Arcobacter sp. F2176	99	178	3,212	3,186	67	57	2	0	28.1	3.27
Arcobacter sp. LA11	53	229	3,006	2,961	49	43	3	0	27.9	3.10
Arcobacter sp. LPB0137	1	–	2,731	2,698	85	64	2	0	27.7	2.87
Arcobacter sp. L[a]	1	–	2,847	2,834	73	56	2	1	26.6	2.95
Arcobacter sp. AF1028[b]	46	148	2,336	2,285	71	51	2	1	27.2	2.41

[a]*Genome sequenced in this study;* [b]*Genome obtained from NCBI database;* [c]*Genome obtained from JGI Gold database. Our results show that these strains belong to the species.* [d]*A. defluvii and* [e]*A. faecis.*

and *A. mytili*, together with two candidates, '*A. neptunis*' and '*A. viscosus.*' Finally, Cluster 4 included the species *A. anaerophilus*, *A. bivalviorum*, and *A. ebronensis*, as well as the candidates '*A. mediterraneus,*' '*A. ponticus,*' and '*A. salis.*' The split decomposition network analysis of the core genome showed that the species *A. lekithochrous* CECT 8942^T and *A. nitrofigilis* DSM 7299^T appeared as orphan species. Furthermore, with this analysis the candidatus '*A. aquaticus*' W112-28 also appeared in a separate branch near to *A. nitrofigilis* DSM 7299^T. On the other hand, both analyses, MLSA and core genome, confirmed the existence of two sub-clusters in Cluster 1 (again *A. butzleri* and '*A. lacus*' were located in the most distant branch within the cluster), and also two subgroups could be observed in Cluster 4, one comprising the species *A. anaerophilus* and *A. ebronensis*, and the other including the rest of species within this cluster (**Figures 1**, **2**). All the clusters and sub-clusters showed a similarity in the concatenated sequences of the 13 housekeeping genes higher than 85% (**Figure 2**).

Phylogenies based on the 16S and 23S rRNA gene sequences, undertaken with the NJ and ML approacheserealso constructed with comparative purposes. 16S rRNA based tree showed also the four major clusters although less defined (**Supplementary Figure S1A**). Species within Cluster 1, showed 16S rRNA gene sequence similarities ranging from 96.1 to 99.9%. Cluster 2 yielded similarities among species for the 16S rRNA gene between 96.7 and 99.6%, whereas within Cluster 3 ranged between 93.0 and 99.1%. Finally, Cluster 4 included species with a range of 16S rRNA sequence similarity from 94.0 to 99.5%. With the exception of Cluster 3, similarity values within the clusters (>94–95%) were within the classical boundaries for genus assignation in bacterial taxonomy (Rosselló-Mora and Amann, 2001; Yarza et al., 2008, 2014; Tindall et al., 2010; Figueras et al., 2011a,b). Our results agree with those from a recent study by Yarza et al. (2014), who investigated 568 taxa and described a threshold in 16S rRNA sequence identity of 94.5% for genus delineation.

Similar groups and topology, with only minor differences, were obtained when the 23S rRNA gene sequences were used to analyze the phylogeny of the genus (**Supplementary Figure S2**). In this analysis, the recently described species *A. acticola*, and *A. pacificus* could not be included because of the unavailability of the type strains and/or whole genome sequences. The same four major clusters formed in the 23S rRNA gene phylogenetic tree, and the species *A. lekithochrous* and *A. nitrofigilis* appeared also as orphan species (**Supplementary Figure S2**). Within Cluster 1 two subgroups could also be obtained, differentiating the species *A. butzleri* and '*A. lacus*' from the rest of the species. Similarly, the species *A. anaerophilus* and *A. ebronensis* formed a differentiated subgroup in Cluster 4.

The visual analysis of the alignments obtained with the sequences of the 16S and 23S rRNA genes allowed the localization of signature motifs, especially in the 16S rRNA gene, for the different clusters established in the phylogenetic analysis. In these sequences, a total of 16 locations were found, presenting nucleotide combinations characteristic for the clusters (**Supplementary Figure S3**). Some of these motifs were located in helix regions as interactions with proteins of the ribosomal 30S subunit, such as helix 21 (region V4) or helix 28/44 (region V9), and therefore had a considerable level of protection against mutations (Adilakshmi et al., 2008; Kitahara et al., 2012). There are some studies on the presence of signature regions with taxonomic/phylogenetic implications in the ribosomal genes (Martínez-Murcia et al., 1992, 2007; Ue et al., 2011;Řeháková et al., 2014; Martínez-Murcia and Lamy, 2015). Some regions with signature motifs detected in the present study have also shown implications for phylogenetic analysis in cyanobacteria, including regions H15, H17, H21, H22-H23, H41, and H44 (Řeháková et al., 2014). A tree was also constructed weighting such positions (**Supplementary Figure S1B**), which allowed a better definition of the main clusters observed with the whole 16S rRNA sequences although, as expected, differentiation among species within each cluster was lower. Two sub-clusters were observed in Cluster 1, where the species *A. butzleri* and '*A. lacus*' grouped into a well-differentiated branch with respect to the other species in the cluster (**Supplementary Figure S1B**). In this analysis, *A. pacificus* was clearly located in the Cluster 3, whereas in Cluster 4, *A anaerophilus* was the borderline species, while *A. ebronensis* and '*A. mediterraneus*' were located in an independent branch (**Supplementary Figure S1B**). Therefore, the signature motifs described here might be a new tool for identification of the different clusters and/or genus.

Genomic Indices

The results of the calculations of the ANI and the *is*DDH among the 36 studied genomes are given in the **Supplementary Table S2** and **Supplementary Figure S4**. The results of the ANI and *is*DDH calculations showed that the genomes grouped into the same clusters observed by the analyses of the MLSA of the 13 housekeeping and core genes (**Figures 1**, **2**). Ranges of ANI within each cluster were from 75.2 to 95.4%, whereas *is*DDH values were between 19.5 and 65.4% (**Figure 2** and **Table 3**). These results confirm the phylogenetic analysis for the 13 new candidate species because all of them showed ANI and *is*DDH values of <96% and <70%, respectively, which are the cut-off values proposed for the delineation of new species (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Figueras et al., 2017). As discussed in other studies, the ANI and *is*DDH indices provided reliable information for the delineation of *Arcobacter* species and are also included in the minimal guidelines to define species using genomes (Whiteduck-Léveillée et al., 2015, 2016; Figueras et al., 2017; Chun et al., 2018). Although those indices are not considered useful for delimiting genera, each of the four clusters showed values that ranged between 75.2 and 81.8% as their lowest ANI, which might be the suitable range for separating different, closely related genera. These values are relatively similar to those reported by Qin et al. (2014) that found 68–82% interspecies ANI values among the genera that they studied. Values of ANI obtained for the candidate species '*A. aquaticus*' were lower than the other results, from 70.0% with *A. cryaerophilus* LMG 24291^T to 71.9% with *A. bivalviorum* CECT 7835^T and more in line with the Qin et al. (2014) results of 68% (**Supplementary Table S2**). In the case of the *is*DDH the lower values among species in the same cluster ranged between 19.5 and 24.8%, and again these might be the levels associated to different genera.

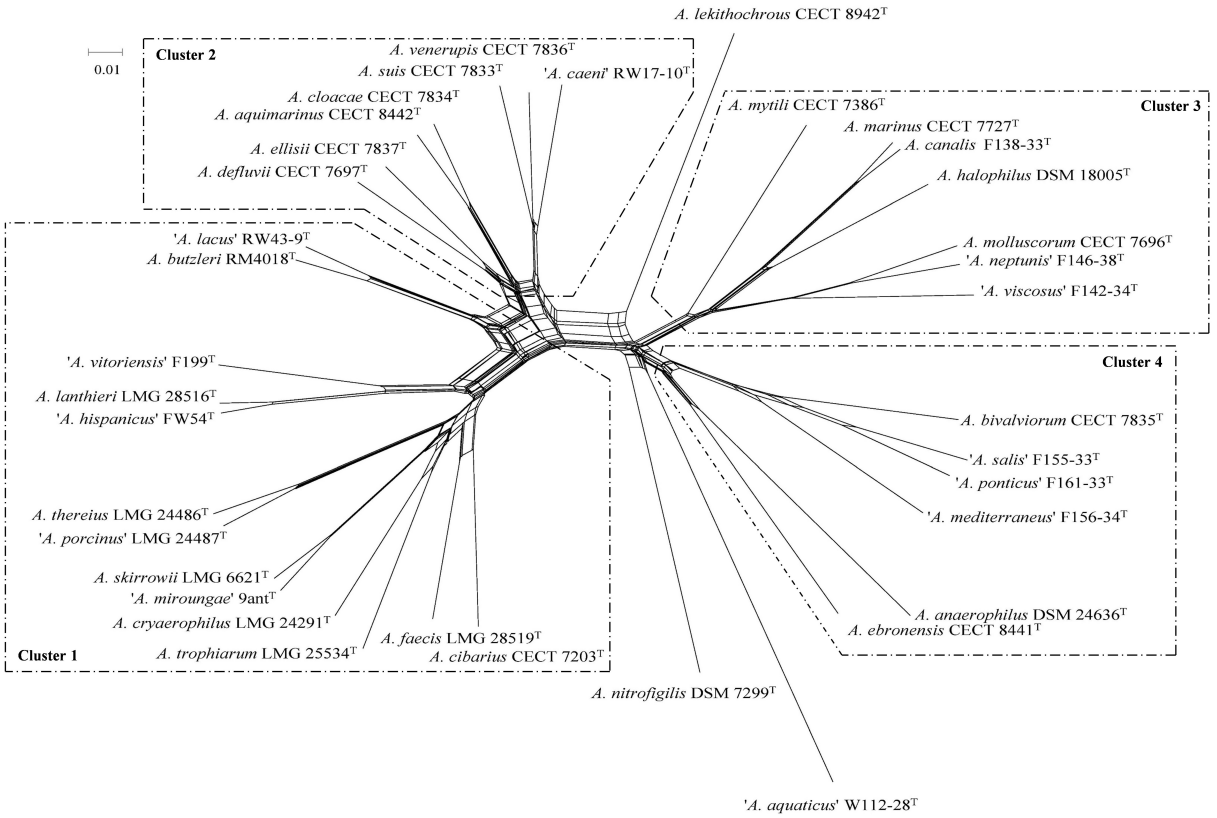


FIGURE 1 | Split decomposition network constructed with the concatenated sequences of 284 core genes from the genomes of 36 type and representative strains of *Arcobacter*. Scale bar, base substitutions per site.

With the aim of confirming if the clusters observed might represent different genera, as suggested by the phylogenetic analyses, the similarity indices AAI and POCP were also calculated (**Supplementary Table S3**). In agreement with the 60–80% AAI that have been described for species belonging to the same genus (Konstantinidis and Tiedje, 2005) all our clusters showed lower ranges of between 67.6 to 80.3% (**Table 3**). All the clusters also complied with the POCP proposed for genus separation above 50% (Luo et al., 2014; Qin et al., 2014) because as shown in **Table 3** all clusters showed the lowest values from 67.0 to 75.4%.

It is widely known that synonymous codon usage varies among organisms and that it is related to differences in G+C content, replication strand skew, or gene expression (Suzuki et al., 2008; Farooqi et al., 2016). The interaction of these factors may vary among species depending on their evolutionary process (Ma et al., 2015). It has also been suggested that the extent of codon usage bias plays a role in the adaptation of prokaryotic organisms to their environments and lifestyles (Botzman and Margalit, 2011). To analyze the overall codon usage trends of the *Arcobacter* species, the frequencies of the different codons were obtained from the whole genomes and the RSCU was computed using the CAI, which is a useful tool for estimating codon usage bias (Ma et al., 2015; Farooqi et al., 2016). A first finding was that all the *Arcobacter* species presented a preferential use of the codons finishing in A or T (**Supplementary Figure S5**), which might be expected due to their low G+C% content. The characteristic pattern showed by *A. aquaticus* is noteworthy (**Supplementary Figure S5**), which supports its differentiation from the other species in Cluster 3 as well as its unique taxonomy. Such difference was the only statistically significant ($p < 0.05$) in the multinomial regression analysis carried out.

Next, the codon usage trends were analyzed by PCA to reveal possible evolutionary relationships. Interestingly, different groups of strains could be observed in the three-dimensional graphic (**Figure 3**), which correlated with those clusters established in the different phylogenetic analyses, as shown above. As reported previously for different species of *Mycoplasma* (Marenda et al., 2005; Ma et al., 2015), PCA provides an additional pathway to investigate the evolutionary direction of the *Arcobacter* species. In addition, similarities in the synonymous codon usage patterns might reflect similar lifestyles (pathogenic vs. non-pathogenic) and adaptation to certain environments (marine water, shellfish, etc.).

Metabolic Inference and Phenotypic Analysis

Phylogenetic and genomic analysis confirmed the existence of four clusters among the validated and candidate *Arcobacter*

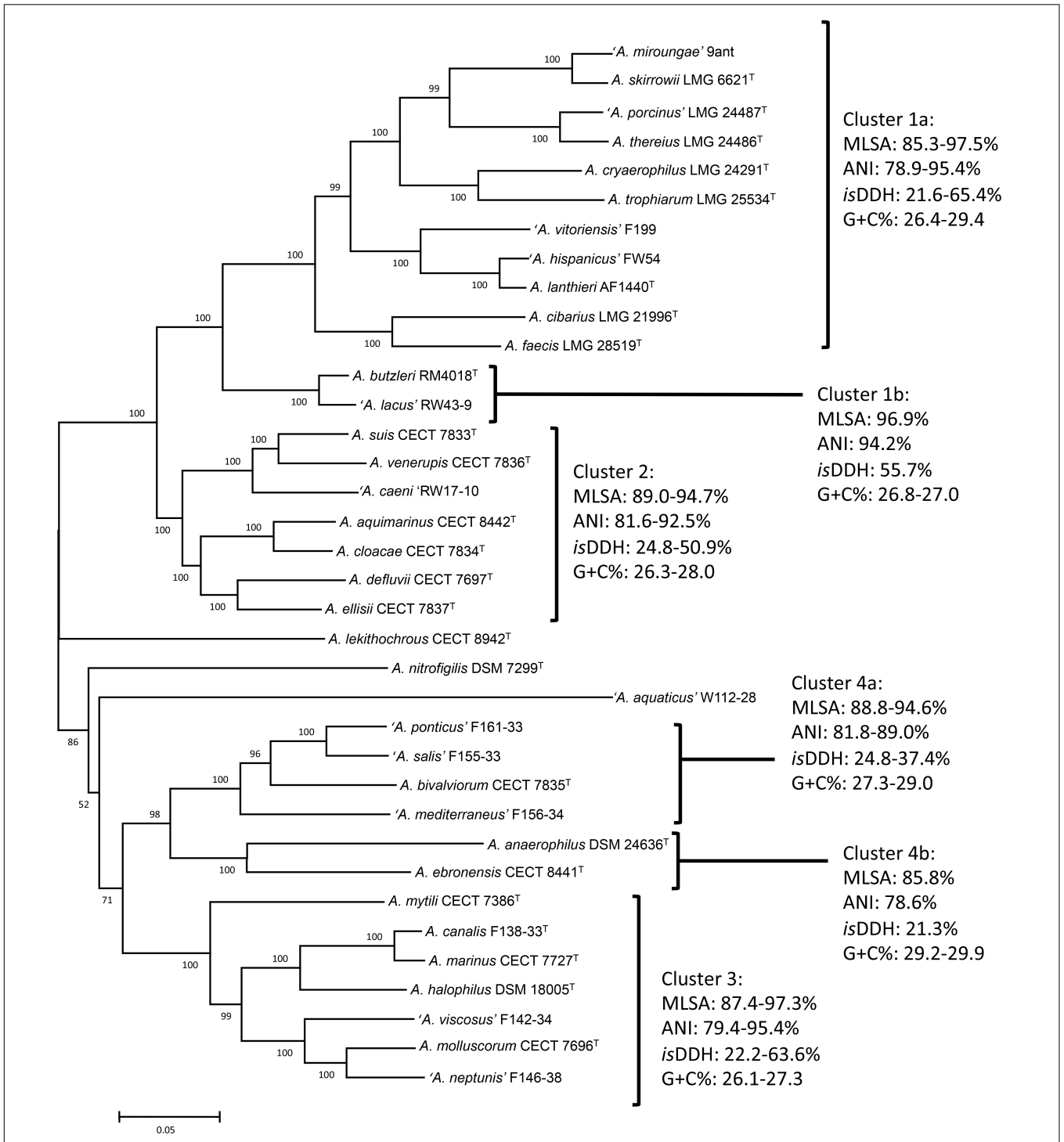


FIGURE 2 | Phylogenetic tree constructed with 36 type and representative strains of *Arcobacter* species based on concatenated sequences of 13 housekeeping genes by the Maximum-Likelihood algorithm (model GTR+G+I). Numbers at nodes denote the level of bootstrap based on 1,000 replicates; only values greater than 50% are shown. Scale bar, base substitutions per site.

species, which comply with the cut-off values established for the differentiation of independent genera. A thorough phenotypic analysis was therefore carried out to determine if the description of new taxa at genus level was possible or if such clusters were only clades or genomovars within the genus *Arcobacter*. In fact, this is what has occurred in a recent polyphasic study of 52 *A. cryaerophilus* strains (including genome information) in which, despite four different genomospecies

TABLE 3 | Intra-cluster similarities (%) obtained for the 16S rRNA gene and for the different genomic indexes analyzed.

	16S RNA gene		MLSA		ANI		*is*DDH		AAI		POCP		G+C% (mol)	
Cluster 1a	96.8–99.8	96.1–99.9	85.3–97.5	85.3–97.5	78.9–95.4	77.7–95.4	21.6–65.4	20.5–65.4	72.5–95.0	71.5–95.0	68.2–95.6	67.0–95.6	26.4–29.4	26.4–29.4
Cluster 1b	99.9		96.9		94.2		55.7		93.7		84.3		26.8–27.0	
Cluster 2	96.7–99.6		89.0–94.7		81.6–92.5		24.8–50.9		73.1–93.5		71.7–87.5		26.3–28.0	
Cluster 3	94.2–99.1		87.4–97.3		79.4–95.4		22.2–63.6		67.6–95.7		75.4–91.4		26.1–27.3	
Cluster 4a	96.6–99.5	94.0–99.5	88.8–94.6	85.8–94.6	81.8–89.0	75.2–89.0	24.8–37.4	19.5–37.4	80.3–83.4	68.7–83.4	74.4–90.7	71.6–90.7	27.3–29.0	27.3–29.9
Cluster 4b	96.9		85.8		78.6		21.3		78.4		77.9		29.2–29.9	

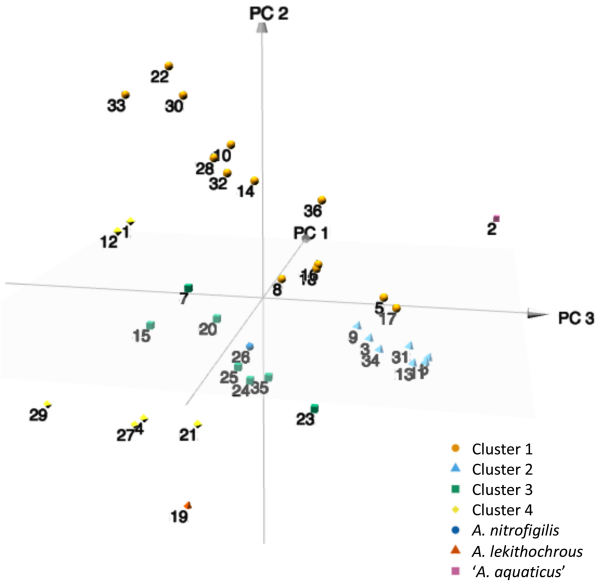


FIGURE 3 | 3D plot of the three major axes generated by principal component analysis (PCA) of the RSCU values computed for the 36 type and representative strains of *Arcobacter* species. 1, *A. anaerophilus* DSM 24636^T; 2, '*A. aquaticus*' W112-28; 3, *A. aquimarinus* CECT 8442^T; 4, *A. bivalviorum*; 5, *A. butzleri* RM4018^T; 6, '*A. caeni*' RW17-10; 7, *A. canalis* CECT 8984^T; 8 *A. cibarius* LMG 21996^T; 9, *A. cloacae* CECT 7834^T; 10, *A. cryaerophilus* LMG 24291^T; 11, *A. defluvii* CECT 7697^T; 12, *A. ebronensis* F128-2^T; 13, *A. ellisii* CECT 7837^T; 14, *A. faecis* AF1078^T; 15, *A. halophilus* DSM 18005^T; 16, '*A. hispanicus*' FW54; 17, '*A. lacus*' RW43-9; 18, *A. lanthieri* AF1440^T; 19, *A. lekithochrous* LFT1.7^T; 20, *A. marinus* CECT 7727^T; 21, '*A. mediterraneus*' F156-34; 22, '*A. miroungae*' 9Ant; 23, *A. molluscorum* CECT 7696^T; 24, *A. mytili* W112-28; 25, '*A. neptunis*' F146-38; 26, *A. nitrofrigilis* DSM 7299^T; 27, '*A. ponticus*' F161-33; 28, '*A. porcinus*' LMG 24487; 29, '*A. salis*' F155-33; 30, *A. skirrowii* LMG 6621^T; 31, *A. suis* CECT 7833^T; 32, *A. thereius* LMG 24486^T; 33, *A. trophiarum* LMG 25534^T; 34, *A. venerupis* CECT 7836^T; 35, '*A. viscosus*' F142-34; 36, '*A. vitoriensis*' F199.

being recognized, the phenotypic characterization did not allow their differentiation into separate species and were therefore considered genomovars (Pérez-Cataluña et al., 2018a).

Phenotypic inference using Traitar confirmed the lack of reaction of *Arcobacter* species to most of the tests commonly used for bacterial identification (**Supplementary Figure S6**). Thus, all the type and representative strains rendered negative results, regardless of the predictor employed, for use as the sole carbon source of sugars (D-Mannitol, D-Mannose, Salicin, or Trehalose, among others) and carboxylic acids (Citrate or Malonate). Such results have been previously reported in the original descriptions of the species (see review of On et al., 2017). On the other hand, there was some incongruence between results from Traitar and those obtained by classical characterization for some tests, including growth on MacConkey agar or urea hydrolysis (data not shown). A possible explanation is related with the macro-accuracy of the predictors employed in the Traitar analysis (82.6–85.5%), as reported in the original description of the microbial trait analyzer (Weimann et al., 2016). The fact that

some of the *Arcobacter* species studied are halophilic cannot be ignored, since some of the media usually employed in the wet-lab characterization are developed for non-halophilic microorganisms.

The heat maps built from the combined results of both predictors in the Traitar analysis revealed the existence of similarity groups regarding the metabolic characteristics of the *Arcobacter* type strains (**Supplementary Figure S6**). In most case, clustering of strains supported the groups obtained with genomic tools, although some incongruence was also observed, such as for *A. butzleri* (better related here to *A. defluvii, A. ellisii* or *A. cloacae*), *A. mytili* (closest Traitar species '*A. caeni*') or *A. venerupis* (forming a branch with *A. ebronensis* and '*A. ponticus*'). In any case, Traitar might be helpful as a first-step method for phenotypic inference, although further verification should be made, especially in environmental bacterial species with special growth requirements (i.e., halophilic conditions).

A deep review of the characteristics reported in the original descriptions of the *Arcobacter* species, together with results obtained in our respective laboratories, allowing phenotypic traits to differentiate the clusters established by the phylogenetic and genomic analyses (**Table 4**). Growth at 37°C in microaerophilic condition, the halophilic character, the ability to grow in presence of glycine, safranin, oxgall, or triphenyltetrazolium chloride (TTC), the presence of some enzymatic activities, such as catalase, urease or indoxyl acetate hydrolysis, and resistance to cefoperazone among others, were the main differentiating traits. Most of these characters are included in the minimal standards for describing new species in the families *Campylobacteraceae* and *Helicobacteraceae* (On et al., 2017), and they should, therefore, also be maintained for the new family *Arcobacteraceae* proposed by Waite et al. (2017), once this taxonomical change is validated. The phenotypic differentiation proposed in **Table 4** enabled to further describe the new genera that corresponded to the different clusters of *Arcobacter* species determined in the present study.

Stability of the Genomic-Based Clustering

In order to test the stability of the new taxonomical scheme proposed, we analyzed the whole genome sequences using second strains from each species or from unassigned sequences obtained from the public databases. That analysis is shown in **Supplementary Figure S7** and included 55 genomes. These new phylogenetic analyses of the core genome also using a Split network showed that the four clusters were maintained, but the two clusters (Clusters 3 and 4) that include species able to grow in media containing 2.5% NaCl appeared in the right place (**Supplementary Figure S7**). The genome of *Arcobacter* sp. LPB0137 obtained from the NCBI database grouped with the species *A. lekithochrous* CECT 8942^T, while the genomes *Arcobacter* sp. LA11 and CAB grouped together in a separate branch near to Cluster 4. Interestingly, the ANI and *is*DDH values of 91.4% and 45.8% between strain F2176, previously identified as *A. nitrofigilis* (Figueras et al., 2008), and the type strains of this species along with the phylogenetic position (**Supplementary Figure S7**), revealed that this strain belonged to another potentially new species. Furthermore, strains L and AF1028, deposited at the NCBI database as *Arcobacter* sp. were identified as *A. defluvii* and *A. faecis*, respectively, because they clustered with the type strains of those species (**Supplementary Figure S7**). This was also confirmed by the ANI and *is*DDH results being above 96% and 70%, respectively.

Collado and Figueras (2011), in their review about the epidemiology and clinical significance of the genus *Arcobacter*, reported that these bacteria should be considered quite atypical within the class *Epsilonproteobacteria* because of the great diversity of hosts and habitats from which they have been isolated. In order to show if the clusters obtained have a relationship with their ecological habitat, the origin of each strain is also given in **Supplementary Figure S7**. Despite the fact that only two strains from each species were included in the analysis, each of the clusters embraced species that had been recovered from common or related origins. Cluster 1 included by strains isolated from humans and animals, from wastewater and from broiler skin (*A. cibarius* CECT 7203^T). The fact that some strains isolated from wastewater that was contaminated by humans or animal excreta, gives evidence of the relationship of these sources. This finding agrees with the high abundance of *Arcobacter* in wastewater and in water contaminated with fecal pollution (Collado et al., 2008, 2010). Among the species of Cluster 1, both by metagenomics analysis or direct plating without enrichment (Fisher et al., 2014; Levican et al., 2016), the species *A. cryaerophilus* was the prevalent species in wastewater, while the species *A. butzleri* is normally predominant in studies that investigate water and food samples of animal origin, such as different types of meats using an enrichment step (Collado et al., 2009b; Collado and Figueras, 2011; Hsu and Lee, 2015; and references therein). So far, only the species *A. cryaerophilus*, *A. thereius*, *A. trophiarum, A. cibarius* or *A. skirrowii* have been recovered from humans or animals (De Smet et al., 2011; Figueras et al., 2014; Van den Abeele et al., 2014) and all these species are as commented in the same cluster.

Cluster 2 included strains from different origins but was dominated by species that came from wastewater, shellfish or food products. In this sense, *A. defluvii* CECT 7697^T and '*A. caeni*' RW17-10 were isolated from wastewater, while the strain *A. defluvii* L was recovered from a microbial fuel cell. Strains of *A. defluvii* have also been recovered from shellfish in other studies (Levican et al., 2014; Salas-Massó et al., 2016). The strain *A. suis* CECT 7833^T was isolated from pork meat, but other isolates have also been obtained from buffalo milk in Italy (Levican et al., 2013; Giacometti et al., 2015). The other five strains in the cluster were isolated from shellfish, wastewater and seawater (**Table 1** and **Supplementary Figure S7**). The other two clusters (Clusters 3 and 4) included strains isolated from seawater shellfish giving evidence of the marine origin of these clusters. The orphan species (*A. nitrofigilis* DSM7299^T, *A. lekithochrous* CECT 8942^T, and '*A. aquaticus*' W112-28) also corresponded

TABLE 4 | Differential phenotypic traits among the different clusters of *Arcobacter* species obtained on the basis of the characteristics of the type and representative strains of the species included in each group.

Test	*A. nitrofigilis*	Cluster 1	Cluster 2	Cluster 3	Cluster 4	*A. lekithochrous*	*A. aquaticus*
Growth at/on							
CO_2 37°C	−	V	V	+	V	−	+
0.5% NaCl	−	+[a]	+	V	−[b]	−[c]	+
4% NaCl	+	−	−	+	+	−	−
1% Glycine	−	V	−	V	V	−	−
0.05% Safranin	−	+	V	V	V	+	+
0.04 TTC	−	V	−	−	−	+	−
1% Oxgall	−	V	V	−	−[d]	−	−
CCDA	−	V	V	−[e]	−	+	+
Enzymatic activities							
Catalase	−	+[f]	+	V	V	+	−
Urease	+	−	V	−	−[d]	−	−
Indoxyl acetate hydrolysis	+	+[f]	+	V	V	−	−
Nitrate reduction	+	V	+	−[g]	V	−	−
Resistance to cefoperazone (64 mg/l)	ND	V	−	V	−	−	+

+, positive result; −, negative result; V, variable result in all the species of the cluster; [a]With the exception of A. skirrowii; [b]With the exception of A. pacificus; [c]A. lekithochrous needs sea salts to grow; [d]With the exception of A. ebronensis; [e]With the exception of A. molluscorum; [f]With the exception of A. cibarius; [g]With the exception of A. anaerophilus. ND, not determined.

to strains isolated from marine environments and their phylogenetic position was close to the two marine clusters (3 and 4).

As indicated in the review by Collado and Figueras (2011), there are many uncultured or not-yet-described species of *Arcobacter*, which have been recognized on the basis of nearly full-length 16S rRNA gene sequences, and which probably outnumber those species that were already known at that time. Their hosts and/or habitats are very diverse and include cod larvae, cyanobacterial mats, activated sludge, tidal and marine sediments, estuarine and river water, plankton, coral, tubeworms, snails, etc. (Collado et al., 2011; and references therein). In the near future new species can be expected to emerge that will reinforce the value of the different genera proposed in this study.

CONCLUSION

Genomic information obtained through next-generation sequencing leads to great advances in the systematics of prokaryotes (Whitman, 2015), not only to the general understanding of prokaryotic biology but also for the resolution of the phylogeny of taxa higher than species. Single gene phylogeny, including 16S rRNA gene, has often limitations that analysis of complete genome sequences can overcome. The study aims to use this modern taxonomy approach to clarify the relationships of the diverse *Arcobacter* species.

The results obtained in the present study confirmed the opinion of some authors on the need for a clarification of the taxonomy of the genus *Arcobacter*. The phylogenetic analyses derived from the MLSA of 13 genes and of the core genome as well as the existence of signature regions in the 16S rRNA gene have shown, together with the genomic indexes ANI (75.2–81.8%), *is*DDH (19.5–24.8%), AAI (67.6–80.3%), and POCP (67.0–75.4%), to be useful tools for delimiting several genomic and phylogenetic groups within this genus. The intra-genus ranges and cut-off values established here might also be helpful for future taxonomic studies in other bacterial groups.

Such genomic variability, together with the determination of combinations of differentiating phenotypic traits allowed the division of the current genus *Arcobacter* in at least six different genera for which the names *Aliiarcobacter* gen. nov., *Pseudoarcobacter* gen. nov., *Haloarcobacter* gen. nov., *Malacobacter* gen. nov., and *Poseidonibacter* gen. nov. are proposed. In addition, the candidate species '*A. aquaticus*' also constitutes a new genus for which the name Candidate '*Arcomarinus*' gen. nov. is proposed, although such proposal should be formulated in parallel to the formal description of the species.

According to Tindall et al. (2010) "*the type strain of a genus is the most important reference organism to which a novel species has to be compared.*" In the case of the genus *Arcobacter*, the type species has rarely been isolated (Collado et al., 2009b; Toh et al., 2011; Levican et al., 2016; Salas-Massó et al., 2016) and in fact, all the analyses show that *A. nitrofigilis* is an orphan species and the only representative of the genus *Arcobacter*, for which an emended description is provided.

The other genera are described here while taking into account the species validated at the time of writing but with the confidence that the formal description of the candidate species would fit in such descriptions. Thus, the genus *Aliiarcobacter* gen. nov. is described comprising seven species *Aliiarcobacter cryaerophilus* comb. nov., *A. butzleri* comb. nov., *A. skirrowii* comb. nov., *A. cibarius* comb. nov., *A. thereius* comb. nov., *A. trophiarum* comb. nov., *A. lanthieri* comb. nov., and *A. faecis* comb. nov. On the other hand, the genus *Pseudoarcobacter* gen. nov. includes

the species *Pseudoarcobacter defluvii* comb. nov., *P. ellisii* comb. nov., *P. venerupis* comb. nov., *P. cloacae* comb. nov., *P. suis* comb. nov., *P. aquimarinus* comb. nov., and *P. acticola* comb. nov. Four species, *Malacobacter halophilus* comb. nov., *M. mytili* comb. nov., *M. marinus* comb. nov., *M. molluscorum* comb. nov., and *M. pacificus* comb. nov. are compiled in the new genus *Malacobacter* gen. nov., whereas the genus *Haloarcobacter* gen. nov. comprises three species *Haloarcobacter bivalviorum* comb. nov., *H. anaerophilus* comb. nov., and *H. ebronensis* comb. nov. Finally, the genus *Poseidonibacter* gen. nov. has a unique species *Poseidonibacter lekithochrous* comb. nov.

Emended Description of the Genus *Arcobacter* Vandamme et al., 1991 emend. Vandamme et al., 1992 and Sasi-Jyothsna et al., 2013

Arcobacter (Ar'co.bac.ter. L. n. *arcus*, bow; Gr. n. *bacter*, rod; M. L. masc. n. *Arcobacter*, bow-shaped rod).

Cells are Gram-negative, curved rods 0.2–0.9 μm in diameter and 1–3 μm long. Coccoid bodies are found in old cultures but are not rapidly produced under aerobic conditions. Motile with a rapid corkscrew motion. Each cell possesses a single polar flagellum. Does not swarm. Chemoorganotrophic. Utilizes organic and amino acids as carbon sources, but not carbohydrates. Respiratory metabolism with oxygen as the terminal electron acceptor; anaerobic growth with aspartate and fumarate, but not with nitrate. Nitrate usually reduced to nitrite. Requires NaCl for growth. Grows at temperatures of 10°C–35°C but not at 42°C. Catalase, oxidase, urease, and nitrogenase positive. Phosphatase, sulfatase and indole negative. Does not hydrolyze esculin, casein, DNA, gelatine, hippurate or starch. Fluorescent pigments are not produced. Unable to grow with glycine (1% wt/vol), safranin (0.05% wt/vol), oxgall (1% wt/vol), or 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol). Positive for the hydrolysis of indoxyl acetate. Poly-β-hydroxybutyrate not produced.

The base composition of the DNA is 28.1–28.4% G+C as determined from the genomes.

The type species is *Arcobacter nitrofigilis*.

Description of *Aliiarcobacter* gen. nov.

Aliiarcobacter (A.li.i.ar.co.bac'ter, L. pronoun *alius* other, another; N.L. masc. n. *Arcobacter* a bacterial generic name; N.L. masc. n. *Aliiarcobacter* the other *Arcobacter*).

Cells are Gram-negative, curved rods 0.2–0.5 μm in diameter and 1–3 μm long. Motile by single polar flagellum. Does not swarm. Chemoorganotrophic. Oxidase and catalase positive. No growth occur at 4% NaCl. Growth occurs at 15°C–42°C. Carbohydrates are not fermented. Nitrate usually reduced to nitrite. Positive for the hydrolysis of indoxyl acetate and negative for urease. Growth does not occur in the presence 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol) or glycine (1% wt/vol). Some species may grow in the presence of safranin (0.05% wt/vol) or oxgall (1% wt/vol). Fluorescent pigments are not produced. Some species are sensitive to cefoperazone (64 mg/l). Range of DNA G+C content is 26.4–29.4 mol%.

The type species is *Aliiarcobacter cryaerophilus*.

Description of *Aliiarcobacter cryaerophilus* comb. nov.

Basonym: *Campylobacter cryaerophila* Neill et al., 1985.

Other synonym: *Arcobacter cryaerophilus* Vandamme et al., 1991.

The description is the same given by Neill et al. (1985). The type strain is A169/B^T (= NCTC 1185^T = ATCC 43158^T).

Description of *Aliiarcobacter butzleri* comb. nov.

Basonym: *Campylobacter butzleri* Kiehlbauch et al., 1991.

Other synonym: *Arcobacter butzleri* Vandamme et al., 1992.

The description is the same given by Vandamme et al. (1992). The type strain is LMG 10828^T (= CDC D2686^T = ATCC 49616^T).

Description of *Aliiarcobacter skirrowii* comb. nov.

Basonym: *Arcobacter skirrowii* Vandamme et al., 1992.

The description is the same given by Vandamme et al. (1992). The type strain is Skirrow 449/80^T (= LMG 6621^T = CCUG 10374^T).

Description of *Aliiarcobacter cibarius* comb. nov.

Basonym: *Arcobacter cibarius* Houf et al., 2005.

The description is the same given by Houf et al. (2005). The type strain is LMG 21996^T (= CCUG 48482^T).

Description of *Aliiarcobacter thereius* comb. nov.

Basonym: *Arcobacter thereius* Houf et al., 2009.

The description is the same given by Houf et al. (2009). The type strain is LMG 24486^T (= CCUG 56902^T).

Description of *Aliiarcobacter trophiarum* comb. nov.

Basonym: *Arcobacter trophiarum* De Smet et al., 2011.

The description is the same given by De Smet et al. (2011). The type strain is 64^T (= LMG 25534^T = CCUG 59229^T).

Description of *Aliiarcobacter lanthieri* comb. nov.

Basonym: *Arcobacter lanthieri* Whiteduck-Léveillée et al., 2015.

The description is the same given by Whiteduck-Léveillée et al. (2015). The type strain is AF1440^T (= LMG 28516^T = CCUG 66485^T).

Description of *Aliiarcobacter faecis* comb. nov.

Basonym: *Arcobacter faecis* Whiteduck-Léveillée et al., 2016.

The description is the same given by Whiteduck-Léveillée et al. (2016). The type strain is AF1078^T (= LMG 28519^T = CCUG 66484^T).

Description of *Pseudoarcobacter* gen. nov.

Pseudoarcobacter (Pseu.do.ar.co.bac'ter, Gr. adj. *pseudes*, false; N.L. masc. n. *Arcobacter* a bacterial generic name; N.L. masc. n. *Pseudoarcobacter*, false *Arcobacter*).

Gram-negative, cells are rod shaped and motile. Cell size 0.2–0.9 μm in diameter and 0.4–2.2 μm long. Some species may present cells up to 10 μm in length. Oxidase and catalase positive. No growth occurs at 4% NaCl. Growth occurs at 15–37°C, but not at 42°C. Carbohydrates are not fermented. Reduce nitrate to nitrite. Positive for the hydrolysis of indoxyl acetate. Some species may hydrolyze urea. Growth does not occur in the presence 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol) or glycine (1% wt/vol). Some species may grow in the presence of safranin (0.05% wt/vol) or oxgall (1% wt/vol). Sensitive to cefoperazone (64 mg/l). Range of DNA G+C content is 26.3–28.0 mol%.

The type species is *Pseudoarcobacter defluvii.*

Description of *Pseudoarcobacter defluvii* comb. nov.

Basonym: *Arcobacter defluvii* Collado et al., 2011.

The description is the same given by Collado et al. (2011). The type strain is SW28-11^T (= CECT 7697^T = LMG 25694^T).

Description of *Pseudoarcobacter ellisii* comb. nov.

Basonym: *Arcobacter ellisii* Figueras et al., 2011b.

The description is the same given by Figueras et al. (2011b). The type strain is F79-6^T (= CECT 7837^T = LMG 26155^T).

Description of *Pseudoarcobacter venerupis* comb. nov.

Basonym: *Arcobacter venerupis* Levican et al., 2012.

The description is the same given by Levican et al. (2012). The type strain is F67-11^T (= CECT 7836^T = LMG 26156^T).

Description of *Pseudoarcobacter cloacae* comb. nov.

Basonym: *Arcobacter cloacae* Levican et al., 2013.

The description is the same given by Levican et al. (2013). The type strain is SW28-13^T (= CECT 7834^T = LMG 26153^T)

Description of *Pseudoarcobacter suis* comb. nov.

Basonym: *Arcobacter suis* Levican et al., 2013.

The description is the same given by Levican et al. (2013). The type strain is F41^T (= CECT 7833^T = LMG 26152^T).

Description of *Pseudoarcobacter aquimarinus* comb. nov.

Basonym: *Arcobacter aquimarinus* Levican et al., 2015.

The description is the same given by Levican et al. (2015). The type strain is W63^T (= CECT 8442^T = LMG 27923^T).

Description of *Pseudoarcobacter acticola* comb. nov.

Basonym: *Arcobacter acticola* Park et al., 2016.

The description is the same given by Park et al. (2016). The type strain is AR-13^T (= KCTC 52212^T = NBRC 112272^T).

Description of *Malacobacter* gen. nov.

Malacobacter (Ma.la.co.bac'ter; Gr. n. *malaco*, soft, with soft boy, mollusc; Gr. n. *bacter*, rod; N.L. masc. n. *Malacobacter*, bacteria isolated from molluscs).

Gram-negative, cells are rod shaped and motile. Cell size 0.1–0.6 μm wide and 0.5–3.6 μm long. Oxidase positive and catalase variable among species. Halophilic, no growth can be obtained without NaCl and capable to grow up to 4% NaCl. Growth occurs at 15°C–37°C. Does not grow at 37°C in microaerophilic conditions nor at 42°C in anaerobiosis. Carbohydrates are not fermented. Does not reduce nitrate to nitrite. Negative for the hydrolysis of urea. Some species may hydrolyze indoxyl acetate. Growth does not occur in the presence of oxgall (1% wt/vol) or 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol). Some species may grow in the presence of glycine (1% wt/vol) or safranin (0.05% wt/vol). Sensitive to cefoperazone (64 mg/l) variable among species. Range of DNA G+C content is 26.1–27.3 mol%.

The type species is *Malacobacter halophilus.*

Description of *Malacobacter halophilus* comb. nov.

Basonym: *Arcobacter halophilus* Donachie et al., 2005.

The description is the same given by Donachie et al. (2005). The type strain is LA31B^T (= ATCC BAA-1022^T = CIP 108450^T).

Description of *Malacobacter mytili* comb. nov.

Basonym: *Arcobacter mytili* Collado et al., 2009a.

The description is the same given by Collado et al. (2009a). The type strain is F2075^T (= CECT 7386^T = LMG 24559^T).

Description of *Malacobacter marinus* comb. nov.

Basonym: *Arcobacter marinus* Kim et al., 2010.

The description is the same given by Kim et al. (2010), with the exception of variable result among strains for the hydrolysis of the indoxyl-acetate under microaerobic conditions (Salas-Massó et al., 2016). The type strain is CL-S1^T (= KCCM 90072^T = JCM 15502^T).

Description of *Malacobacter canalis* comb. nov.

Basonym: *Arcobacter canalis* Pérez-Cataluña et al., 2018b.

The description is the same given by Pérez-Cataluña et al. (2018b). The type strain is F138-33^T (= CECT 8984^T = LMG 29148^T).

Description of *Malacobacter molluscorum* comb. nov.

Basonym: *Arcobacter molluscorum* Figueras et al., 2011a.

The description is the same given by Figueras et al. (2011a). The type strain is F98-3^T (= CECT 7696^T = LMG 25693^T).

Description of *Malacobacter pacificus* comb. nov.

Basonym: *Arcobacter pacificus* Zhang et al., 2015.

The description is the same given by Zhang et al. (2015). The type strain is SW028^T (= DSM 25018T = JCM 17857^T = LMG 26638^T).

Description of *Haloarcobacter* gen. nov.

Haloarcobacter (Ha.lo.ar.co.bac'ter, Gr. n. *halo*, salt; N.L. masc. n. *Arcobacter*, a bacterial generic name; N.L. masc. n. *Haloarcobacter, Arcobacter* salt loving).

Gram-negative, cells are rod shaped and motile. Cell size 0.1–0.5 μm in diameter and 0.9–2.5 μm in length. Oxidase positive and catalase variable among species. Halophilic, growth can be obtained within the range of 0.5% (variable among species) and up to 4% NaCl. Growth occurs at 15–42°C. Growth at 37°C in microaerophilic conditions or at 42°C in anaerobiosis variable among species. Carbohydrates are not fermented. Some species may reduce nitrate to nitrite. Negative for the hydrolysis of urea (with the exception of *H. ebronensis*). Some species may hydrolyze indoxyl acetate. Growth does not occur in the presence of oxgall (1% wt/vol) (with the exception of *H. molluscorum*) or 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol). No growth on CCDA. Some species may grow in the presence of glycine (1% wt/vol) or safranin (0.05% wt/vol). Sensitive to cefoperazone (64 mg/l). Range of DNA G+C content is 27.3–29.9 mol%.

The type species is *Haloarcobacter bivalviorum*.

Description of *Haloarcobacter bivalviorum* comb. nov.

Basonym: *Arcobacter bivalviorum* Levican et al., 2012.

The description is the same given by Levican et al. (2012). The type strain is F4^T (= CECT 7835^T = LMG 26154^T).

Description of *Haloarcobacter anaerophilus* comb. nov.

Basonym: *Arcobacter anaerophilus* Sasi-Jyothsna et al., 2013.

The description is the same given by Sasi-Jyothsna et al. (2013). The type strain is JC84^T (= KCTC 15071^T = MTCC 10956^T = DSM 24636^T).

Description of *Haloarcobacter ebronensis* comb. nov.

Basonym: *Arcobacter ebronensis* Levican et al., 2015.

The description is the same given by Levican et al. (2015). The type strain is F128-2^T (= CECT 8441^T = LMG 27922^T).

Description of *Poseidonibacter* gen. nov.

Poseidonibacter (Po.se.i.do.ni.bac'ter, Gr. n. *Poseidon*, God of the sea; Gr. n. *bacter*, rod; N.L. masc. n. *Poseidonibacter* referring to the marine habitat of this bacteria).

Gram-negative, cells are rod shaped and motile. Oxidase and catalase positive. Halophilic, no growth can be obtained without seawater or the addition of combined marine salts to the medium. Growth occurs at 15°C–25°C, but not at 37°C or 42°C. Range of pH for growth is 6–8. Carbohydrates are not fermented. Reduce nitrate to nitrite. Negative for the hydrolysis of indoxyl acetate and urea. Growth occurs in the presence of safranin (0.05% wt/vol), and 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol), but not in the presence of glycine (1% wt/vol) sensitive to cefoperazone (30 μg). Possess ubiquinone MK-6 as a respiratory quinone. DNA G+C content is 28.7 mol%.

The type species is *Poseidonibacter lekithochrous*.

Description of *Poseidonibacter lekithochrous* comb. nov.

Basonym: *Arcobacter lekithochrous* Diéguez et al., 2017.

The description is the same given by Diéguez et al. (2017). The type strain is LFT1.7^T (= CECT 8942^T = DSM 100870^T).

AUTHOR CONTRIBUTIONS

MF and JR designed the work. AP-C, NS-M, and AD performed the phenotypic and phylogenetic experiments. AP-C and SB carried out the genome sequencing and analysis. AP-C, AL, and JR performed the bioinformatic work. JR, MF, AP-C, and AD wrote the paper.

FUNDING

This work was supported in part by Grants JPIW2013-69095-C03-03 from the Ministerio de Economía y Competitividad (MINECO), AQUAVALENS of the Seventh Framework Program (FP7/2007-2013) grant agreement 311846 from the European Union and AGL2013-42628-R and AGL2016-77539-R (AEI/FEDER UE) from the Agencia Estatal de Investigación (Spain).

ACKNOWLEDGMENTS

The authors thank Dr. F. J. García (Laboratorio Central de Veterinaria de Algete, MAGRAMA, Madrid, Spain) and Drs. R. Alonso, I. Martinez-Malaxetxebarria, and A. Fernandez-Astorga [Faculty of Pharmacy, University of the Basque Country (UPV-EHU), Vitoria-Gasteiz, Spain], for kindly providing some of the

Arcobacter strains. AP-C thanks Institut d'Investigació Sanitària Pere Virgili (IISPV) for her Ph.D. fellowship and NS-M thanks the Universitat Rovira i Virgili (URV), the Institut de Recerca i Tecnologia Agroalimentària (IRTA) and the Banco Santander for her Ph.D. fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02077/full#supplementary-material

REFERENCES

Adilakshmi, T., Bellur, D. L., and Woodson, S. A. (2008). Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature* 455, 1268–1272. doi: 10.1038/nature07298

Bandelt, H. J., and Dress, A. W. M. (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol. Phylogenet. Evol.* 1, 242–252. doi: 10.1016/1055-7903(92)90021-8

Botzman, M., and Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol.* 12:R109. doi: 10.1186/gb-2011-12-10-r109

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijs.0.054171-0

Collado, L., Cleenwerck, I., Van Trappen, S., De Vos, P., and Figueras, M. J. (2009a). *Arcobacter mytili* sp. nov., an indoxyl acetate –hydrolysis- negative bacterium isolated from mussels. *Int. J. Syst. Evol. Microbiol.* 59, 1391–1396. doi: 10.1099/ijs.0.003749-0

Collado, L., Guarro, J., and Figueras, M. J. (2009b). Prevalence of *Arcobacter* in meat and shellfish. *J. Food Prot.* 72, 1102–1106.

Collado, L., and Figueras, M. J. (2011). Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clin. Microbiol. Rev.* 24, 174–192. doi: 10.1128/CMR.00034-10

Collado, L., Inza, I., Guarro, J., and Figueras, M. J. (2008). Presence of *Arcobacter* spp. in environmental waters correlates with high levels of fecal pollution. *Environ. Microbiol.* 10, 1635–1640. doi: 10.1111/j.1462-2920.2007.01555.x

Collado, L., Kasimir, G., Perez, U., Bosch, A., Pinto, R., Saucedo, G., et al. (2010). Occurrence and diversity of *Arcobacter* spp. along the Llobregat River catchment, at sewage effluents and in a drinking water treatment plant. *Water Res.* 44, 3696–3702. doi: 10.1016/j.watres.2010.04.002

Collado, L., Levican, A., Pérez, J., and Figueras, M. J. (2011). *Arcobacter defluvii* sp. nov., isolated from sewage samples. *Int. J. Syst. Evol. Microbiol.* 61, 2155–2161. doi: 10.1099/ijs.0.025668-0

De Smet, S., Vandamme, P., De Zutter, L., On, S. L. W., Douidah, L., and Houf, K. (2011). *Arcobacter trophiarum* sp. nov., isolated from fattening pigs. *Int. J. Syst. Evol. Microbiol.* 61, 356–361. doi: 10.1099/ijs.0.022665-0

Diéguez, A. L., Balboa, S., Magnesen, T., and Romalde, J. L. (2017). *Arcobacter lekithochrous* sp. nov., isolated from a molluscan hatchery. *Int. J. Syst. Evol. Microbiol.* 67, 1327–1332. doi: 10.1099/ijsem.0.001809

Diéguez, A. L., Pérez-Cataluña, A., Figueras, M. J., and Romalde, J. L. (2018). *Arcobacter haliotis* Tanaka et al., 2017 is a later heterotypic synonym of *Arcobacter lekithochrous* Diéguez et al., 2017. *Int. J. Syst. Evol. Microbiol.* doi: 10.1099/ijsem.0.002909 [Epub ahead of print].

Donachie, S. P., Bowman, J. P., On, S. L. W., and Alam, M. (2005). *Arcobacter halophilus* sp. nov., the first obligate halophile in the genus *Arcobacter*. *Int. J. Syst. Evol. Microbiol.* 55, 1271–1277. doi: 10.1099/ijs.0.63581-0

Farooqi, M. S., Mishra, D., Rai, N., Singh, D., Rai, A., Chaturvedi, K., et al. (2016). Genome-wide relative analysis of codon usage bias and codon context pattern in the bacteria *Salinibacter ruber*, *Chromohalobacter salexigens* and *Rhizobium etli*. *Biochem. Anal. Biochem.* 5:257. doi: 10.4172/2161-1009.1000257

Figueras, M. J., Collado, L., and Guarro, J. (2008). A new 16S rDNA-RFLP method for the discrimination of the accepted species of *Arcobacter*. *Diagn. Microbiol. Infect. Dis.* 62, 11–15. doi: 10.1016/j.diagmicrobio.2007.09.019

Figueras, M. J., Collado, L., Levican, A., Pérez, J., Solsona, M. J., and Yustes, C. (2011a). *Arcobacter molluscorum* sp. nov., a new species isolated from shellfish. *Syst. Appl. Microbiol.* 34, 105–109. doi: 10.1016/j.syapm.2010.10.001

Figueras, M. J., Levican, A., Collado, L., Inza, M. I., and Yustes, C. (2011b). *Arcobacter ellisii* sp. nov., isolated from mussels. *Syst. Appl. Microbiol.* 34, 414–418. doi: 10.1016/j.syapm.2011.04.004

Figueras, M. J., Levican, A., Pujol, I., Ballester, F., Rabada Quilez, M. J., and Gomez-Bertomeu, F. (2014). A severe case of persistent diarrhoea associated with *Arcobacter cryaerophilus* but attributed to *Campylobacter* sp. and a review of the clinical incidence of *Arcobacter* spp. *New Microbes New Infect.* 2, 31–37. doi: 10.1002/2052-2975.35

Figueras, M. J., Pérez-Cataluña, A., Salas-Massó, N., Levican, A., and Collado, L. (2017). '*Arcobacter porcinus*' sp. nov., a novel *Arcobacter* species uncovered by *Arcobacter thereius*. *New Microbe New Infect.* 15, 104–106. doi: 10.1016/j.nmni.2016.11.014

Fisher, J. C., Levican, A., Figueras, M. J., and McLellan, S. L. (2014). Population dynamics and ecology of *Arcobacter* in sewage. *Front. Microbiol.* 5:525. doi: 10.3389/fmicb.2014.00525

Giacometti, F., Salas-Massó, N., Serraino, A., and Figueras, M. J. (2015). Characterization of *Arcobacter* suis isolated from water buffalo (*Bubalus bubalis*) milk. *Food Microbiol.* 51, 186–191. doi: 10.1016/j.fm.2015.06.004

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483r-0

Horikoshi, M., and Tang, Y. (2015). *ggfortify: Data Visualization Tools for Statistical Analysis Results*. Available at: https://CRAN.R-project.org/package=ggfortify

Houf, K., On, S. L. W., Coenye, T., Debruyne, L., De Smet, S., and Vandamme, P. (2009). *Arcobacter thereius* sp. nov., isolated from pigs and ducks. *Int. J. Syst. Evol. Microbiol.* 59, 2599–2604. doi: 10.1099/ijs.0.006650-0

Houf, K., On, S. L. W., Coenye, T., Mast, J., Van Hoof, J., and Vandamme, P. (2005). *Arcobacter cibarius* sp. nov., isolated from boiled carcasses. *Int. J. Syst. Evol. Microbiol.* 55, 713–717. doi: 10.1099/ijs.0.63103-0

Hsu, T. T. D., and Lee, J. (2015). Global distribution and prevalence of *Arcobacter* in food and water. *Zoonoses Public Health* 62, 579–589. doi: 10.1111/zph.12215

Huson, D. H., and Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

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

Kiehlbauch, A., Brenner, D. J., Nicholson, M. A., Baker, C. N., Patton, C. M., Steigerwalt, A. G., et al. (1991). *Campylobacter butzleri* sp. nov., isolated from humans and diarrheal illness. *J. Clin. Microbiol.* 29, 376–385.

Kim, H. M., Hwang, C. Y., and Cho, B. C. (2010). *Arcobacter marinus* sp. nov. *Int. J. Syst. Evol. Microbiol.* 60, 531–536. doi: 10.1099/ijs.0.007740-0

Kitahara, K., Yasutake, Y., and Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *PNAS* 109, 19220–19225. doi: 10.1073/pnas.1213609109

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, 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

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Levican, A., Collado, L., Aguilar, C., Yustes, C., Diéguez, A. L., Romalde, J. L., et al. (2012). *Arcobacter bivalviorum* sp. nov. and Arcobacter venerupis sp. nov., new species isolated from shellfish. *Syst. Appl. Microbiol.* 35, 133–138. doi: 10.1016/j.syapm.2012.01.002

Levican, A., Collado, L., and Figueras, M. J. (2013). *Arcobacter cloacae* sp. nov. and *Arcobacter suis* sp. nov., two new species isolated from food and sewage. *Syst. Appl. Microbiol.* 36, 22–27. doi: 10.1016/j.syapm.2012.11.003

Levican, A., Collado, L., and Figueras, M. J. (2016). The use of two culturing methods in parallel reveals a high prevalence and diversity of *Arcobacter* spp. in a wastewater treatment plant. *Biomed Res. Int.* 2016:8132058. doi: 10.1155/2016/8132058

Levican, A., Collado, L., Yustes, C., Aguilar, C., and Figueras, M. J. (2014). Higher water temperature and incubation under aerobic and microaerobic conditions increase the recovery and diversity of *Arcobacter* spp. from shellfish. *Appl. Environ. Microbiol.* 80, 385–391. doi: 10.1128/AEM.03014-13

Levican, A., Rubio-Arcos, S., Martínez-Murcia, A., Collado, L., and Figueras, M. J. (2015). *Arcobacter ebronensis* sp. nov. and *Arcobacter aquimarinus* sp. nov., two new species isolated from marine environment. *Syst. Appl. Microbiol.* 38, 30–35. doi: 10.1016/j.syapm.2014.10.011

Luo, C., Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res.* 42:e73. doi: 10.1093/nar/gku169

Ma, X. X., Feng, Y. P., Bai, J. L., Zhang, D. R., Lin, X. S., and Ma, Z. R. (2015). Nucleotide composition bias and codon usage trends of gene populations in *Mycoplasma capriolum* subsp. *capriolum* and *M. agalactiae*. *J. Genetics* 94, 251–260. doi: 10.1007/s12041-015-0512-2

Marenda, M. S., Sagne, E., Poumarat, F., and Citti, C. (2005). Suppression subtractive hybridization as a basis to assess *Mycoplasma agalactiae* and *Mycoplasma bovis* genomic diversity and species-specific sequences. *Microbiology* 151, 475–489. doi: 10.1099/mic.0.27590-0

Martínez-Murcia, A., and Lamy, B. (2015). "Molecular diagnostics by genetic methods," in *Aeromonas*, ed. J. Graf (Poole: Caister Academic Press), 155–200.

Martínez-Murcia, A. J., Benlloch, S., and Collins, M. D. (1992). Phylogenetic interrelationships of members of the genera Aeromonas and Plesiomonas as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. *Int. J. Syst. Bacteriol.* 42, 412–421. doi: 10.1099/00207713-42-3-412

Martínez-Murcia, A. J., Figueras, M. J., Saavedra, M. J., and Stackebrandt, E. (2007). The recently proposed species *Aeromonas sharmana* sp. nov., isolate GPTSA-6T, is not a member of the genus Aeromonas. *Int. Microbiol.* 10, 61–64.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Neill, S. D., Campbell, J. N., O'Brien, J. J., Weatherup, S. T. C., and Ellis, W. A. (1985). Taxonomic position of *Campylobacter cryaerophila* sp. *nov*. *Int. J. Syst. Bacteriol.* 35, 342–356. doi: 10.1099/00207713-35-3-342

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20, 714–737. doi: 10.1089/cmb.2013.0084

On, S. L. W., Miller, W. G., Houf, K., Fox, J. G., and Vandamme, P. (2017). Minimal standards for describing new species belonging to the families Campylobacteraceae and Helicobacteraceae: *campylobacter, Arcobacter, Helicobacter* and *Wolinella* spp. *Int. J. Syst. Evol. Microbiol.* 67, 5296–5311. doi: 10.1099/ijsem.0.002255

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

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Park, S., Jung, Y. T., Kim, S., and Yoon, J. H. (2016). *Arcobacter acticola* sp. nov., isolated from seawater of the East Sea in South Korea. *J. Microbiol.* 54, 655–659. doi: 10.1007/s12275-016-6268-4

Pérez-Cataluña, A., Collado, L., Salgado, O., Lefiñanco, V., and Figueras, M. J. (2018a). A polyphasic and taxogenomic evaluation uncovers *Arcobacter cryaerophilus* as a species complex that embraces four genomovars. *Front. Microbiol* 9:805. doi: 10.3389/fmicb.2018.00805

Pérez-Cataluña, A., Salas-Massó, N., and Figueras, M. J. (2018b). *Arcobacter canalis* sp. nov., isolated from a water canal contaminated with urban sewage. *Int. J. Syst. Evol. Microbiol.* 68, 1258–1264. doi: 10.1099/ijsem.0.002662

Puigbò, P., Bravo, I. G., and Garcia-Vallve, S. (2008). CAIcal: a combined set of tools to assess codon usage adaptation. *Biol. Dir.* 3:38. doi: 10.1186/1745-6150-3-38

Qin, Q.-L., Xie, B. B., Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196, 2210–2215. doi: 10.1128/JB.01688-14

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Řeháková, K., Johansen, J. R., Bowen, M. B., Martin, M. P., and Sheil, C. A. (2014). Variation in secondary structure of the 16S rRNA molecule in Cyanobacteria with implications for phylogenetic analysis. *Fottea* 14, 161–178. doi: 10.5507/fot.2014.013

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67. doi: 10.1111/j.1574-6976.2001.tb00571.x

Salas-Massó, N., Andree, K. B., Furones, M. D., and Figueras, M. J. (2016). Enhanced recovery of *Arcobacter* spp. using NaCl in culture media and re-assessment of the traits of *Arcobacter marinus and Arcobacter halophilus* isolated from marine water and shellfish. *Sci. Total Environ.* 56, 1355–1361. doi: 10.1016/j.scitotenv.2016.05.197

Salas-Massó, N., Pérez-Cataluña, A., Collado, L., Levican, A., and Figueras, M. J. (2018). "Arcobacter", in *Handbook of Foodborne Diseases*, ed. D. Liu (Boca Raton, FL: CRC Press, Taylor & Francis Group).

Sasi-Jyothsna, T. S., Rahul, K., Ramaprasad, E. V. V., Sasikala, C., and Ramana, C. V. (2013). *Arcobacter anaerophilus* sp. nov., isolated from an estuarine sediment and emended description of the genus Arcobacter. *Int. J. Syst. Evol. Microbiol.* 63, 4619–4625. doi: 10.1099/ijs.0.054155-0

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

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sharp, P. M., and Li, W. H. (1987). The codon adaptation index – a mesure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acid Res.* 15, 1281–1295. doi: 10.1093/nar/15.3.1281

Suzuki, H., Brown, C. J., Fornay, L. J., and Top, E. M. (2008). Comparison of correspondence analysis methods for synonymous codon usage in bacteria. *DNA Res.* 15, 357–365. doi: 10.1093/dnares/dsn028

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

Tanaka, R., Cleenwerck, I., Mizutani, Y., Iehata, S., Bossier, P., and Vandamme, P. (2017). *Arcobacter haliotis* sp. nov., isolated from abalone species *Haliotis gigantea*. *Int. J. Syst. Evol. Microbiol* 67, 3050–3056. doi: 10.1099/ijsem.0.002080

Tang, Y., Horikoshi, M., and Li, W. (2016). Ggfortify: unified interface to visualize statistical result of popular R Packages. *The R J.* 8, 474–485.

Tindall, B. J., Rosselló-Mora, R., Busse, H. J., Ludwig, W., and Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. doi: 10.1099/ijs.0.016949-0

Toh, H., Sharma, V. K., Oshima, K., Kondo, S., Hattori, M., Ward, F. B., et al. (2011). Complete genome sequences of *Arcobacter butzleri* ED-1 and *Arcobacter* sp. strain L, both isolated from a microbial fuel cell. *J. Bacteriol.* 193, 6411–6412. doi: 10.1128/JB.06084-11

Ue, H., Matsuo, Y., Kasai, H., and Yokota, A. (2011). Demequina globuliformis sp. nov., *Demequina oxidasica* sp. nov. and *Demequina aurantiaca* sp. nov., actinobacteria isolated from marine environments, and proposal of

Demequinaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* 61, 1322–1329. doi: 10.1099/ijs.0.024299-0

Ursing, J. B., Lior, H., and Owen, R. J. (1994). Proposal of minimal standards for describing new species of the family Campylobacteraceae. *Int. J. Syst. Bacteriol.* 44, 842–845. doi: 10.1099/00207713-44-4-842

Van den Abeele, A. M., Vogelaers, D., Van Hende, J., and Houf, K. (2014). Prevalence of *Arcobacter* species among humans, Belgium, 2008-2013. *Emerg. Infect. Dis.* 20, 1731–1734. doi: 10.3201/eid2010.140433

Vandamme, P., Falsen, E., Rossau, R., Hoste, B., Segers, P., Tytgat, R., et al. (1991). Revision of Campylobacter, *Helicobacter* and *Wolinella* taxonomy: emendation of generic descriptions and proposal of *Arcobacter* gen. *nov*. *Int. J. Syst. Bacteriol.* 41, 88–103. doi: 10.1099/00207713-41-1-88

Vandamme, P., Vancanneyt, M., Pot, B., Mels, L., Hoste, B., Dewettinck, D., et al. (1992). Polyphasic taxonomic study of the emended genus *Arcobacter* with *Arcobacter butzleri* comb. nov. and *Arcobacter skirrowii* sp. nov., an aerotolerant bacterium isolated from veterinary specimens. *Int. J. Syst. Bacteriol.* 42, 344–356. doi: 10.1099/00207713-42-3-344

Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., et al. (2017). Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to Epsilonbacteraeota (phyl. *nov.)*. *Front. Microbiol.* 8:682. doi: 10.3389/fmicb.2017.00682

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* 45, D535–D542. doi: 10.1093/nar/gkw1017

Weimann, A., Mooren, K., Frank, J., Pope, P. B., Bremges, A., and McHardy, A. C. (2016). From genomes to phenotypes: traitar, the microbial trait analyzer. *mSystems* 1, e101–e116. doi: 10.1128/mSystems.00101-16

Weiner, J. (2017). *Pca3d: Three Dimensional PCA Plots*. Available at: https://CRAN.R-project.org/package=pca3d.

Whiteduck-Léveillée, K., Whiteduck-Léveillée, J., Cloutier, M., Tambong, J. T., Xu, R., Topp, E., et al. (2015). *Arcobacter lanthieri* sp. nov., isolated from pig and dairy cattle manure. *Int. J. Syst. Evol. Microbiol.* 65, 2709–2716. doi: 10.1099/ijs.0.000318

Whiteduck-Léveillée, K., Whiteduck-Léveillée, J., Cloutier, M., Tambong, J. T., Xu, R., Topp, E., et al. (2016). Identification, characterization and description of *Arcobacter* faecis sp. nov., isolated from a human waste septic tank. *Syst. Appl. Microbiol.* 39, 93–99. doi: 10.1016/j.syapm.2015.12.002

Whitman, W. B. (2015). Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Syst. Appl. Microbiol.* 38, 217–222. doi: 10.1016/j.syapm.2015.02.003

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-0-387-98141-3

Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K. H., et al. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol.* 31, 241–250. doi: 10.1016/j.syapm.2008.07.001

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330

Zhang, Z., Yu, C., Wang, X., Yu, S., and Zhang, X. H. (2015). *Arcobacter pacificus* sp. nov., isolated from seawater of the South Pacific Gyre. *Int. J. Syst. Evol. Microbiol.* 66, 542–547. doi: 10.1099/ijsem.0.000751

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 Pérez-Cataluña, Salas-Massó, Diéguez, Balboa, Lema, Romalde and Figueras. 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

CORRECTION
published: 21 December 2018
doi: 10.3389/fmicb.2018.03123

Corrigendum: Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Caos

Alba Pérez-Cataluña[1], Nuria Salas-Massó[1], Ana L. Diéguez[2], Sabela Balboa[2], Alberto Lema[2], Jesús L. Romalde[2] and Maria José Figueras[1]**

[1] Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina, Institut d'Investigació Sanitària Pere Virgili, Universitat Rovira i Virgili, Reus, Spain, [2] Departamento de Microbiología y Parasitología, CIBUS-Facultad de Biología, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Keywords: ***Arcobacter***, ***Aliarcobacter*** **gen. nov.,** ***Pseudoarcobacter*** **gen. nov.,** ***Haloarcobacter*** **gen. nov.,** ***Malacobacter*** **gen. nov.,** ***Poseidonibacter*** **gen. nov., taxonomic criteria**

OPEN ACCESS

Edited and reviewed by:
Yasir Muhammad,
King Abdulaziz University, Saudi Arabia

***Correspondence:**
Jesús L. Romalde
jesus.romalde@usc.es
Maria José Figueras
mariajose.figueras@urv.cat

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

Received: *02 November 2018*
Accepted: *03 December 2018*
Published: *21 December 2018*

Citation:
Pérez-Cataluña A, Salas-Massó N, Diéguez AL, Balboa S, Lema A, Romalde JL and Figueras MJ (2018) Corrigendum: Revisiting the Taxonomy of the Genus Arcobacter: Getting Order From the Caos. Front. Microbiol. 9:3123. doi: 10.3389/fmicb.2018.03123

A Corrigendum on

Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos.
by Pérez-Cataluña, A., Salas-Massó, N., Diéguez, A. L., Balboa, S., Lema, A., Romalde, J. L., et al. (2018). Front. Microbiol. 9:2077. doi: 10.3389/fmicb.2018.02077

In the original article, there was an error. The NCTC number for the strain type *Arcobacter cryaerophilus*, was incorrect. A correction has been made to the **Conclusion, Description of *Aliiarcobacter cryaerophilus* comb. nov.:**

The description is the same given by Neill et al. (1985). The type strain is A169/B^T (= NCTC 11885^T = ATCC 43158^T).

There were also typographical errors of some Basonym descriptions.

A correction has been made to the Conclusion, Description of *Pseudoarcobacter ellisii* comb. nov., *Malacobacter mytili* comb. nov., *Malacobacter canalis* comb. nov., *Malacobacter molluscorum* comb. nov and Malacobacter pacificus comb. nov.:

Basonym: *Arcobacter ellisii* Figueras et al., 2011b
Basonym: *Arcobacter mytili* Collado et al., 2009
Basonym: *Arcobacter canalis* Pérez-Cataluña et al., 2018
Basonym: *Arcobacter molluscorum* Figueras et al., 2011a
and, Basonym: *Arcobacter pacificus* Zhang et al., 2016.

After indication of the nomenclature editors for validation of new bacterial names, the term *Aliiarcobacter* has been corrected to *Aliarcobacter* throughout the article.

In the original article, the reference for Zhang et al., 2015 was incorrectly written as Zhang, Z., Yu, C., Wang, X., Yu, S., and Zhang, X. H. (2015), It should be Zhang, Z., Yu, C., Wang, X., Yu, S., and Zhang, X. H. (2016).

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Collado, L., Cleenwerck, I., Van Trappen, S., De Vos, P., and Figueras, M. J. (2009). Arcobacter mytili sp. nov., an indoxyl acetate -hydrolysis- negative bacterium isolated from mussels. *Int. J. Syst. Evol. Microbiol.* 59, 1391–1396. doi: 10.1099/ijs.0.003749-0

Figueras, M. J., Collado, L., Levican, A., Pérez, J., Solsona, M. J., and Yustes, C. (2011a). Arcobacter molluscorum sp. nov., a new species isolated from shellfish. *Syst. Appl. Microbiol.* 34, 105–109. doi: 10.1016/j.syapm.2010.10.001

Figueras, M. J., Levican, A., Collado, L., Inza, M. I., and Yustes, C. (2011b). Arcobacter ellisii sp. nov., isolated from mussels. *Syst. Appl. Microbiol.* 34, 414–418. doi: 10.1016/j.syapm.2011.04.004

Neill, S. D., Campbell, J. N., ÓBrien, J. J., Weatherup, S. T. C., and Ellis, W. A. (1985). Taxonomic position of Campylobacter cryaerophila sp. nov. *Int. J. Syst. Bacteriol.* 35, 342–356. doi: 10.1099/00207713-35-3-342

Pérez-Cataluña, A., Salas-Massó, N., and Figueras, M. J. (2018). Arcobacter canalis sp. nov., isolated from a water canal contaminated with urban sewage. *Int. J. Syst. Evol. Microbiol.* 68, 1258–1264. doi: 10.1099/ijsem.0.002662

Zhang, Z., Yu, C., Wang, X., Yu, S., and Zhang, X. H. (2016). *Arcobacter pacificus* sp. *nov.,* isolated from seawater of the South Pacific Gyre. *Int. J. Syst. Evol. Microbiol.* 66, 542–547. doi: 10.1099/ijsem.0.000751

Copyright © 2018 Pérez-Cataluña, Salas-Massó, Diéguez, Balboa, Lema, Romalde and Figueras. 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

CORRECTION
published: 01 October 2019
doi: 10.3389/fmicb.2019.02253

Corrigendum (2): Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos

Alba Pérez-Cataluña[1], Nuria Salas-Massó[1], Ana L. Diéguez[2], Sabela Balboa[2], Alberto Lema[2], Jesús L. Romalde[2] and María J. Figueras[1]**

[1] Departament de Ciències Mèdiques Bàsiques, Facultat de Medicina, Institut d'Investigació Sanitària Pere Virgili, Universitat Rovira i Virgili, Reus, Spain, [2] Departamento de Microbiología y Parasitología, CIBUS-Facultad de Biología, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Keywords: ***Arcobacter***, ***Aliiarcobacter*** **gen. nov.,** ***Pseudoarcobacter*** **gen. nov.,** ***Haloarcobacter*** **gen. nov.,** ***Malacobacter*** **gen. nov.,** ***Poseidonibacter*** **gen. nov., taxonomic criteria**

OPEN ACCESS

Edited by:
Martin G. Klotz,
Washington State University, United States

Reviewed by:
David W. Ussery,
University of Arkansas for Medical Sciences, United States
Iain Sutcliffe,
Northumbria University, United Kingdom
George Garrity,
Michigan State University, United States

****Correspondence:***
Jesús L. Romalde
jesus.romalde@usc.es
María J. Figueras
mariajose.figueras@urv.cat

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

Received: *01 August 2019*
Accepted: *17 September 2019*
Published: *01 October 2019*

Citation:
Pérez-Cataluña A, Salas-Massó N, Diéguez AL, Balboa S, Lema A, Romalde JL and Figueras MJ (2019) Corrigendum (2): Revisiting the Taxonomy of the Genus Arcobacter: Getting Order From the Chaos. Front. Microbiol. 10:2253. doi: 10.3389/fmicb.2019.02253

A Corrigendum on

Revisiting the Taxonomy of the Genus *Arcobacter*: Getting Order From the Chaos
by Pérez-Cataluña, A., Salas-Massó, N., Diéguez, A. L., Balboa, S., Lema, A., Romalde, J. L., et al. (2018) Front. Microbiol. 9:2077. doi: 10.3389/fmicb.2018.02077

The original paper, which contained a description of five new genera with 25 species (all new combinations) contained many errors that prevented the proposed names from being included in a Validation List in the *International Journal of Systematic and Evolutionary Microbiology*. The List Editors could correct part of the problems, so that the generic names *Pseudarcobacter*, *Malaciobacter*. *Halarcobacter*, and *Poseidonibacter* and the species assigned as "comb. nov." to these genera could be validly published [*Oren, A. and Garrity, G. M. (2019). List of new names and new combinations previously effectively, but not validly, published. Validation List no. 185. Int. J. Syst. Evol. Microbiol. 69, 5–9*]. The corrections made are explained in footnotes to the list. However, because of the nature of some of the changes required, the List Editors could not make the corrections for the proposed genus "Aliiarcobacter" and the eight proposed new combinations in the Validation List. The corrigendum published [**Revisiting the Taxonomy of the Genus Arcobacter: Getting Order From the Caos (sic),** *by Pérez-Cataluña, A., Salas-Massó, N., Diéguez, A. L., Balboa, S., Lema, A., Romalde, J. L., et al. (2018). Front. Microbiol. 9:3123. doi: 10.3389/fmicb.2018.03123*] failed to correct the remaining errors and introduced new problems. This necessitated a new Corrigendum in order to effectively publish the names *Aliarcobacter* and eight "comb. nov." species to be submitted subsequently for List validation in the *International Journal of Systematic and Evolutionary Microbiology*.

Description of *Aliarcobacter* gen. nov.
Aliarcobacter (A.li.ar.co.bac'ter. L. pronoun *alius* other, another; N.L. masc. n. *Arcobacter* a bacterial generic name; N.L. masc. n. *Aliarcobacter* the other *Arcobacter*). Cells are Gram-negative, curved rods 0.2–0.5 mm in diameter and 1–3 mm long. Motile by single polar flagellum. Does not swarm. Chemoorganotrophic. Oxidase and catalase positive. No growth occur at 4% NaCl. Growth occurs at 15°C−42°C. Carbohydrates are not fermented. Nitrate usually reduced to nitrite. Positive for the hydrolysis of indoxyl acetate and negative for urease. Growth does not occur in the presence 2,3,5-triphenyltetrazolium chloride (0.04%, wt/vol) or glycine (1% wt/vol). Some species may grow in the presence of safranin (0.05% wt/vol) or oxgall (1% wt/vol). Fluorescent pigments are not produced. Some species are sensitive to cefoperazone (64 mg/l). Range of DNA GC+C content is 26.4–29.4 mol%. The type species is *Aliarcobacter cryaerophilus*.

Description of *Aliarcobacter cryaerophilus* comb. nov.
Basonym: *Campylobacter cryaerophila* Neill et al. 1985.
The description is the same given by Neill et al. (1985). The type strain is A169/B^T (= NCTC 11885^T = ATCC 43158^T).

Description of *Aliarcobacter butzleri* comb. nov.
Basonym: *Campylobacter butzleri* Kiehlbauch et al. 1991.
The description is the same given by Vandamme et al. (1992). The type strain is LMG 10828^T (= CDC D2686^T = ATCC 49616^T).

Description of *Aliarcobacter skirrowii* comb. nov.
Basonym: *Arcobacter skirrowii* Vandamme et al. 1992.
The description is the same given by Vandamme et al. (1992). The type strain is Skirrow 449/80^T (= LMG 6621^T = CCUG 10374^T).

Description of *Aliarcobacter cibarius* comb. nov.
Basonym: *Arcobacter cibarius* Houf et al. 2005.
The description is the same given by Houf et al. (2005). The type strain is LMG 21996^T (= CCUG 48482^T).

Description of *Aliarcobacter thereius* comb. nov.
Basonym: *Arcobacter thereius* Houf et al. 2009.
The description is the same given by Houf et al. (2009). The type strain is LMG 24486^T (= CCUG 56902^T).

Description of *Aliarcobacter trophiarum* comb. nov.
Basonym: *Arcobacter trophiarum* De Smet et al. 2011.
The description is the same given by De Smet et al. (2011). The type strain is 64^T (= LMG 25534^T = CCUG 59229^T).

Description of *Aliarcobacter lanthieri* comb. nov.
Basonym: *Arcobacter lanthieri* Whiteduck-Léveillée et al. 2015.
The description is the same given by Whiteduck-Léveillée et al. (2015). The type strain is AF1440^T (= LMG 28516^T = CCUG 66485^T).

Description of *Aliarcobacter faecis* comb. nov.
Basonym: *Arcobacter faecis* Whiteduck-Léveillée et al. 2016.
The description is the same given by Whiteduck-Léveillée et al. (2016). The type strain is AF1078^T (= LMG 28519^T = CCUG 66484^T).

The first Corrigendum stated that the original article had been updated. This action was reversed with the publication of the Erratum. The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way.

ACKNOWLEDGMENTS

The authors are indebted to Prof. Aharon Oren (Jerusalem, Israel) for expert guidance on following the rules on bacterial nomenclature as provided by The International Code of Nomenclature of Prokaryotes (ICNP).

REFERENCES

De Smet, S., Vandamme, P., De Zutter, L., On, S. L. W., Douidah, L., and Houf, K. (2011). *Arcobacter trophiarum* sp. nov., isolated from fattening pigs. *Int. J. Syst. Evol. Microbiol.* 61, 356–361. doi: 10.1099/ijs.0.022665-0

Houf, K., On, S. L. W., Coenye, T., Debruyne, L., De Smet, S., and Vandamme, P. (2009). *Arcobacter thereius* sp. nov., isolated from pigs and ducks. *Int. J. Syst. Evol. Microbiol.* 59, 2599–2604. doi: 10.1099/ijs.0.006650-0

Houf, K., On, S. L. W., Coenye, T., Mast, J., Van Hoof, J., and Vandamme, P. (2005). *Arcobacter cibarius* sp. nov., isolated from boiled carcasses. *Int. J. Syst. Evol. Microbiol.* 55, 713–717. doi: 10.1099/ijs.0.63103-0

Neill, S. D., Campbell, J. N., O'Brien, J. J., Weatherup, S. T. C., and Ellis, W. A. (1985). Taxonomic position of *Campylobacter cryaerophila* sp. *nov*. *Int. J. Syst. Bacteriol.* 35, 342–356. doi: 10.1099/00207713-35-3-342

Vandamme, P., Vancanneyt, M., Pot, B., Mels, L., Hoste, B., Dewettinck, D., et al. (1992). Polyphasic taxonomic study of the emended genus *Arcobacter* with *Arcobacter butzleri* comb. nov. and *Arcobacter skirrowii* sp. nov., an aerotolerant bacterium isolated from veterinary specimens. *Int. J. Syst. Bacteriol.* 42, 344–356. doi: 10.1099/00207713-42-3-344

Whiteduck-Léveillée, K., Whiteduck-Léveillée, J., Cloutier, M., Tambong, J. T., Xu, R., Topp, E., et al. (2015). *Arcobacter lanthieri* sp. nov., isolated from pig and dairy cattle manure. *Int. J. Syst. Evol. Microbiol.* 65, 2709–2716. doi: 10.1099/ijs.0.000318

Whiteduck-Léveillée, K., Whiteduck-Léveillée, J., Cloutier, M., Tambong, J. T., Xu, R., Topp, E., et al. (2016). Identification, characterization and description of *Arcobacter* faecis sp. nov., isolated from a human waste septic tank. *Syst. Appl. Microbiol.* 39, 93–99. doi: 10.1016/j.syapm.2015.12.002

Copyright © 2019 Pérez-Cataluña, Salas-Massó, Diéguez, Balboa, Lema, Romalde and Figueras. 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: 20 April 2018
doi: 10.3389/fmicb.2018.00770

Species Delimitation, Phylogenetic Relationships, and Temporal Divergence Model in the Genus *Aeromonas*

J. G. Lorén[1], *Maribel Farfán*[1,2]* *and M. C. Fusté*[1,2]

[1] *Departament de Biologia, Sanitat i Medi Ambient, Secció de Microbiologia, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Spain,* [2] *Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain*

The definition of species boundaries constitutes an important challenge in biodiversity studies. In this work we applied the Generalized Mixed Yule Coalescent (GMYC) method, which determines a divergence threshold to delimit species in a phylogenetic tree. Based on the tree branching pattern, the analysis fixes the transition threshold between speciation and the coalescent process associated with the intra-species diversification. This approach has been widely used to delineate eukaryote species and establish their diversification process from sequence data. Nevertheless, there are few examples in which this analysis has been applied to a bacterial population. Although the GMYC method was originally designed to assume a constant (Yule) model of diversification at between-species level, it was later evaluated simulating other conditions. Our aim was therefore to determine the species delineation in *Aeromonas* using the GMYC method and asses which model best explains the speciation process in this bacterial genus. The application of the GMYC method allowed us to clearly delineate the *Aeromonas* species boundaries, even in the controversial groups, such as the *A. veronii* or *A. media* species complexes.

Keywords: species delimitation, GMYC, diversification model, *Aeromonas*, *mdh*, *recA*

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Tomochika Fujisawa, Institut Pasteur, France
David R. Arahal, Universitat de València, Spain
Antony T. Vincent, Institut de Biologie Intégrative et des Systèmes, Canada

***Correspondence:**
Maribel Farfán
mfarfan@ub.edu

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

Received: *12 January 2018*
Accepted: *05 April 2018*
Published: *20 April 2018*

Citation:
Lorén JG, Farfán M and Fusté MC (2018) Species Delimitation, Phylogenetic Relationships, and Temporal Divergence Model in the Genus Aeromonas. Front. Microbiol. 9:770. doi: 10.3389/fmicb.2018.00770

INTRODUCTION

As in other organisms, diversification in bacteria leads to entities that we call species, which group together organisms that have evolved separately from others. Species can be differentiated from populations of individual strains by a maximum likelihood method that determines the point of transition of the evolutionary processes from the level of species (speciation and extinction) to the population (coalescence). The entities determined in this way maintain the biological properties and the levels of sequence divergence of the traditionally defined species.

Several methods have been proposed for species delimitation in bacteria. The classical approach uses a sequence divergence of 3 per cent pairwise distance in 16S rRNA (Schloss and Handelsman, 2006) or the 1 percent recommended by Acinas et al. (2004) as a threshold for separating species. However, rates of substitution may vary among lineages, as can the levels of variation within and between species, which makes it difficult to assume a universal threshold. Other widely applied methods for species delimitation are based on multilocus sequences, or more recently whole

genome sequencing of several individuals has been used to determine the average nucleotide identity (ANI) (Richter and Roselló-Mora, 2009) and *in silico* DNA-DNA hybridization (*is*DDH) (Meier-Kolthoff et al., 2013) values to separate species. When comparing these methods with traditional phenotypic approaches, discrepancies can sometimes arise.

The study of evolutionary patterns in prokaryotes is hampered by the lack of a reliable fossil record, limited morphological differentiation and frequently complex taxonomic relationships. The few studies in this field suggest a constant rate of diversification in free-living or symbiotic bacteria (Martin et al., 2004; Vinuesa et al., 2005; Sanglas et al., 2017), while in some pathogens, such as *Borrelia burgdorferi,* it seems to follow a radiation pattern (Morlon et al., 2012). Differences in the cladogenesis of *B. burgdorferi* could be explained by its association with vertebrate and arthropod hosts, which could restrict the gene flow between populations. Understanding the evolution of prokaryote biological diversity therefore remains a significant challenge for biologists (Barraclough et al., 2009).

The main objective of our research in the last years has been the study of diversification in the genus *Aeromonas*, a γ-*Proteobacteria* that comprises a group of Gram-negative, rod-shaped bacteria, which are found in aquatic environments worldwide and are members of the microbiota (as well as primary or secondary pathogens) of fish, amphibians and other animals (Martin-Carnahan and Joseph, 2005; Janda and Abbott, 2010). *Aeromonas* is an ideal genus to study the diversification processes in bacteria because its species, a combination of free-living bacteria and host-associated strains, can be isolated from a wide variety of habitats. We began by investigating the rate and pattern of cladogenesis in this bacterial genus from a collection of strains including the type strains of all the *Aeromonas* species, using the sequences of five housekeeping genes (Lorén et al., 2014). However, the frequently high intra-specific diversity in bacteria is not always adequately reflected by the type strain.

We therefore performed a second analysis using molecular data from the sequences of two housekeeping genes (*mdh* and *recA*) obtained from 150 strains belonging to 27 species of *Aeromonas* (Sanglas et al., 2017). Phylogenies that mix variation between and within species often need to be reduced to trees with only one sequence per species to avoid invalidating the results (Fontaneto et al., 2012). To fulfill these conditions we constructed a tree with the consensus sequence for each species and used the BEAST program to obtain the species tree (Sanglas et al., 2017). In both cases the results of the analysis allowed us to determine that the process of speciation in *Aeromonas* follows a constant model of diversification.

The genus *Aeromonas*, includes 30 taxonomic species, many of them described recently. However, some descriptions were based on only one strain, or using techniques such as the 16S rRNA gene sequences, which have been demonstrated to have a low discriminatory power for delimiting species in *Aeromonas* (Figueras et al., 2000). Consequently, the improper characterization of some *Aeromonas* species has led to several reclassifications, as in the case of *A. aquariorum* (Beaz-Hidalgo et al., 2013), *A. ichthiosmia* (Huys et al., 2001), *A. culicicola* (Huys et al., 2005), among others. In addition, the genus includes several species complexes, such as *A. hydrophila* (Martin-Carnahan and Joseph, 2005), *A. veronii* (Silver et al., 2011) or *A. media* (Talagrand-Reboul et al., 2017), which are constituted by groups of strains with high intra-specific heterogeneity, giving rise to controversies about the species delineation.

In this study, we aimed to determine the species delineation and corroborate the diversification process in *Aeromonas* by applying a sequence-based method that allows the use of several strains per species, the Generalized Mixed Yule Coalescent (GMYC) method. This approach has been widely used to delineate eukaryote species and establish the diversification process from sequence data (Pons et al., 2006; Fontaneto et al., 2007; Lahaye et al., 2008). Nevertheless, there are only a few previous examples in which this analysis has been applied to a bacterial population (Barraclough et al., 2009; Powell, 2012). The method determines a divergence threshold to delimit species in a phylogenetic tree. Based on the tree branching pattern, the analysis fixes the transition threshold between speciation and the coalescent process associated with the intra-species diversification. The pruned tree provided by the threshold allows the use of classical methods of diversification analysis to determine the speciation process followed by the species in the analyzed population.

It is presupposed that branch lengths between species are derived from speciation and extinction rates, whereas branch lengths within a species reflect a coalescence process at the population level (Pons et al., 2006). The method determines the locations of ancestral nodes that define putative species and applies a likelihood ratio test to assess the fit of the branch lengths to a mixed lineage birth-population coalescent model (Pons et al., 2006). As the analysis to determine the transition threshold is based on the tree branching pattern, the tree has to accomplish certain requirements, for example, be ultrametric, fully resolved (without multifurcations) and reliable (with a high support in the nodes).

Our aim was therefore to determine the species delineation in *Aeromonas* based on the GMYC method. Although initially this method assumed diversification at between-species level follows a constant (Yule) model (Pons et al., 2006; Fontaneto et al., 2007), it was later evaluated simulating other diversification models (Fujisawa and Barraclough, 2013). Despite previous results indicate that *Aeromonas* follows a constant (Yule) model of diversification (Lorén et al., 2014; Sanglas et al., 2017), we also determined which model best explains the speciation process in this bacterial genus.

MATERIALS AND METHODS

Gene Sequences

A collection of 147 *Aeromonas* strains, representative of the 30 species recognized up to August 2017, was selected for the study. Two housekeeping genes (*mdh* and *recA*) were chosen for the analysis; for each strain, the full-length sequences for both genes were previously obtained (Sanglas et al., 2017). In the case of the more recently accepted *Aeromonas* species: *A. lacus*, *A. aquatica* and *A. finlandiensis*, sequences were obtained from genomes

available in the GenBank[1]. The strains and sequences are listed in Supplementary Table S1, including the GenBank accession numbers.

Data Sets

Phylogenetic reconstruction was carried out from the concatenated sequences of *mdh* and *recA* genes. For each gene, the translated sequences were aligned using the ClustalW program implemented in MEGA6 (Tamura et al., 2013) and translated back to obtain the nucleotide alignments. Both alignments were concatenated with the DAMBE program (v5.3.10; Xia, 2013) and later the sequences were checked for the presence of incongruences or gaps. Divergent and ambiguously aligned blocks were also removed using the Gblocks program (Castresana, 2000).

Sequence Analysis

To ascertain if the sequences used allowed a good species discrimination we determined the intra- and inter-specific distances among the different *Aeromonas* species. Data were graphically depicted with box plots obtained using the R package ggplot2 (Wickham, 2016). Sequences were also analyzed with different distance models that consider the influence of saturation, base frequency, transition–transversions or multiple substitutions. All the analyses were conducted with the function dist.dna implemented in the R package ape (Paradis et al., 2004). To determine the polymorphic sites along the sequences we construct a dots plot graph with the R package phyclust (Chen, 2010, 2011). The R package NbClust (Charrad et al., 2014) was used to determine the number of clusters from the *p* distances of the alienated sequences (Supplementary Table S2).

Phylogenetic Reconstruction

Bayesian phylogenetic trees were reconstructed with the BEAST program (v1.8.1; Drummond and Rambaut, 2007; Drummond et al., 2012) from the data sets. The model of evolution for the each gene was determined using the jModelTest 2 program (Darriba et al., 2012). The general time-reversible model with discrete gamma distribution and invariant sites (GTR + G + I) was selected as the best-fit model of nucleotide substitution. The Bayesian analysis was performed using a GTR model with four gamma categories, a Yule process of speciation, and a constant clock model of evolution as the tree priors, as well as other default parameters. We used the divergence time between *Escherichia coli* and *Salmonella enterica* estimated by Ochman and Wilson as the calibration point (Ochman and Wilson, 1987a,b). Accordingly, we calibrated the divergence of *Aeromonas* with a normally distributed prior with a mean of 140 Ma and a standard deviation of 10 Ma. We performed three independent Markov Chain Monte Carlo (MCMC) runs 20 million generations, sampling every 2,000 generations. Posterior distributions for parameter estimates and likelihood scores to approximate convergence were visualized with the Tracer program (v1.6.0; Rambaut et al., 2014). Visual inspection of traces within and across runs, as well as the effective sample sizes (ESS) of each parameter (>200), allowed us to confirm that the analysis was adequately sampled. A maximum clade credibility (MCC) tree was chosen by TreeAnnotator (v1.8.1; Drummond et al., 2012) from the combined output of the three MCMC runs using the LogCombiner program[2] after the removal of the initial trees (20–25%) as burn-in. The MCC tree was visualized with the program FigTree (v1.4.2)[3].

GMYC Analysis

The GMYC analysis uses the information contained in a tree to delimit species and determine the diversification model. To analyze the influence of the priors chosen, we constructed 3 different trees from the same DNA sequence alignment, changing the model used to express the branching pattern of the tree (the Yule or a coalescent model) and the rate of molecular evolution, considering a constant or a relaxed clock, as recommended by Michonneau (unpublished). The topology of the trees would be almost identical in all cases; however, the branch lengths would vary.

The analysis was conducted using the function gmyc available in the splits package implemented in R (Ezard et al., 2014), with the single threshold option as recommended by Fujisawa and Barraclough (2013). The method compares the null (no threshold) and alternative hypothesis (one threshold) and infers the number of genetic entities. Branching events between species are modeled with a Yule model (Barraclough and Nee, 2001), while branching events within species are adjusted to a neutral coalescent process (Hudson, 1991).

[1]http://www.ncbi.nlm.nih.gov/genbank/

[2]http://beast.bio.ed.ac.uk/logcombiner

[3]http://beast.bio.ed.ac.uk/figtree

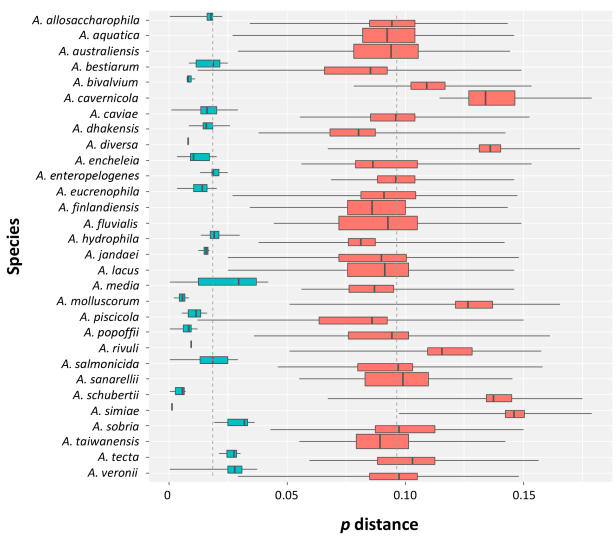


FIGURE 1 | Intra- and inter-specific distances in *Aeromonas* species. Box plots showing the intra- (turquoise) and inter-specific (salmon) *p* distances obtained from the data set. The ends of the boxes correspond to the first and third quartiles. The ends of the horizontal lines indicate the highest and lowest values. The black vertical line dividing the boxes represents the median of the data.

Results are evaluated by a likelihood ratio test (LRT) between the null hypothesis, which considers that all strains analyzed constitute a single species, and the alternative hypothesis (GMYC model) that assumes the existence of a species-delimiting threshold. The LRT significance was calculated using a chi-square test with 2 degrees of freedom.

The gmyc function gives a likelihood score for the model that considers all sequences belonging to the same species, and a likelihood score considering the sequences split in different species. The output also lists how many clusters and entities are associated with the highest likelihood score with the corresponding confidence intervals (CI) and the estimated threshold time when there was a transition between the speciation- and coalescent-level events. The R package also contains functions that plot (1) the number of lineages-through-time (LTT), with the inferred position of the threshold (red vertical line); (2) the likelihood profile through time; (3) the tree with the clusters highlighted in red. Additionally, the "support" for the delineated species can be plotted, indicating whether the results are reliable or not.

The use of only one indirect calibration point in the construction of the phylogeny, due to the absence of more reliable calibration data, can be a source of uncertainty. As the GMYC method relies on relative rather than absolute branch lengths, we repeated the analysis with the same alignment but removing the outgroup, obtaining the corresponding tree. Following the suggestion of Fujisawa and Barraclough (2013), we scaled branch lengths in the tree to have a root age of 1.0 before running the analysis.

Diversification Analyses

A standard lineage-through-time (LTT) plot was constructed using the R package ape (Paradis et al., 2004) to graphically visualize and evaluate the temporal pattern of lineage diversification in *Aeromonas*.

We used the birth–death likelihood (BDL) tests implemented in laser (Rabosky, 2009) to detect the diversification model and the speciation and extinction rates (λ and μ) from the pruned tree obtained by cutting the original tree at the threshold given by the GMYC analysis. To test the null hypothesis of no-rate change versus variable-rate change in diversification, we have applied the ML approach of Rabosky, the test ΔAIC_{RC} (Rabosky, 2009). This statistic is calculated as: $\Delta AIC_{RC} = AIC_{RC} - AIC_{RV}$, where AIC_{RC} is the Akaike information criterion (AIC) score for the best fitting

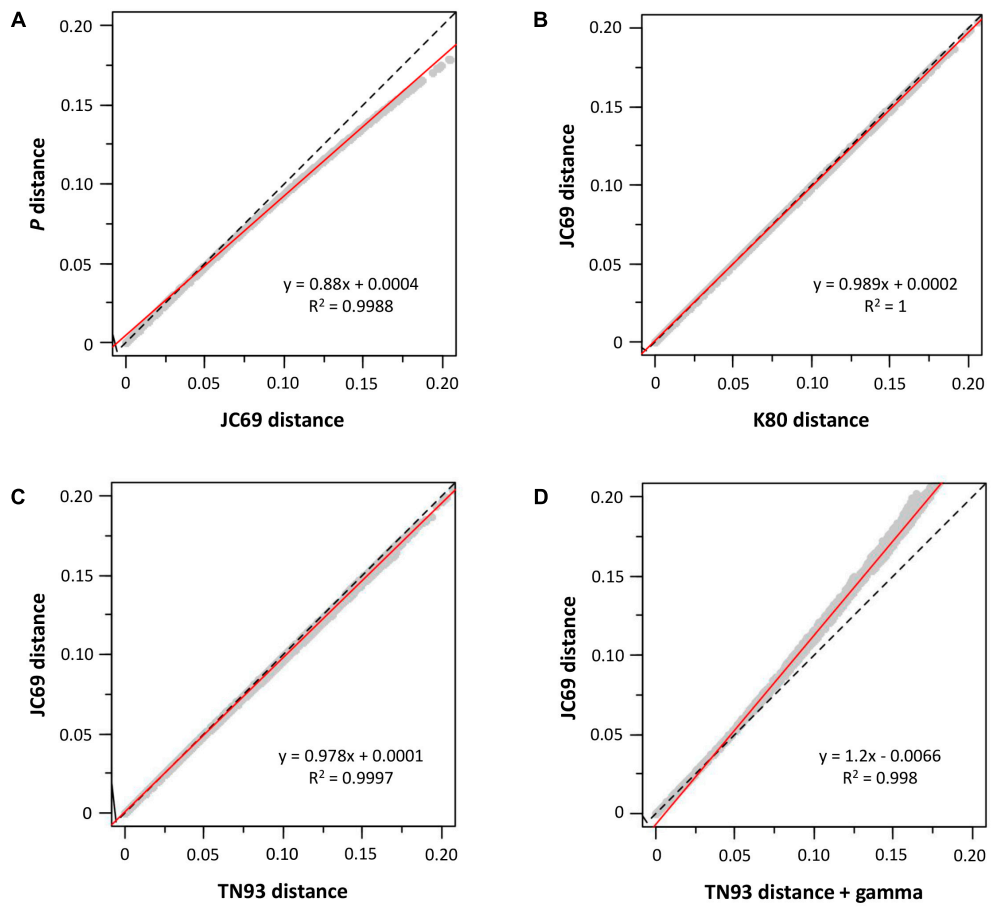


FIGURE 2 | Sequence analysis. Graphs **(A–D)** show the influence of different substitution models. The dashed line represents the equation y = x, and the red continuous line is the regression line whose equation appears in the corresponding graph.

rate-constant diversification model, and AIC_{RV} is the AIC for the best fitting variable-rate diversification model. Thus, a positive value for AIC_{RV} indicates that the data are best approximated with a rate-variable model, while a negative AIC_{RV} value suggests a rate-constant model of diversification. We tested eight different models, two of which were rate-constant (pure-birth or Yule and birth-death) and six were rate-variable (DDL, DDX, Yule 2-,3-,4- and 5- rates) (Lorén et al., 2014).

We calculated the gamma (γ) statistic (Pybus and Harvey, 2000; Fordyce, 2010) and its significance by simulating 5,000 phylogenies, as described previously (Lorén et al., 2014). This statistic compares the relative node positions in a phylogeny with those expected under a constant diversification rate model, in which the statistic follows a standard normal distribution. Positive γ values evidence that nodes are closer to the tips than expected under the constant rate model. When γ is negative, the internal nodes are closer to the root than expected under a constant model, indicating a decrease in diversification through time. In addition, we compared the observed empirical gamma value with a distribution of the gamma statistics obtained by simulation.

RESULTS

Sequence Analysis

The analysis involved 147 *Aeromonas* strains. The number of total positions analyzed was 1,979 bp. All positions containing gaps and missing data were eliminated in the construction of the phylogenetic tree.

The intra- and inter-specific distances determined from the sequences (Supplementary Figure S1 and Supplementary Table S3), with the exception of the conflicting *A. veronii* group and the *A. bestiarum*/*A. piscicola* cluster, allowed a good discrimination of the species with no overlap among the distances (**Figure 1**).

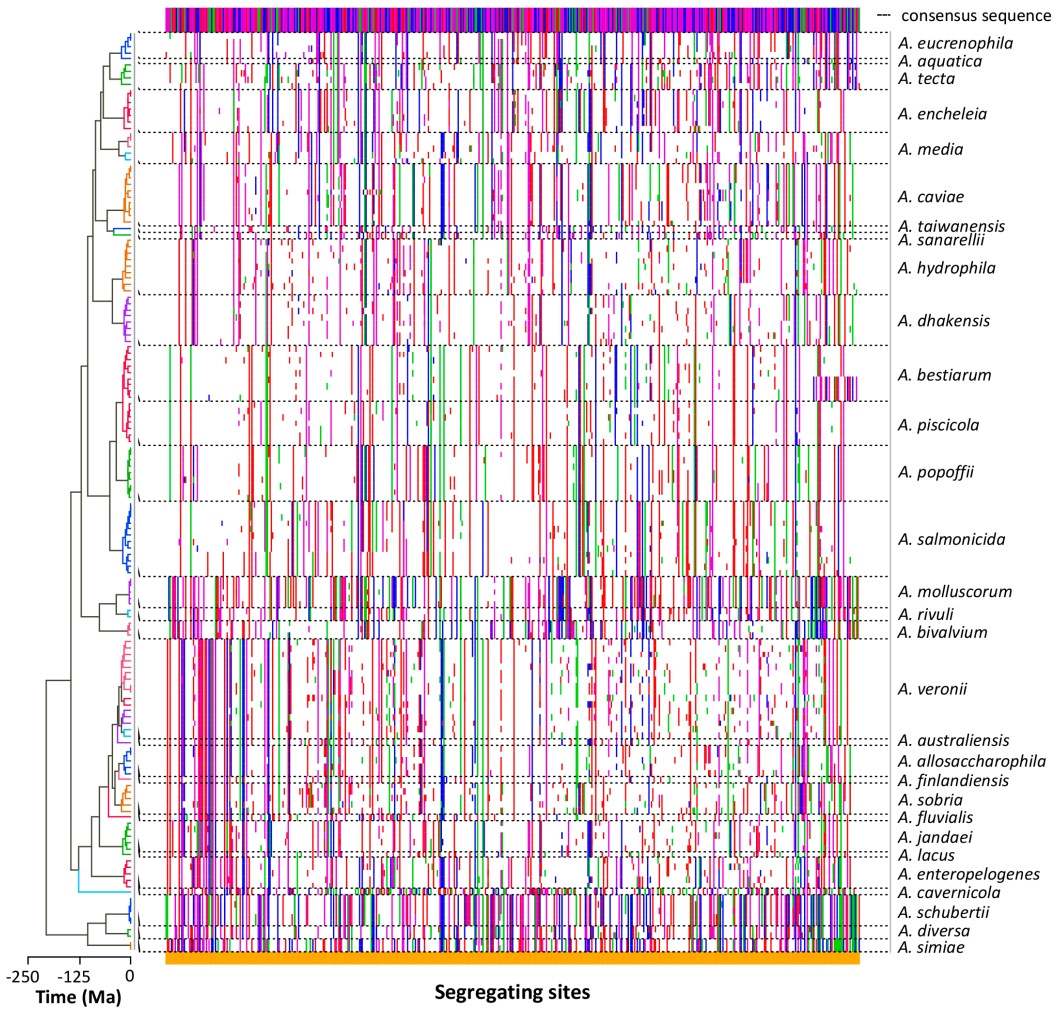


FIGURE 3 | Sequence polymorphism. The dots plot shows the segregating sites determined in each sequence arranged in the same order as in the tree **(left)**. Polymorphic differences between the sequences and the consensus sequence **(top)** are marked in colors depending on the base (A green, G blue, C purple, T red). Dashed lines delimit the sequences belonging to the same *Aeromonas* species.

We also determined the sequence evolution of the concatenated sequences of *mdh* and *recA*. As can be seen in **Figure 2**, we analyzed our sequences for the influence of saturation (**Figure 2A**), base frequencies (**Figure 2B**), transitions and transversions (**Figure 2C**), and the gamma correction for multiple substitutions (**Figure 2D**). The sequences showed no influence of saturation, base frequency bias or the transition-transversion ratio; only the heterogeneity in the inter-site substitution rate could have some effect (**Figure 2D**).

The nucleotide substitution analysis (**Figure 3**) showed a clear diversity among the sequences of the strains belonging to the same species. Nevertheless, the different species exhibited segregating site patterns that allowed them to be separated.

Phylogenetic Reconstruction

The best-fit models of sequence evolution were implemented according to the Akaike Information Criterion (AIC) scores for the substitution models evaluated, using jModeltest. The general time reversible (GTR) model was selected as the best model of evolution for the concatenated sequences using a discrete gamma distribution and a fraction of invariable sites (GTR + G + I). **Figure 4** shows the *Aeromonas* Bayesian phylogeny, all the strains belonging to the same species clustering together in the same group. The posterior values obtained for each node were close to 1 for the majority of the main clades. The figure also shows the clusters obtained with the *p* distance analysis (Supplementary Table S2), which fully matched those established from the phylogeny.

Figure 5 depicts the phylogeny obtained with the corresponding lineage-through-time plot (LTT plot). The LTT plot (**Figure 5B**) clearly shows two different linear relationships with a sudden change in the slopes at approximately −25 Ma. This breakpoint roughly matches the region of the tree between −25 Ma and the present (**Figure 5A**), which accumulates the majority of the nodes (75%). We applied a linear regression

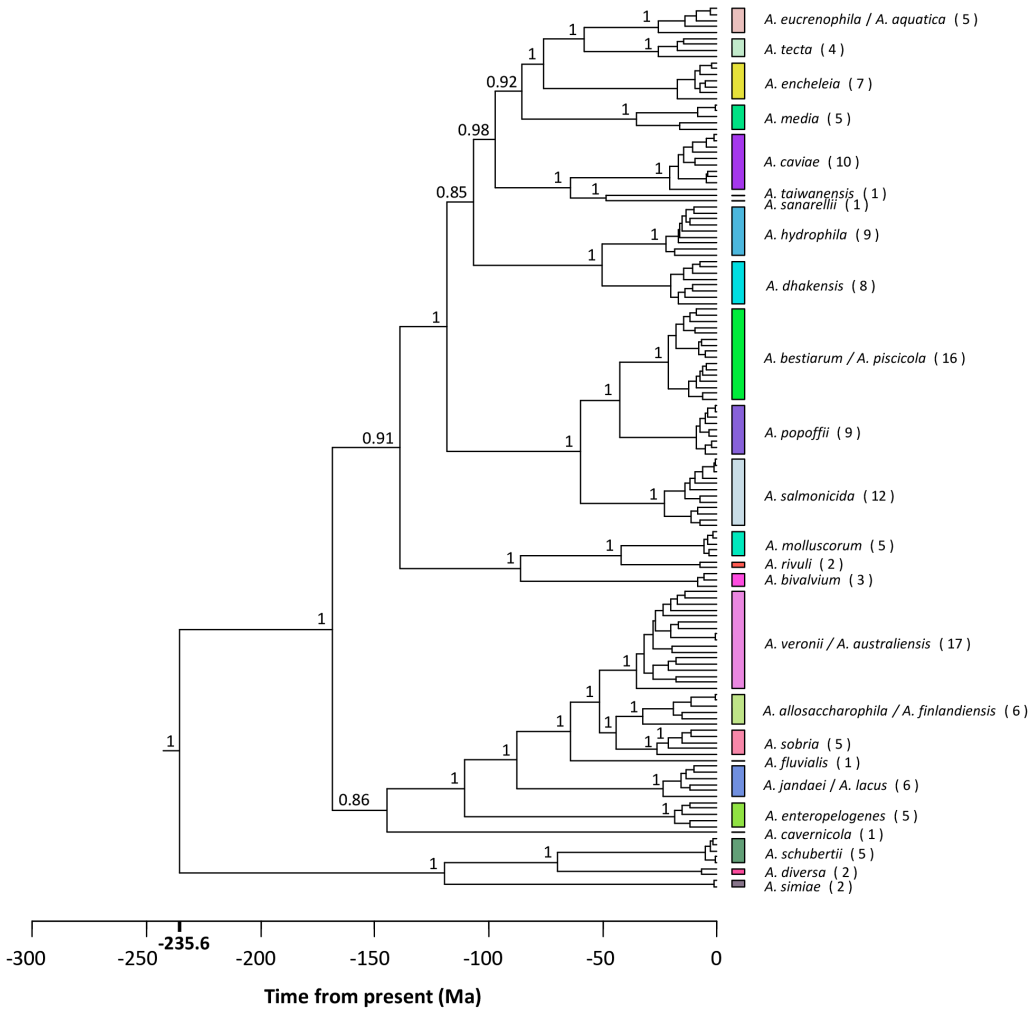


FIGURE 4 | *Aeromonas* phylogeny. Bayesian chronogram with the posterior probability values for the main clusters. Colored vertical bars on the right correspond to the clusters obtained by the NbClust analysis (Supplementary Table S3). The number of strains belonging to the same species is given in brackets. Scale bar at the bottom indicates the divergence time in millions of years (Ma, Mega annum).

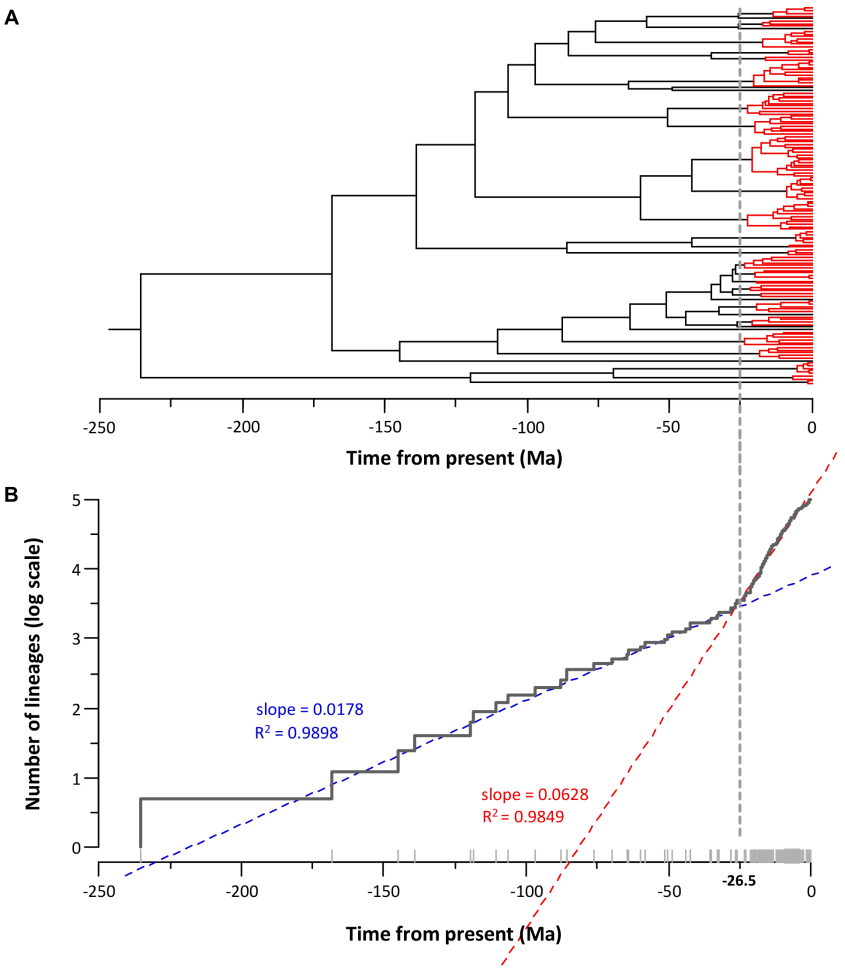


FIGURE 5 | *Aeromonas* diversification. **(A)** *Aeromonas* chronogram showing the temporal node distribution. Nodes posterior to –25 Ma (breakpoint) are marked in red. **(B)** LTT plot corresponding to the chronogram showing the breakpoint and the linear adjustment of the two segments (before and after the breakpoint).

model to the LTT plot points between the crown age (−235.6 Ma) and the breakpoint (−25 Ma) (segment I), and those obtained from the breakpoint to the present (segment II). The fit of the two segments to a linear model was very good, with R squared values of 0.9898 and 0.9849 for segments I and II, respectively. The two regression lines intersected at −26.5 Ma, and the slopes for segment I and II were 0.0178 and 0.0628 (**Figure 5B**). Using the Chow test (Chow, 1960), we verified that the two linear regression slopes differed significantly (F test = 1715.8, P value = $1.12e^{-100}$).

Generalized Mixed Yule Coalescent

The existence of distinct *Aeromonas* lineages was confirmed by the branch length analysis. The resulting LTT plot showed a steep upturn in branching rates toward the present, marking the transition from the between- to within-species rate of lineage branching (**Figure 6**). The GMYC model fitted a transition in branching rate occurring at −26.6 Ma. The support in the majority of the GMYC clusters was high, confirming the reliability of the results (**Figure 6**). The number of ML entities determined was 31 with a confident interval (CI) of 24–36 (**Table 1**). The λ values for the diversification and coalescent processes were 0.0092 and 0.2779, respectively. The scaling parameter p for the diversification process was close to 1 (1.154), which indicates a constant speciation rate model with no extinction (Yule model), while the value for the coalescent process was clearly below 1 ($6.186e^{-08}$) indicating a deficit of recent coalescent events (Fujisawa and Barraclough, 2013). No substantial differences were detected between Bayesian models using a coalescent or Yule prior when constructing the tree (**Table 1**).

The calibration point used to date the phylogeny also had no influence. Identical results were obtained when the dated tree was compared with the one obtained by repeating the analysis with the same alignment but removing the outgroup.

Diversification Analysis

As suggested by Rabosky (2006), we calculated the significance of ΔAIC_{RC} for the set of analyzed models by using the Yule model

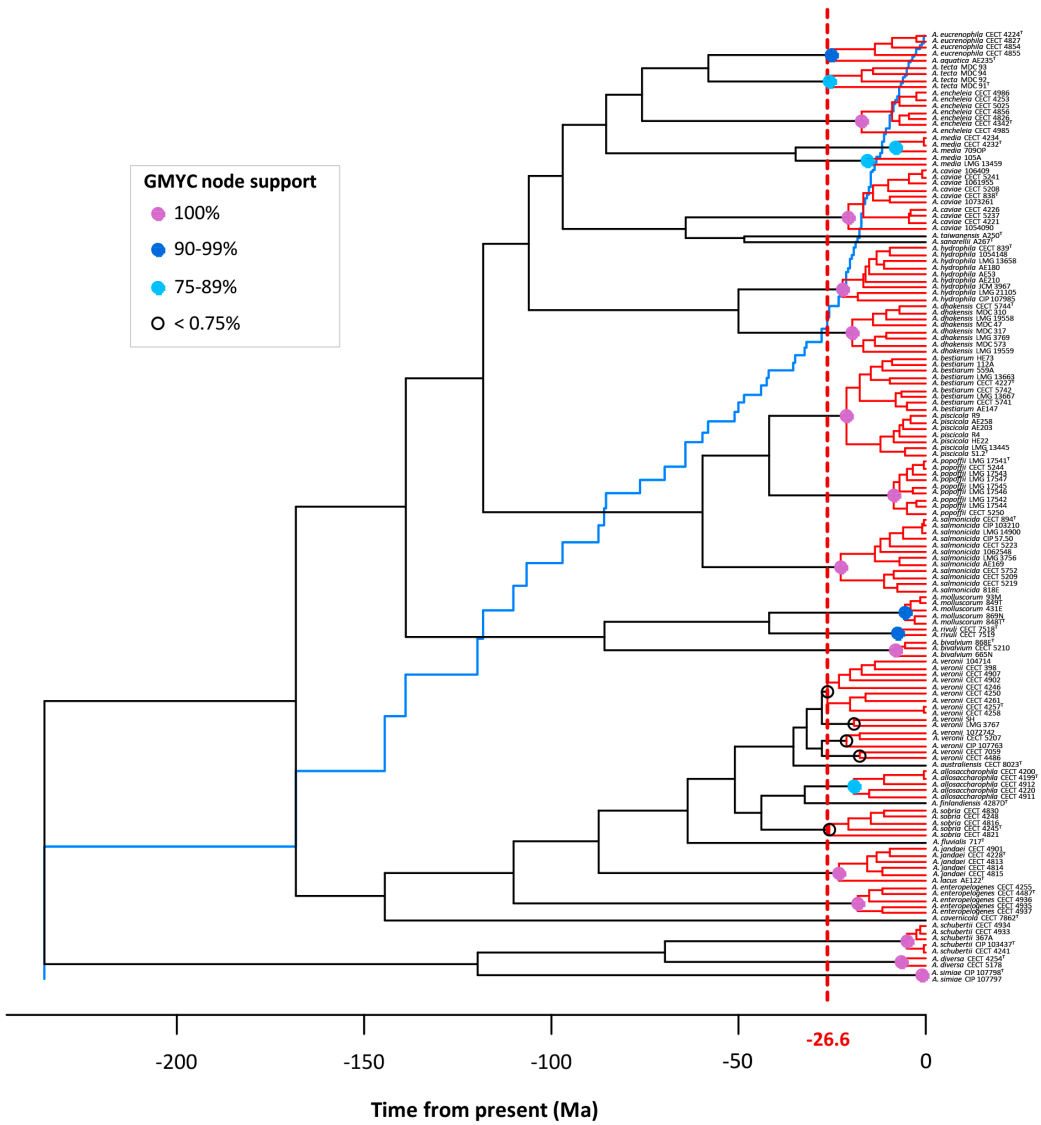


FIGURE 6 | GMYC analysis. The LTT plot (blue line) showing a clear abrupt change from the threshold (vertical dashed red line). Node support of the main GMYC clusters (red) is illustrated by colored circles (see legend in box).

to simulate 5,000 phylogenies of the same size and diversification rate as those obtained from our data, and determined the *P* value from the resulting distributions. As can be seen in **Table 2**, we cannot reject the null hypothesis of a Yule model to a level of significance of $\alpha = 0.05$, which means that the diversification in *Aeromonas* is constant.

For further corroboration, we determined the gamma statistic of Pybus and Harvey, a powerful tool principally used for comparing models of decreasing speciation rate through time and a constant rate of diversification. The estimated γ value from the chronogram (pruned tree) was 0.798 with a *P* value of 0.425, indicating that we cannot reject the null hypothesis of constant diversification for our phylogeny. Data obtained from the simulation also corroborate a constant diversification process in *Aeromonas* (**Figure 7**).

DISCUSSION

The choice of appropriate genes is crucial for the reconstruction of reliable molecular phylogenies. They should be housekeeping genes, evolving at a constant rate (good molecular clocks), and have unsaturated synonymous and non-synonymous substitutions. Unlike the majority of phylogenetic studies, in our analysis we checked our sequences to ensure that the genes used in this work fulfilled the above conditions and allowed a good species separation.

The phylogeny constructed from the concatenated sequences corroborates the monophyletic origin of this group of bacteria. In the chronogram obtained, the majority of the nodes were strongly supported, with posterior values close to 1. In addition, the main clade distribution was in

TABLE 1 | GMYC analysis.

	GMYC	
	Yule Constant clock	**Coalescent Constant clock**
Number of ML entities (CI)[a]	31 (24 – 36)	31 (25 – 38)
Threshold time	−26.60	−24.60
pdiv value[b]	1.1548	1.2170
pcoal value[c]	6.1860e-08	2.0497e-09
λ_{div}[b]	0.0092	0.0080
λ_{coal}[c]	0.2779	0.3026

[a]*ML, maximum likelihood; CI, confidence interval.* [b]*div, diversification process.* [c]*coal, coalescent process.*

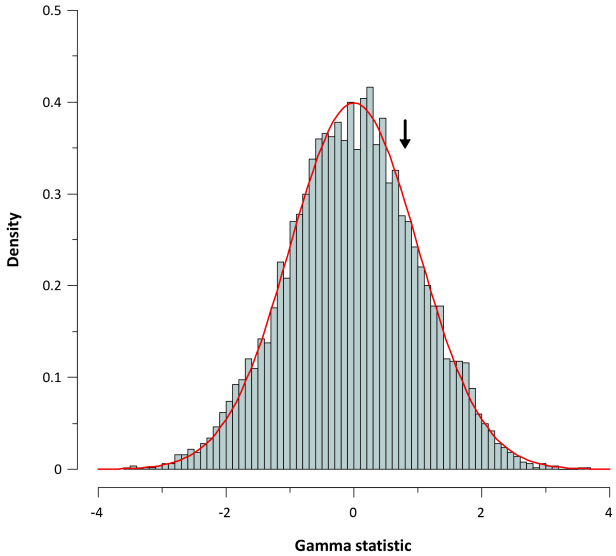


FIGURE 7 | Gamma statistic distribution. Gamma statistic distribution obtained by simulating 5,000 phylogenies under the Yule model. The arrow indicates the empirical gamma (γ) value obtained.

agreement with previously published phylogenies (Roger et al., 2012; Colston et al., 2014; Lorén et al., 2014; Sanglas et al., 2017). We obtained a perfect clustering of the strains belonging to the same species, including those considered synonymous.

The analyses based on the LTT plot detected an increase in the diversification rates from a certain point in time. As **Figure 5** illustrates, a high percentage of nodes accumulate close to the present due to the greater similarity of the sequences belonging to the same species.

The LTT plot constructed from the phylogenetic tree clearly shows two different linear relationships with a sudden change in the slopes at −26.5 Ma. The fit of the two segments to a linear model was very good, with high *R* squared values. The Chow test verified that the two linear regression slopes differed significantly (*P* value = $1.12e^{-100}$), which indicates a change in the rate of lineage branching.

In this work we carried out a quantitative analysis of sequence data to delimit putative *Aeromonas* species by detecting shifts in the rate of lineage branching. The branch-length analysis is based on a probabilistic model that distinguishes between species diversification and coalescent processes. This approach compensates for undefined species limits by including confidence intervals when allocating species-defining nodes and does not rely on population limits; thus, species represented by a single strain can be included (Pons et al., 2006).

Some studies recommend the use of BEAST for the tree construction as input for the GMYC analysis (Monaghan et al., 2009; Tang et al., 2014). From a set of given data (sequences) and a substitution model, the Bayesian Inference performs a probability analysis in search of the best set of trees that maximize the posterior probability. From an initial tree, it uses the Markov Chains Monte Carlo to evaluate the posterior probability of the different states proposed. After generating a large number of trees (frequently around 10,000), it uses the subsequent probability to ascertain how many times a node is repeated in each tree and this is done for each node. Subsequently, the Bayes theorem is used to build a single phylogenetic tree (maximum clade credibility, MCC). The same method provides an estimate of the reliability of the results obtained through the posterior probability values of each node, without the need to evaluate it *a posteriori*.

Although Monaghan et al. (2009) recommended the use of the coalescent tree prior instead of the Yule prior for constructing the tree for the GMYC analysis, in our work, in agreement with Talavera et al. (2013), we obtained identical results with the different BEAST options.

Barraclough et al. (2009) used the GMYC method to delimit bacterial species based on 16S rRNA sequences, concluding that

TABLE 2 | Diversification models.

Models	**AIC[a]**							
	Pure birth	**Birth-death**	**DDX**	**DDL**	**Yule-2-rates**	**Yule-3-rates**	**Yule-4-rates**	**Yule-5-rates**
	150.55	151.83	152.21	152.55	152.80	152.30	153.02	155.56
ΔAIC_{RC} test[b]	**Best constant model**		**Best variable model**	**ΔAIC_{RC}**	***P* value[c]**	**Best model**		**λ (ML)**
	Pure birth		DDX	−1.6569	0.753	Pure birth (Yule)		0.017

[a] *AIC, Akaike Information Criterium.* [b] *ΔAIC_{RC} test (Rabosky, 2006).* [c] *P value obtained by simulation of 5,000 phylogenies.*

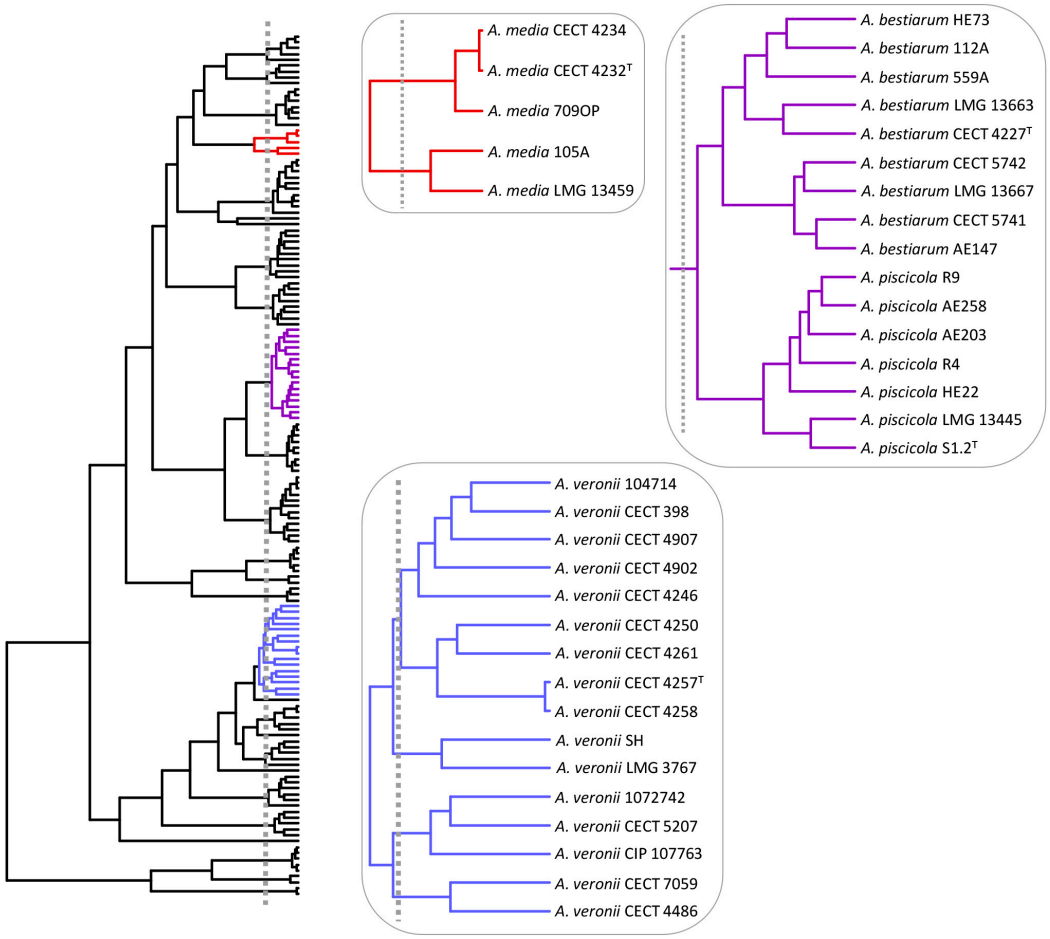


FIGURE 8 | *Aeromonas* species delimitation. Phylogenetic tree obtained with the GMYC method. In boxes are the species clusters that present taxonomic uncertainties when using the estimated breakpoint to delimit species.

the 16S rRNA is a rather conservative molecule for surveying bacterial diversity and cannot be used to distinguish putative species. They suggest that studies using sequences of multiple genes or whole genomes of individuals would be more useful. Accordingly, we applied the GMYC method to delimit species in the genus *Aeromonas,* and the results corroborate the species previously described with other methods (phenotypic analysis, MLSA, genome sequence analysis) for this bacterial genus. The resulting tree exhibited the branching rate pattern of the species-to-population transition. Only the controversial *A. veronii* and *A. media* species complexes, the *A. piscicola*/*A. bestiarum* group, and two of the most recently proposed species were not entirely resolved.

The *A. veronii* group includes two biovars, *A. veronii* Veronii and *A. veronii* Sobria, as well as two other *Aeromonas* species considered as synonymous, *A. culicicola* (Huys et al., 2005) and *A. ichthiosmia* (Huys et al., 2001). When we analyzed the *A. veronii* species complex, four groups of strains were revealed: a main cluster consisting of 9 strains, including the two biovars, *A. veronii* Veronii and *A. veronii* Sobria; a second group with two strains, including *A. ichthiosmia*; a third comprising three strains, including *A. culicicola*; and the fourth with two strains (**Figure 8**).

Considering the GMYC results, it would be interesting to review the inclusion of *A. culicicola* and *A. ichthiosmia* in the already controversial *A. veronii* group (Huys et al., 2001; Esteve et al., 2003). In addition, Colston et al. (2014) determined the ANI and *is*DDH values from the genomes of different *Aeromonas* species. When the authors compared the type strain of *A. culicicola* and *A. ichthiosmia* with all the strains defined as *A. veronii*, the *is*DDH values obtained were below the threshold of 70% (69.1–69.6% for *A. culicicola* and 67.4–68.2% for *A. ichthiosmia*) and ANI values close to 96%, suggesting these species are different from *A. veronii*.

Aeromonas media described by Allen et al. (1983) traditionally comprised two hybridization groups: *A. media* HG5A and HG5B (Hänninen and Siitonen, 1995), with the strains LMG 13459 (CDC 0862-83) and CECT 4232 (ATCC 33907) representing both groups (Popoff et al., 1981). This was later corroborated by Küpfer et al. (2006) and more recently by Talagrand-Reboul et al. (2017). Our results reveal the existence of two groups of strains outside the GMYC threshold that determine the

species in *Aeromonas* (**Figure 8**), which is in accordance with Talagrand-Reboul et al. (2017), who analyzed the relationships of 40 *A. media* strains with the recently proposed species of *A. rivipollensis*. These authors also identified two groups of strains inside the *A. media* cluster with ANIb (ANI calculator and EzGenome) and *is*DDH values between groups of ≤94.6% and ≤61.2%, respectively. These values, which are below the considered threshold for isolates of the same species, support the existence of two different species within this group.

Another controversial GMYC clade grouped together *A. bestiarum*/*A. piscicola* (**Figure 8**). Colston et al. (2014) determined 0.04 substitutions per site for this pair of strains, which was clearly lower than those determined for other *Aeromonas* species, such as *A. eucrenophila*/*A. tecta* (1.0) or *A. schubertii*/*A. diversa* (0.9). Furthermore, the ANI or *is*DDH values for *A. bestiarum*/*A. piscicola* were close to the threshold to be considered as members of the same species.

Recently, three new *Aeromonas* species have been formally accepted: *A. finlandensis*, *A. lacus* and *A. aquatica* (Beaz-Hidalgo et al., 2015). Nevertheless, our analysis showed that *A. lacus* and *A. aquatica* should probably not be considered new species (**Figure 6**). In the case of *A. lacus*, the results obtained are in agreement with the ANI values calculated from the genomes in the original description of these species. When *A. jandaei* was compared with *A. lacus*, the ANI value obtained was 95.38%, similar to the species cutoff of 96%.

The GMYC analysis also reveals that diversification in *Aeromonas* is a constant process and follows a Yule model. This result is independent of the number of sequences per species used, and agrees with a study of macroevolutionary dynamic models in a huge number of eukaryote and prokaryote taxa (Maruvka et al., 2013). Prokaryote populations are less affected by extinction and founder effects than larger less abundant organisms (Lynch and Conery, 2003; Butterfield, 2011). Bacterial species may be described as metapopulations extending over time and evolving separately from other species (Cohan, 2002a,b; Achtman and Wagner, 2008). Although microbial dispersion can be limited by geographical barriers and the resulting physical isolation may influence microbial evolution, due to their small size bacteria generally have an unrestricted capacity for dispersion (Finlay and Clarke, 1999) via a variety of passive mechanisms and over long distances (Fierer, 2008).

CONCLUSION

The GMYC method clearly delineated the species boundaries in *Aeromonas*, even in the controversial groups, such as the *A. veronii* or *A. media* species complexes. Some of these taxonomic uncertainties are due to the use of inappropriate methods for defining species [biochemical tests with miniaturized methods (API) or 16S rRNA sequences]. It would therefore be interesting to revise bacterial species separation using the emerging new analyses based on genome sequences (ANI, *is*DDH). The method presented here, widely used to delimit species in eukaryotes, offers an alternative that is easy to use and fully practicable for all laboratories. In addition, the GMYC method has promising potential for phylogenetic community ecology studies.

AUTHOR CONTRIBUTIONS

JGL and MCF conceived and designed the research. JGL and MF performed the computations. JGL, MCF, and MF analyzed the data and discussed the results. All authors contributed to and revised the final manuscript version. They all approved the final version to be published.

FUNDING

This work was supported by a grant from the University of Barcelona for Open Access Publishing in scientific journals.

ACKNOWLEDGMENTS

We thank Dr. Ariadna Sanglas for her invaluable help and contribution of sequences to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00770/full#supplementary-material

REFERENCES

Achtman, M., and Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* 6, 431–440. doi: 10.1038/nrmicro1872

Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, L., Distel, D. L., et al. (2004). Fine scale phylogenetic architecture of a complex bacterial community. *Nature* 430, 551–554. doi: 10.1038/nature02649

Allen, D. A., Austin, B., and Colwell, R. R. (1983). *Aeromonas media*, a new species isolated from river water. *Int. J. Syst. Bacteriol.* 33, 599–604. doi: 10.1099/00207713-33-3-599

Barraclough, T. G., Hughes, M., Ashford-Hodges, N., and Fujisawa, T. (2009). Inferring evolutionary significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biol. Lett.* 5, 425–428. doi: 10.1098/rsbl.2009.0091

Barraclough, T. G., and Nee, S. (2001). Phylogenetics and speciation. *Trends Ecol. Evol.* 16, 391–399. doi: 10.1016/S0169-5347(01)02161-9

Beaz-Hidalgo, R., Latif-Eugenin, F., Hossain, M. J., Berg, K., Niemi, R. M., Rapala, J., et al. (2015). *Aeromonas aquatica* sp. nov., *Aeromonas finlandiensis* sp. nov. and *Aeromonas lacus* sp. nov. isolated from Finnish waters associated with cyanobacterial blooms. *Syst. Appl. Microbiol.* 38, 161–168. doi: 10.1016/j.syapm.2015.02.005

Beaz-Hidalgo, R., Martínez-Murcia, A., and Figueras, M. J. (2013). Reclassification of *Aeromonas hydrophila* subsp. *dhakensis* Huys et al. 2002 and *Aeromonas aquariorum* Martínez-Murcia et al. 2008 as *Aeromonas dhakensis* sp. nov. comb. nov. and emendation of the species *Aeromonas hydrophila*. *Syst. Appl. Microbiol.* 36, 171–176. doi: 10.1016/j.syapm.2012.12.007

Butterfield, N. J. (2011). Animals and the invention of the Phanerozoic Earth system. *Trends Ecol. Evol.* 26, 81–87. doi: 10.1016/j.tree.2010.11.012

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

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, W. C. (2010). *Phylogenetic Clustering with R Package Phyclust. Phyloclustering-Phylogenetic Clustering Website*. Available at: https://CRAN.R-project.org/package=phyclust

Chen, W. C. (2011). *Overlapping Codon Model, Phylogenetic Clustering, and Alternative Partial Expectation Conditional Maximization Algorithm*. Ph.D. dissertation, Iowa State University, Ames, IA.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regression. *Econometrica* 28, 591–605. doi: 10.2307/1910133

Cohan, F. M. (2002a). What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487. doi: 10.1146/annurev.micro.56.012302.160634

Cohan, F. M. (2002b). Sexual isolation and speciation in bacteria. *Genetica* 116, 359–370. doi: 10.1023/A:1021232409545

Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., and Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. *mBio* 5:e02136. doi: 10.1128/mBio.02136-14

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Esteve, C., Valera, L., Gutiérrez, C., and Ventosa, A. (2003). Taxonomic study of sucrose-positive *Aeromonas jandaei*-like isolates from faeces, water and eels: emendation of *A. jandaei* Carnahan et al. 1992. *Int. J. Syst. Evol. Microbiol.* 53, 1411–1419. doi: 10.1099/ijs.0.02504-0

Ezard, T., Fujisawa, T., and Barraclough, T. (2014). *Splits: Species' Limits by Threshold Statistics. R Package Version 1.0–19*. Available at: http://R-Forge.R-project.org/projects/splits/

Fierer, N. (2008). "Microbial biogeography: patterns in microbial diversity across space and time," in *Accessing Uncultivated Microorganisms: From the Environment to Organisms and Genomes and Back*, ed. K. Zengler (Washington DC: ASM Press), 95.

Figueras, M. J., Soler, L., Chacón, M. R., Guarro, J., and Martínez-Murcia, A. J. (2000). Extended method for discrimination of *Aeromonas* spp. by 16S rDNA RFLP analysis. *Int. J. Syst. Evol. Microbiol.* 50, 2069–2073. doi: 10.1099/00207713-50-6-2069

Finlay, B. J., and Clarke, K. J. (1999). Ubiquitous dispersal of microbial species. *Nature* 400:828.

Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., et al. (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS Biol.* 5:e87. doi: 10.1371/journal.pbio.0050087

Fontaneto, D., Tang, C. Q., Obertegger, U., Leasi, F., and Barraclough, T. G. (2012). Different diversification rates between sexual and asexual organisms. *Evol. Biol.* 39, 262–270. doi: 10.1007/s11692-012-9161-z

Fordyce, J. A. (2010). Interpreting the gamma statistic in phylogenetic diversification rate studies: a rate decrease does not necessarily indicate an early burst. *PLoS One* 5:e11781. doi: 10.1371/journal.pone.0011781

Fujisawa, T., and Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* 62, 707–724. doi: 10.1093/sysbio/syt033

Hänninen, M. L., and Siitonen, A. (1995). Distribution of *Aeromonas* phenospecies and genomospecies among strains isolated from water, foods or from human clinical samples. *Epidem. Infect.* 115, 39–50. doi: 10.1017/S0950268800058106

Hudson, R. R. (1991). Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7, 1–44.

Huys, G., Cnockaert, M., and Swings, J. (2005). *Aeromonas culicicola* Pidiyar et al. 2002 is a later subjective synonym of *Aeromonas veronii* Hickman-Brenner et al. 1987. *Syst. Appl. Microbiol.* 28, 604–609. doi: 10.1016/j.syapm.2005.03.012

Huys, G., Kämpfer, P., and Swings, J. (2001). New DNA-DNA hybridization and phenotypic data on the species *Aeromonas ichthiosmia* and *Aeromonas allosaccharophila*: *A. ichthiosmia* Schubert et al. 1990 is a later synonym of *A. veronii* Hickman-Brenner et al. 1987. *Syst. Appl. Microbiol.* 24, 177–182. doi: 10.1078/0723-2020-00038

Janda, J. M., and Abbott, S. L. (2010). The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clin. Microbiol. Rev.* 23, 35–73. doi: 10.1128/CMR.00039-09

Küpfer, M., Kuhnert, P., Korczak, B. M., Peduzzi, R., and Demarta, A. (2006). Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, gyrB and rpoD gene sequences. *Int. J. Syst. Evol. Microbiol.* 28, 604–609.

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., et al. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2923–2928. doi: 10.1073/pnas.0709936105

Lorén, J. G., Farfán, M., and Fusté, M. C. (2014). Molecular phylogenetics and temporal diversification in the genus *Aeromonas* based on the sequences of five housekeeping genes. *PLoS One* 9:e88805. doi: 10.1371/journal.pone.0088805

Lynch, M., and Conery, J. S. (2003). The origins of genome complexity. *Science* 302, 1401–1404. doi: 10.1126/science.1089370

Martin, A. P., Costello, E. K., Meyer, A. F., Nemergut, D. R., and Schmidt, S. K. (2004). The rate and pattern of cladogenesis in microbes. *Evolution* 58, 946–955. doi: 10.1111/j.0014-3820.2004.tb00429.x

Martin-Carnahan, A., and Joseph, S. W. (2005). "Genus I. *Aeromonas* stanier 1943, 213AL," in *Bergey's Manual of Systematic Bacteriology*, Vol. 2, eds G. M. Garrity, D. J. Brenner, N. R. Krieg, and J. T. Staley (New York, NY: Springer), 557–578.

Maruvka, Y. E., Shnerb, N. M., Kessler, D. A., and Ricklefs, R. E. (2013). Model for macroevolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2460–E2469. doi: 10.1073/pnas.1220014110

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequences-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., et al. (2009). Accelerated species inventory on Madagascar using coalescent-based models of species delimitation. *Syst. Biol.* 58, 298–311. doi: 10.1093/sysbio/syp027

Morlon, H., Kemps, B. D., Plotkin, J. B., and Brisson, D. (2012). Explosive radiation of a bacterial species group. *Evolution* 66, 2577–2586. doi: 10.1111/j.1558-5646.2012.01598.x

Ochman, H., and Wilson, A. C. (1987a). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* 26, 74–86.

Ochman, H., and Wilson, A. C. (1987b). "Evolutionary history of enteric bacteria," in *Escherichia coli* and *Salmonella typhimurium*: *Molecular and Cellular Aspects*, eds F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (Washington, DC: ASM Publications), 1649–1654.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in r language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., et al. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609. doi: 10.1080/10635150600852011

Popoff, M. Y., Coynault, C., Kiredjian, M., and Lemelin, M. (1981). Polynucleotide sequence relatedness among motile *Aeromonas* species. *Curr. Microbiol.* 5, 109–114. doi: 10.1007/BF01567430

Powell, J. R. (2012). Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods Ecol. Evol.* 30, 1–11. doi: 10.1111/j.2041-210X.2011.00122.x

Pybus, O. G., and Harvey, P. H. (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. Biol. Sci.* 267, 2267–2272. doi: 10.1098/rspb.2000.1278

Rabosky, D. (2009). *LASER: Likelihood Analysis of Speciation/Extinction Rates from Phylogenies. R Package Version 2.3*. Available at: https://cran.r-project.org/src/contrib/Archive/laser

Rabosky, D. L. (2006). Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60, 1152–1164. doi: 10.1111/j.0014-3820.2006.tb01194.x

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v1.6*. Available at: http://beast.bio.ed.ac.uk/Tracer

Richter, M., and Roselló-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Roger, F., Marchandin, H., Jumas-Bilak, E., Kodjo, A., ColBVH study group, and Lamy, B. (2012). Multilocus genetics to reconstruct aeromonad evolution. *BMC Microbiol.* 12:62. doi: 10.1186/1471-2180-12-62

Sanglas, A., Albarral, V., Farfán, M., Lorén, J. G., and Fusté, M. C. (2017). Evolutionary roots and diversification of the genus *Aeromonas*. *Front. Microbiol.* 8:127. doi: 10.3389/fmicb.2017.00127

Schloss, P. D., and Handelsman, J. (2006). Toward a census of bacteria in soils. *PLoS Comput. Biol.* 2:e92. doi: 10.1371/journal.pcbi.0020092

Silver, A. C., Williams, D., Faucher, J., Horneman, A. J., Gogarten, J. P., and Graf, J. (2011). Complex evolutionary history of the *Aeromonas veronii* group revealed by host interaction and DNA sequence data. *PLoS One* 6:e16751. doi: 10.1371/journal.pone.0016751

Talagrand-Reboul, E., Roger, F., Kimper, J. L., Colston, S. M., Graf, J., Ltif-Eugenin, F., et al. (2017). Delineation of taxonomic species within complex of species: *Aeromonas media* and related species as a test case. *Front. Microbiol.* 8:621. doi: 10.3389/fmicb.2017.00621

Talavera, G., Dinca, V., and Vila, R. (2013). Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods Ecol. Evol.* 4, 1101–1110. doi: 10.1111/2041-210X.12107

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, C. Q., Humphreys, A. M., Fontaneto, D., and Barraclough, T. (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods Ecol. Evol.* 5, 1086–1094. doi: 10.1111/2041-210X.12246

Vinuesa, P., Silva, C., Werner, D., and Martínez-Romero, E. (2005). Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Mol. Phylogenet. Evol.* 34, 29–54. doi: 10.1016/j.ympev.2004.08.020

Wickham, H. (2016). *ggplot: Elegant Graphics for Data Analysis*, 2nd Edn. New York, NY: Springer-Verlag. doi: 10.1007/978-3-319-24277-4

Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 30, 1720–1728. doi: 10.1093/molbev/mst064

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 Lorén, Farfán and Fusté. 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 June 2017
doi: 10.3389/fmicb.2017.01154

Genome Data Provides High Support for Generic Boundaries in *Burkholderia* Sensu Lato

Chrizelle W. Beukes[1], Marike Palmer[1], Puseletso Manyaka[1], Wai Y. Chan[1], Juanita R. Avontuur[1], Elritha van Zyl[1], Marcel Huntemann[2], Alicia Clum[2], Manoj Pillay[2], Krishnaveni Palaniappan[2], Neha Varghese[2], Natalia Mikhailova[2], Dimitrios Stamatis[2], T. B. K. Reddy[2], Chris Daum[2], Nicole Shapiro[2], Victor Markowitz[2], Natalia Ivanova[2], Nikos Kyrpides[2], Tanja Woyke[2], Jochen Blom[3], William B.Whitman[4], Stephanus N. Venter[1] and Emma T. Steenkamp[1]*

[1] Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa, [2] DOE Joint Genome Institute, Walnut Creek, CA, United States, [3] Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany, [4] Department of Microbiology, University of Georgia, Athens, GA, United States

OPEN ACCESS

Edited by:
Sabela Balboa Méndez, Universidade de Santiago de Compostela, Spain

Reviewed by:
Paulina Estrada De Los Santos, Instituto Politécnico Nacional, Mexico
Radhey S. Gupta, McMaster University, Canada

***Correspondence:**
Stephanus N. Venter
fanus.venter@up.ac.za

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

Received: *20 April 2017*
Accepted: *07 June 2017*
Published: *26 June 2017*

Citation:
Beukes CW, Palmer M, Manyaka P, Chan WY, Avontuur JR, van Zyl E, Huntemann M, Clum A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Stamatis D, Reddy TBK, Daum C, Shapiro N, Markowitz V, Ivanova N, Kyrpides N, Woyke T, Blom J, Whitman WB, Venter SN and Steenkamp ET (2017) Genome Data Provides High Support for Generic Boundaries in Burkholderia Sensu Lato. Front. Microbiol. 8:1154. doi: 10.3389/fmicb.2017.01154

Although the taxonomy of *Burkholderia* has been extensively scrutinized, significant uncertainty remains regarding the generic boundaries and composition of this large and heterogeneous taxon. Here we used the amino acid and nucleotide sequences of 106 conserved proteins from 92 species to infer robust maximum likelihood phylogenies with which to investigate the generic structure of *Burkholderia* sensu lato. These data unambiguously supported five distinct lineages, of which four correspond to *Burkholderia* sensu stricto and the newly introduced genera *Paraburkholderia*, *Caballeronia,* and *Robbsia*. The fifth lineage was represented by *P. rhizoxinica*. Based on these findings, we propose 13 new combinations for those species previously described as members of *Burkholderia* but that form part of *Caballeronia*. These findings also suggest revision of the taxonomic status of *P. rhizoxinica* as it is does not form part of any of the genera currently recognized in *Burkholderia* sensu lato. From a phylogenetic point of view, *Burkholderia* sensu stricto has a sister relationship with the *Caballeronia*+*Paraburkholderia* clade. Also, the lineages represented by *P. rhizoxinica* and *R. andropogonis,* respectively, emerged prior to the radiation of the *Burkholderia* sensu stricto+*Caballeronia*+*Paraburkholderia* clade. Our findings therefore constitute a solid framework, not only for supporting current and future taxonomic decisions, but also for studying the evolution of this assemblage of medically, industrially and agriculturally important species.

Keywords: *Burkholderia*, *Paraburkholderia*, *Caballeronia*, phylogenomics, *Robbsia andropogonis*, *Burkholderia rhizoxinica*

INTRODUCTION

The genus *Burkholderia* was originally introduced to accommodate an assemblage of seven *Pseudomonas* species (Yabuuchi et al., 1992), two of which were later transferred to *Ralstonia* (Gillis et al., 1995; Yabuuchi et al., 1995). Since then, the number of *Burkholderia* species has grown substantially, to about 108 in 2015 (Estrada-de los Santos et al., 2016), spanning a range of

human, animal and plant pathogens, as well as numerous strains with significant biotechnological potential (Depoorter et al., 2016; Estrada-de los Santos et al., 2016). The latter includes the so-called plant beneficial and environmental (PBE) species (Suárez-Moreno et al., 2012), many of which are plant-associated (e.g., those with plant growth promoting activities, the symbiotic diazotrophs and free-living species with diazotrophic, bioremedial and antibiotic activities) (Depoorter et al., 2016; Estrada-de los Santos et al., 2016). Because of this heterogeneity, new genera [e.g., '*Caballeronia*' (Gyaneshwar et al., 2011) and '*Paraburkholderia*' (Sawana et al., 2014)] has been introduced to accommodate most of the PBE species (Oren and Garrity, 2015a,b, 2017), while retaining the pathogens in *Burkholderia* sensu stricto. Most recently, a third genus, *Robbsia* was introduced to accommodate the phytopathogen previously referred to as *B. andropogonis* (Lopes-Santos et al., 2017).

Overall, the taxonomy of *Burkholderia* sensu lato remains in significant flux (Estrada-de los Santos et al., 2016). With their review of the group, Estrada-de los Santos et al. (2016) recognized two monophyletic groups [Groups A and B; A consists of *Caballeronia* and *Paraburkholderia* as circumscribed by Gyaneshwar et al. (2011) and Sawana et al. (2014), respectively, while Group B includes most of the notable human, animal and plant pathogens, as well as the so-called "*B. cepacia* complex"]. They showed that *B. andropogonis* (now *Robbsia andropogonis*) is separated into its own group, and they designated two so-called "Transition Groups" (i.e., 1 and 2; neither were supported as monophyletic and both contained mainly environmental species). Since then, Dobritsa and Samadpour (2016) have proposed the transfer of species in Transition Group 2 to a new genus. However, to complicate the issue, this new genus was named "*Caballeronia*" although its proposed usage is not synonymous with the one previously proposed by Gyaneshwar et al. (2011) for accommodating all the PBE isolates.

The proposals to split *Burkholderia* sensu lato were based almost entirely on evidence from 16S ribosomal RNA (rRNA) phylogenetic trees with limited and in some cases no statistical support (Gyaneshwar et al., 2011; Dobritsa and Samadpour, 2016; Eberl and Vandamme, 2016). Even phylogenies based on conventional multilocus sequence analysis (MLSA) using the combined sequence information for 4–7 genes (Gevers et al., 2005) produced phylogenies in which the major groups were not supported as monophyletic (Estrada-de los Santos et al., 2013). Also, the most comprehensive phylogenetic hypothesis to date (based on 21 conserved gene sequences) lacked sufficient representation across this diverse assemblage (Sawana et al., 2014). Thus, uncertainties remain regarding the genomic and evolutionary coherence of *Burkholderia* sensu lato and its lineages. This, in turn, blurs the boundaries of the *Burkholderia* sensu lato genera currently recognized and also casts doubt on the appropriateness and legitimacy of their taxonomic circumscriptions.

In this study, we aimed to resolve the relationships within *Burkholderia* sensu lato, particularly those pertaining to *Paraburkholderia* and *Caballeronia*, by making use of whole genome sequence data. For this purpose, we utilized all of the sequences for type strains (or appropriate representatives) available in the public domain. To increase representation of the so-called environmental species, we also determined the sequences for eight additional taxa via Phase III of the GEBA (Genomic Encyclopedia of Bacterial and Archaeal type strains) project (Whitman et al., 2015). These included the rhizobial species *P. aspalathi* (Mavengere et al., 2014) and *P. diazotrophica* (Sheu et al., 2013), and the soil bacteria *P. hospita* (Goris et al., 2002), *P. phenazinium* (Viallard et al., 1998), *P. sartisoli* (Vanlaere et al., 2008), *P. terricola* (Goris et al., 2002), as well as the plant-associated diazotrophic species *P. caballeronis* (Martínez-Aguilar et al., 2013) and *P. tropica* (Reis et al., 2004).

MATERIALS AND METHODS

Whole-Genome Sequencing of Eight *Paraburkholderia* Type Strains

The eight type strains (*P. aspalathi* LMG 27731^T, *P. hospita* LMG 20598^T, *P. diazotrophica* LMG 206031^T, *P. phenazinium* LMG 2247^T, *P. sartisoli* LMG 24000^T, *P. terricola* LMG 20594^T, *P. tropica* LMG 22274^T and *P. caballeronis* LMG 26416^T) were obtained from the Belgian Coordinated Collections of Microorganisms (University of Gent, Belgium). Routine growth of these bacteria in the laboratory and extraction of high quality genomic DNA were completed as described previously (Steenkamp et al., 2015). Whole genome sequencing was performed by the Joint Genome Institute (JGI) following standard protocols[1] and using the Illumina HiSeq-2500 1TB platform with an Illumina 300 base pair (bp) insert standard shotgun library.

All raw sequences were filtered using BBDuk (Bushnell, 2017), which removes known Illumina artifacts, and PhiX. Reads with more than one "N" or with quality scores (before trimming) averaging less than 8 or reads shorter than 51 bp (after trimming) were discarded. The remaining reads were mapped to masked versions of human, cat and dog reference sequences using BBMap (Bushnell, 2017) and discarded if identity values exceeded 93%. The remaining reads were then assembled into contigs using Velvet version 1.2.07 (Zerbino and Birney, 2008) (the settings used were velveth: 63 –shortPaired and velvetg: –very clean yes –exportFiltered yes –min contig lgth 500 –scaffolding no –cov cutoff 10). The Velvet contigs were then used to generate 1–3 kbp simulated paired end reads using wgsim version 0.3.0[2] (the settings used were –e 0 –1 100 –2 100 –r 0 R 0 –X 0). We then assembled the quality filtered Illumina reads with the simulated read pairs using Allpaths-LG version r46652 (Gnerre and MacCallum, 2011) (the settings used were PrepareAllpathsInputs: PHRED 64 = 0 PLOIDY = 1 FRAG COVERAGE = 125 JUMP COVERAGE = 25 LONG JUMP COV = 50 and RunAllpathsLG: THREADS = 8 RUN = std shredpairs TARGETS = standard VAPI WARN ONLY = True OVERWRITE = True).

The standard JGI microbial genome annotation pipeline (Huntemann et al., 2015) was used to predict and annotate

[1] http://www.jgi.doe.gov

[2] https://github.com/lh3/wgsim

genes in each of the eight assembled genomes. For this purpose, we specifically used the Prodigal algorithm to identify protein-coding genes (Hyatt et al., 2010). Additional annotation was performed using JGI's Integrated Microbial Genomes (IMG) system (Markowitz et al., 2014).

Sequence Datasets and Multiple Alignments

Protein-coding gene datasets were generated for the eight bacteria sequenced here, as well as all the *Burkholderia* sensu lato type strains (or suitable conspecific strains) for which whole genome sequences were available (Supplementary Table S1). This was achieved by using the EDGAR (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios) server[3] (Blom et al., 2016) to identify single-copy orthologous genes shared among all of the genomes examined. The respective amino acid and nucleotide sequences for each gene dataset were then batch-aligned using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004) iteration-based alignment tool implemented in CLC Main Workbench 7.6 (CLC Bio).

Individual alignments were manually curated in BioEdit version 7.2.5 (Hall, 2011), during which we discarded those genes for which one or more taxa contained more than 5% missing data. The pair-wise protein similarity for the remaining genes (i.e., those for which the datasets were ≥95% complete) were individually determined with Geneious v. 6.1 (Biomatters Limited[4]), followed by concatenation with FASconCAT-G v. 1.02 (Kuck and Longo, 2014). The total pair-wise similarity among the various taxa included in the study was also calculated by making use of the concatenated nucleotide and amino acid datasets using Geneious v. 6.1.

We also evaluated the genomic distribution and functional roles for the genes with ≥95% complete sequence data. The putative function of each gene product was inferred using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases and the GhostKoala mapping tool[5] (Kanehisa et al., 2016), as well as through comparison with the annotated genome of the type species *Burkholderia cepacia* ATCC 25416^T (Yabuuchi et al., 1992). This genome was also used to determine the relative genomic position of each gene used in our dataset. This was done by making use of Geneious v. 6.1 and the publicly available annotations of the ATCC 25416^T genome on the National Center for Biotechnology Information (NCBI[6]) website.

The level of substitution saturation in the various nucleotide and amino acid datasets were evaluated as described before (Palmer et al., 2017). For this purpose, distances based on actual substitutions (p-distance) were compared to those inferred using an appropriate substitution model (Jeffroy et al., 2006; Philippe et al., 2011). The modeled distances for the nucleotide data were inferred using the General Time Reversible (GTR) substitution model (Tavaré, 1986) and the minimum-evolution distance algorithm (Desper and Gascuel, 2002). Both the p- and GTR-distances were determined in DAMBE v. 6.0.1 (Xia and Xie, 2001) and were calculated for the full nucleotide datasets and for the third codon positions only. For the amino acid datasets, MEGA v.6.06 (Tamura et al., 2013) was used to calculate the p-distances and those based on the Jones-Taylor-Thornton (JTT) model (Jones et al., 1992). Graphical representations of the correlation between the respective distances for each dataset were constructed in Microsoft Excel 2013, followed by linear regression analyses.

Phylogenetic Analyses

The respective nucleotide and amino acid alignments for the ≥95% complete protein-coding genes were concatenated and subjected to maximum likelihood phylogenetic analyses with RAxML v. 8.2.1 (Stamatakis, 2014). For this purpose, the sequences were concatenated and partitioned using FASconCAT-G. For the amino acid data, each partition employed the best-fit substitution model as indicated by ProtTest v. 3.4 (Abascal et al., 2005). For the nucleotide data, we used the GTR model with independent parameter estimation for each partition. Branch support was estimated in RAxML using the estimated model parameters, the rapid hill-climbing algorithm and non-parametric bootstrap analyses of 1000 repetitions.

RESULTS

Whole-Genome Sequences for Eight Type Strains of *Paraburkholderia*

Illumina sequencing allowed assembly of high-coverage (i.e., 67.4 to 119.3 X) draft genomes for the type strains of eight *Paraburkholderia* species (**Table 1**). The number of contigs for each genome ranged from 22 to 188 where more than 50% of the individual genomes were incorporated into relatively large contigs (i.e., respective N50-values ranged from 144482 to 573607). The assembled genomes ranged in size from 5.9 for *P. sartisoli* LMG 24000^T to 11.2 million bases for *P. hospita* LMG 20498^T. The number of genes predicted for each genome also corresponded well with their overall sizes (e.g., 5407 genes were predicted for *P. sartisoli* LMG 24000^T and 10534 for *P. hospita* LMG 20498^T). The G+C content for the eight species ranged from 61.09% for *P. aspalathi* to 67.03% for *P. caballeronis*. The assembled genome sequences for all eight species are available from NCBI (see **Table 1** for accession numbers). Overall the sizes and GC content were comparable to previously sequenced genomes of other Burkholderia sensu lato species (**Table 2**).

Sequence Datasets and Multiple Alignments

A set of 106 genes with ≥95% complete sequences were identified among the genomes of 86 *Burkholderia* sensu lato species and the 6 outgroup taxa. The 106 genes were identified using a strict orthology estimation performed in EDGAR (Blom et al., 2016). Only those sequences with a mean % identity of 60.22 (median 54.63%) and a mean Expect(*E*)-value of 6.494625e-09

[3]https://edgar.computational.bio.uni-giessen.de

[4]http://www.geneious.com

[5]www.kegg.jp/blastkoala/

[6]https://www.ncbi.nlm.nih.gov/

TABLE 1 | Details regarding the eight *Paraburkholderia* type strains sequenced at JGI for the GEBA Phase III project.

Type strain[a] (Strain number)	N50 (bp)	Number of contigs	Largest contig size (bp)	Average contig size (bp)	Genome size (bp)	Sequencing depth	G+C (%)	Number of genes	Gold ID	NCBI BioProject Accession
P. hospita (LMG 20598)	141903	188	624080	59577	11200469	73.3X	61.87	10534	Gp0116482	PRJNA323250
P. diazotrophica (LMG 26031)	180974	112	586763	77577	8688577	102.3X	62.58	8041	Gp0116479	PRJNA323247
P. phenazinium (LMG 2247)	342088	52	868734	165344	8597887	67.4X	62.34	7927	Gp0116485	PRJNA323253
P. sartisoli (LMG 24000)	573607	22	1593786	269570	5930529	197.0X	63.52	5407	Gp0116487	PRJNA323254
P. terricola (LMG 20594)	144482	114	380545	64223	7321401	94.4X	63.62	6748	Gp0116489	PRJNA323256
P. tropica (LMG 22274)	341815	61	531529	140659	8580172	119.3X	64.77	7611	Gp0116490	PRJNA323257
P. caballeronis (LMG 26416)	339592	42	770885	168261	7066941	104.1X	67.03	6258	Gp0116481	PRJNA323249
P. aspalathi (LMG 27731)	216058	103	588940	96042	9892286	112.6X	61.09	9038	Gp0116483	PRJNA323251

[a]*LMG = Belgian Coordinated Collections of Microorganisms, Laboratorium voor Microbiologie, Universiteit Gent.*

(median 1.00e-101) of the accepted BLAST hits were included in the final datasets. Although the full set of shared genes among these taxa would be considerably larger, the examined genomes differed substantially in their level of completeness and the annotation approaches utilized. Our conservative approach for generating these datasets therefore attempted to avoid inadvertently including phylogenetic noise caused by potential sequencing and annotation inconsistencies.

The concatenated dataset for the 106 genes consisted of 92 taxa with 25499 residues in the amino acid alignment and 80027 bases in the nucleotide alignment. The amino acid dataset consisted of 99.1% coding characters with 0.9% of the dataset consisting of alignment gaps, while the nucleotide dataset consisted of 98.6% coding characters with 1.4% of the dataset consisting of alignment gaps. Neither of these datasets included any poorly aligned regions because of the absence of more divergent sequences. For example, the amino acid and nucleotide similarity across the entire dataset (including *Ralstonia* and *Cupriavidus* outgroups) were >77% and >73%, respectively (**Figure 1** and Supplementary File S1). Within each of the main phylogenetic clades inferred from the data (see below), these values were generally >92% and >84%, respectively.

Despite the high-level of conservation observed in the 106 genes, both the nucleotide and the amino acid data were free from significant levels of substitution saturation (Supplementary File S2). For both datasets, this was evident from the slope of the linear regression line for the plot between actual and modeled distances. However, compared to the nucleotide dataset, the amino acid dataset was least saturated, as the slope of its regression line was closest to 1. Our results also suggest the limited saturation present in the nucleotide data may be ascribed to multiple substitutions primarily occurring at third codon positions (Supplementary File S2).

We investigated the genomic distribution of the 106 genes by mapping them to those in the annotated genome of strain ATCC 25416^T of *Burkholderia cepacia*, which is also the type species for *Burkholderia* (Yabuuchi et al., 1992). These analyses showed that 101 of the genes mapped to chromosome 1 of this species (Supplementary Figure S1), where they appeared to be scattered throughout the replicon (see Supplementary Table S2 for the nucleotide positions and orientation of the respective genes). The remaining five genes mapped to chromosome 2 (Supplementary Figure S2 and Table S2).

Analysis of the putative functions of the 106 genes revealed that they are likely involved in a multitude of diverse functions. Based on both the original annotations for *B. cepacia* ATCC 25416^T and the KEGG analysis with GhostKOALA, only four of the 106 gene were classified as having unknown or hypothetical functions (Supplementary Table S3). About 44% of the remaining 102 genes represented "informational genes" (*sensu* Jain et al., 1999) and encoded products involved in processes relating to nucleotide synthesis, DNA replication and repair, transcription, translation and related processes. A further 35% of the genes encoded products involved in carbohydrate, lipid and amino acid metabolism, while the remaining 21% encoded products involved in diverse functions (e.g., signal transduction, membrane transport, iron scavenging, etc.) (Supplementary Table S3).

Phylogenetic Analyses

Because of the limited substitution saturation detected in the concatenated amino acid and nucleotide datasets, both datasets were subjected to maximum likelihood phylogenetic analysis in RAxML "as is" (i.e., no attempt was made to exclude saturated sites). However, these analyses were conducted using substitution models specific for each gene, which in all cases accounted for invariable sites and included gamma correction to account for among site rate variation. Although the nucleotide data partitions utilized the GTR model, each partition used independent model parameters (i.e., each gene partition utilized the six nucleotide substitution rates specific to it) (see Supplementary Table S2 for details on the substitution models used for the respective amino acid partitions).

TABLE 2 | Genome properties for all the investigated species forming part of *Burkholderia* sensu stricto, *Caballeronia* and *Paraburkholderia*.

Species[a]	Size (Mb)	G+C (mol%)	Source
P. caledonica	7.3	62	Soil
P. bryophila	7.4	62	Moss
P. kirstenboschensis	8.3	62	Root nodules
P. dilworthii	7.7	62	Root nodules
P. phenoliruptrix	7.8	63	Chemostat
P. graminis	7.5	63	Plant roots
P. terricola	7.3	64	Soil
P. ginsengisoli	6.5	64	Soil
P. monticola	7.9	64	Soil
P. tuberum	9.0	63	Root nodules
P. sprentiae	7.8	63	Root nodules
P. ginsengiterrae	8.5	63	Soil
P. xenovorans	9.7	63	Soil
P. phytofirmans	8.2	62	Plant roots
P. aspalathi	9.9	61	Root nodules
P. fungorum	8.7	62	Fungus
P. phenazinium	8.6	62	Soil
P. sartisoli	5.9	64	Soil
P. diazotrophica	8.7	63	Root nodules
P. phymatum	8.7	62	Root nodules
P. caribensis	9.0	63	Soil
P. terrae	9.9	62	Soil
P. hospita	11.2	62	Soil
P. kururiensis	6.8	64	Water
P. caballeronis	7.1	67	Soil
'*P. acidipaludis*'	6.5	65	Plant stem
P. ferrariae	7.9	65	Iron ore
P. heleia	8.0	65	Plant tissue
P. nodosa	9.5	64	Root nodules
P. mimosarum	8.3	64	Root nodules
P. oxyphila	10.7	64	Soil
P. sacchari	7.3	64	Soil
P. tropica	8.6	65	Plant stem
P. bannensis	8.7	64	Plant roots
C. humi	7.6	63	Soil
C. terrestris	8.2	63	Soil
C. choica	9.8	63	Soil
C. telluris	7.1	64	Soil
'*C. arationis*'	9.4	63	Soil
C. glathei	8.6	64	Soil
C. sordidicola	6.9	60	Fungus
C. udeis	10.1	60	Soil
'*C. concitans*'	6.2	63	Clinical sample
C. grimmiae	6.7	63	Moss
C. zhejiangensis	7.8	63	Wastewater
'*C. fortuita*'	7.4	63	Soil
'*C. temeraria*'	8.3	63	Soil
C. cordobensis	8.2	64	Soil
'*C. hypogeia*'	8.3	63	Soil
'*C. calidae*'	9.6	63	Water
C. megalochromosomata	9.5	63	Soil
C. jiangsuensis	8.6	63	Soil
'*C. pedi*'	9.1	63	Soil
'*C. turbans*'	7.4	63	Clinical sample
'*C. peredens*'	6.7	63	Soil
'*C. ptereochthonis*'	7.7	64	Soil
'*C. glebae*'	7.8	63	Soil
'*C. arvi*'	9.7	62	Soil
'*C. catudaia*'	7.7	63	Soil
B. plantarii	8.1	69	Soil
B. glumae	5.8	68	Plant leaves
B. gladioli	8.8	68	Plant roots
B. mallei	5.8	69	Clinical sample
B. pseudomallei	7.0	68	Clinical sample
B. thailandensis	6.4	68	Soil
B. oklahomensis	7.1	67	Clinical sample
B. stagnalis	7.5	68	Soil
B. ubonensis	6.9	67	Soil
B. dolosa	6.3	67	Clinical sample
B. multivorans	6.2	67	Clinical sample
B. pseudomultivorans	7.9	67	Clinical sample
B. territorii	6.9	66	Water
B. diffusa	6.9	66	Clinical sample
B. latens	6.5	67	Clinical sample
B. vietnamiensis	6.9	67	Soil
B. ambifaria	7.5	67	Soil
B. pyrrocinia	8.0	66	Soil
B. stabilis	8.0	66	Clinical sample
B. cenocepacia	8.1	67	Clinical sample
B. anthina	7.3	67	Soil
B. seminalis	7.6	67	Clinical sample
B. cepacia	8.6	67	Plant tissue
B. contaminans	9.3	66	Veterinary sample
B. lata	8.7	66	Soil

[a]*Species names in inverted commas ('. . .') have not yet been validly published.*

Highly similar and congruent topologies were inferred from the amino acid and nucleotide data for the 106 genes included in this study (**Figure 2** and Supplementary Figure S3). All of the branches in the two trees further received bootstrap support values exceeding 90% (with most supported by values of 100%). The only differences observed between the two trees were in terms of the placement of some species within certain terminal clades (e.g., in the nucleotide phylogeny *P. ginsengisoli* forms a distinct lineage within a larger clade containing *P. caledonica*, *P. bryophila*, *P. kirstenboschensis*, *P. dilworthii*, *P. phenoliruptrix*, *P. graminis*, *P. terricola*, *P. aspalathi*, *P. fungorum*, *P. ginsengiterrae*, *P. phytofirmans*, *P. xenovorans*, *P. monticola*, *P. tuberum,* and *P. sprentiae* but in the amino acid tree it is basal to a smaller clade consisting of *P. monticola*, *P. tuberum,* and *P. sprentiae*). These small topological differences probably reflect limited phylogenetic signal in the datasets for resolving more recent divergences. No disparities were observed regarding the composition of the main clades recovered from the two datasets.

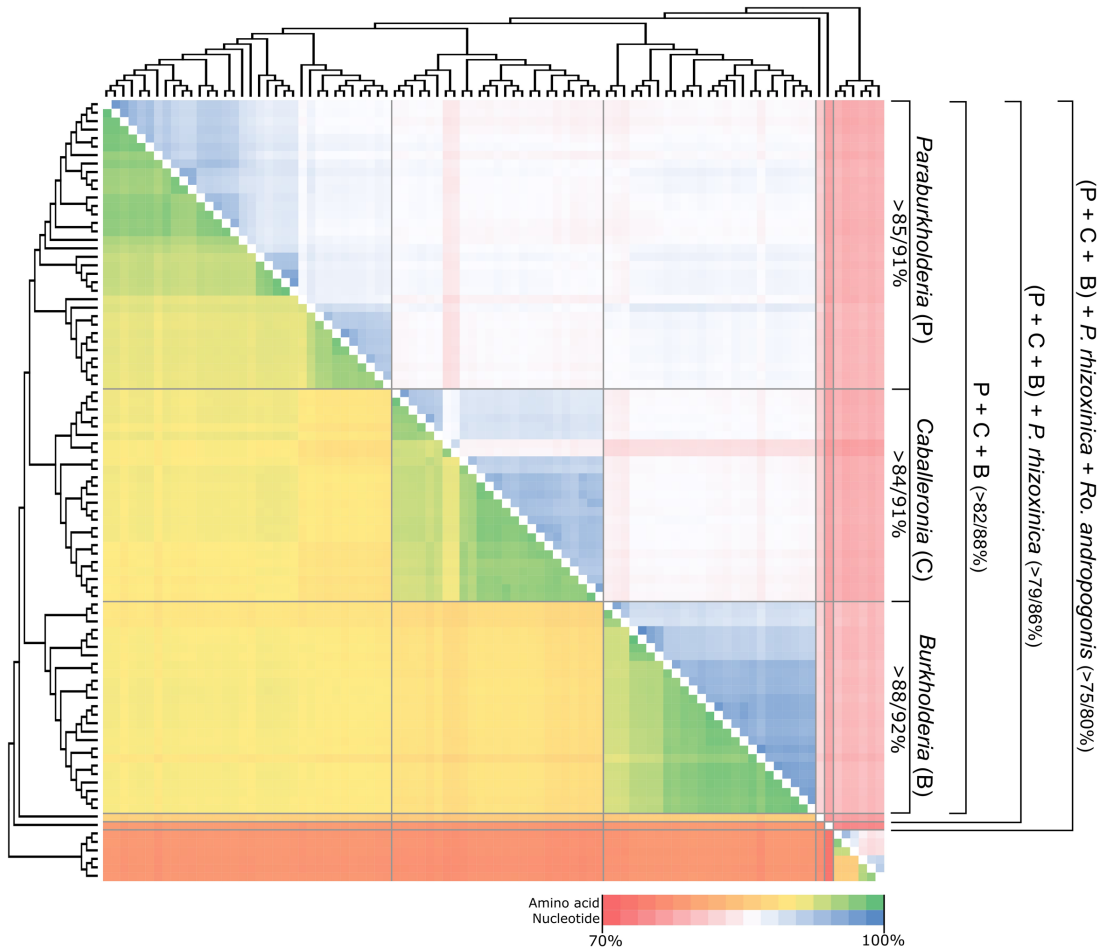


FIGURE 1 | A heat map depicting the sequence similarity of the concatenated sequence of the conserved 106 genes used for phylogenetic analysis. The cladogram indicating the various intra- and intergeneric relationships were inferred from the amino acid based ML topology. Nucleotide similarity values are indicated in the upper triangle of the map, with amino acid similarity values indicated in the lower triangle of the map. A summary of the similarity values for the 5 lineages of interest are indicated for each group (nucleotide/amino acid %), in the panel on the right. For specific values, refer to Supplementary File S1.

In terms of the phylogenetic relationships among the taxa, both trees separated the *Burkholderia* sensu lato species into five distinct lineages (**Figure 1** and Supplementary Figure S3). Three of these corresponded to clades, respectively, representing *Paraburkholderia*, *Caballeronia,* and *Burkholderia* sensu stricto. The remaining two lineages were represented by *R. andropogonis* and *P. rhizoxinica*. Within this phylogeny, *Paraburkholderia* and *Caballeronia* were recovered as sister groups that shared an origin with *Burkholderia* sensu stricto. In turn, these three clades shared a most recent common ancestor with the lineage represented by *P. rhizoxinica*. Based on our analyses, the lineage represented by *R. andropogonis* is the most basal taxon in the *Burkholderia* sensu lato tree.

The *Paraburkholderia* clade consisted of 34 species. Of these, 33 were recently formally transferred to *Paraburkholderia* and the new combinations have been validated. Our data show that the novel combination (suggested by Sawana et al., 2014) requires *P. acidipaludis* still awaits validation. A similar situation exists for the *Caballeronia* clade. Of the 25 species it included, 12 were recently formally transferred to *Caballeronia*, but our results suggest that a further 13 (recently accepted as *Burkholderia* species) also need to be incorporated in this genus (**Table 3**).

DISCUSSION

To achieve our primary goal of resolving the generic boundaries and relationships within *Burkholderia* sensu lato, we endeavored to use as wide a taxon selection as possible. Therefore, to complement the genome data already in the public domain for 78 species in this assemblage, we determined the whole genome sequences for an additional eight PBE species. The genomes for these species exhibited similar characteristics as those of other members of *Burkholderia* sensu lato (see **Table 2**). This was particularly true in terms of genome size and total numbers of genes encoded. Some differences were observed in G+C content. As have been observed before (Gyaneshwar et al.,

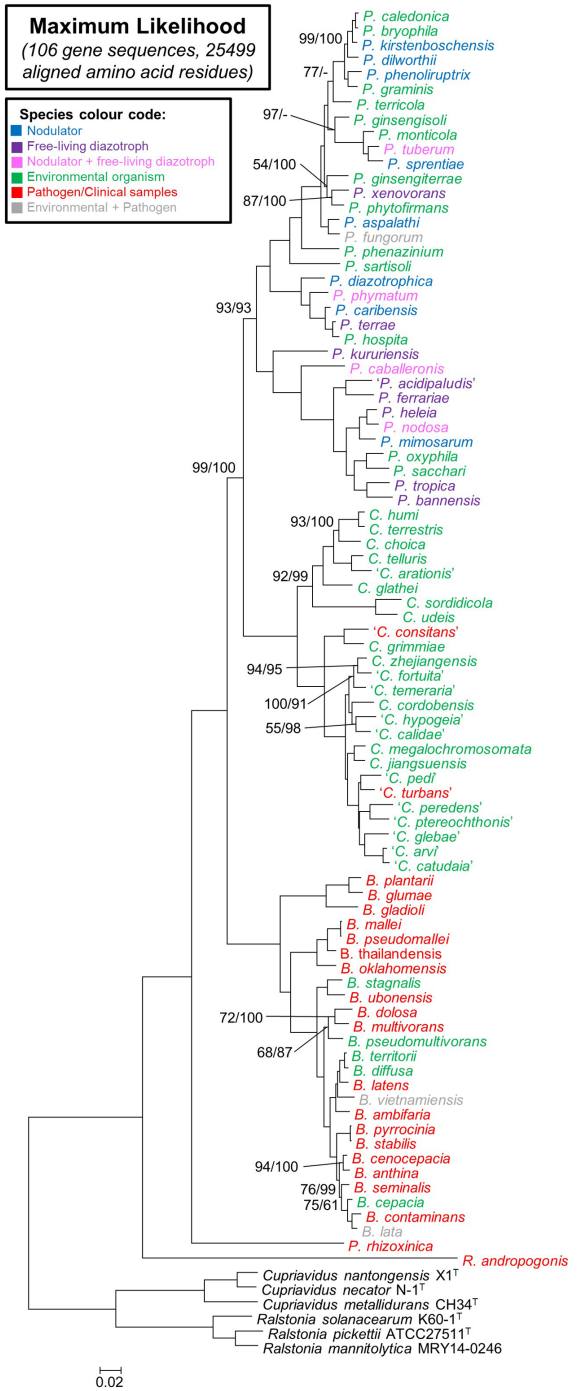


FIGURE 2 | A maximum-likelihood phylogeny of the amino acid sequences of 106 concatenated genes for the 92 strains used in this study. A similar topology was obtained using the nucleotide sequences for these genes (Supplementary Figure S3). New combinations that have not yet been validated are indicated in inverted commas. General species substrates and origins are color coded according to the key provided. The majority of branches received 100% bootstrap in both the amino acid and nucleotide phylogenies and therefore only those branches in which 100% was not calculated for both analyses are indicated. Support is indicated in the order amino acid/nucleotide. The scale bar indicates the number of changes per site.

2011; Estrada-de los Santos et al., 2013; Sawana et al., 2014), the *Burkholderia* sensu stricto genomes were higher in G+C content than *Paraburkholderia* and *Caballeronia*, which were similar in G+C content. Future studies aimed at exploring genome architecture and the functions encoded on these genomes will undoubtedly reveal traits and processes that more clearly characterize the various lineages of this economically important assemblage of bacteria.

For inferring a robust phylogeny that are congruent with the evolutionary history of *Burkholderia* sensu lato, we attempted to avoid or limit the effect of factors known to negatively impact phylogenetic trees (Philippe et al., 2011). The criteria used for generating the respective datasets therefore focused on the use of orthologous loci and on limiting the effects of non-phylogenetic signal. The former was accomplished by using EDGAR to identify orthologous protein-coding genes (Blom et al., 2016). The orthologous nature of a large proportion of the genes included in our final dataset was also congruent with expectations of the so-called complexity hypothesis (Jain et al., 1999; Cohen et al., 2011). *In silico* functional analysis showed that about 44% of these genes represented "informational genes" with products that potentially participate in processes related to DNA replication and repair, transcription and translation. Due to the complexity of their interactions with different proteins and other cellular constituents, these genes are typically less prone to horizontal gene transfer (Jain et al., 1999; Cohen et al., 2011). Our approach for identifying suitable gene sequences from which to infer the phylogeny thus lessened the chances considerably of accidentally using paralogous or xenologous gene copies (Koonin, 2005).

To limit the amount of non-phylogenetic signal in the data, a three-tiered approach was used. [i] The final dataset was large, almost devoid of missing sites (i.e., where genes in some taxa were not sequenced in their entirety) and consisted of the sequences for 106 genes common to *Burkholderia* sensu lato and its *Ralstonia* and *Cupriavidus* outgroups. Such large datasets typically outperform smaller datasets that only contain the sequences for one or a few genes (Daubin et al., 2002; Coenye et al., 2005; Galtier and Daubin, 2008; Bennet et al., 2012; Chan et al., 2012). This is because the "true" phylogenetic signal inherent to orthologs included in such a large dataset will dominate the analysis and typically attenuate or dilute the effects of spurious non-phylogenetic signal associated with one or a few genes (Daubin et al., 2002; Andam and Gogarten, 2011). [ii] Lack of evolutionary independence among loci may contribute to non-phylogenetic signal during tree inference (Gevers et al., 2005). For example, genes that are clustered or whose products are involved in similar or linked processes typically experience similar evolutionary forces, which is accordingly also reflected in their phylogenies (i.e., these reflect the linked evolutionary history of the genes and not the evolutionary history of the species or genus). However, the 106 genes used for resolving *Burkholderia* sensu lato were not significantly clustered (see Supplementary Figures S1, S2), while their inferred products were predicted to participate in diverse functions (see Supplementary Table S3). [iii] Substitution saturation is another important source of non-phylogenetic

TABLE 3 | Summary of the novel combinations proposed for 13 species of *Caballeronia*.

New Combination	Basonym	Type Strain[a]	Reference
Caballeronia arvi comb. nov.	*Burkholderia arvi*	LMG 29317, CCUG68412, MAN34	Peeters et al., 2016
Caballeronia arationis comb. nov.	*Burkholderia arationis*	LMG 29324, CCUG 68405	Peeters et al., 2016
Caballeronia calidae comb. nov.	*Burkholderia calidae*	LMG 29321, CCUG 68408	Peeters et al., 2016
Caballeronia catudaia comb. nov.	*Burkholderia catudaia*	LMG 29318, CCUG 68411	Peeters et al., 2016
Caballeronia concitans comb. nov.	*Burkholderia concitans*	LMG 29315, CCUG 68414, AU12121	Peeters et al., 2016
Caballeronia fortuita comb. nov.	*Burkholderia fortuita*	LMG 29320, CCUG 68409	Peeters et al., 2016
Caballeronia glebae comb. nov.	*Burkholderia glebae*	LMG 29325, CCUG 68404	Peeters et al., 2016
Caballeronia hypogeia comb. nov.	*Burkholderia hypogeia*	LMG 29322, CCUG 68407	Peeters et al., 2016
Caballeronia pedi comb. nov.	*Burkholderia pedi*	LMG 29323, CCUG 68406	Peeters et al., 2016
Caballeronia peredens comb. nov.	*Burkholderia peredens*	LMG 29314, CCUG 68415, NF100	Peeters et al., 2016
Caballeronia ptereochthonis comb. nov.	*Burkholderia ptereochthonis*	LMG 29326, CCUG 68403	Peeters et al., 2016
Caballeronia temeraria comb. nov.	*Burkholderia temeraria*	LMG 29319, CCUG 68410	Peeters et al., 2016
Caballeronia turbans comb. nov.	*Burkholderia turbans*	LMG 29316, CCUG 68413, HI4065	Peeters et al., 2016

[a]*LMG = Belgian Coordinated Collections of Microorganisms, Laboratorium voor Microbiologie, Universiteit Gent; CCUG = Culture Collection, University of Göteborg, Department of Clinical Bacteriology, Institute of Clinical Bacteriology, Immunology, and Virology, University of Göteborg; The strain numbers starting with the abbreviations 'MAN,' 'AU,' 'NF,' and 'HI' are not part of international culture collections.*

signal (Philippe and Forterre, 1999; Xia et al., 2003; Jeffroy et al., 2006; Philippe et al., 2011), and to compensate for its limited occurrence in our datasets, all phylogenetic analyses utilized independent substitution models for each gene partition. This approach proved fairly successful as both the nucleotide and amino acid data supported congruent trees with highly similar topologies.

Our maximum likelihood analyses of the aligned amino acid and nucleotide sequences for 106 genes produced a highly supported phylogeny for *Burkholderia* sensu lato (see **Figure 2**). Most of the branches on this 92-taxon phylogeny received full (100%) bootstrap support. The generation of such a well-resolved phylogeny is, however, not unusual when large datasets containing the information of numerous genes are used. Various previous studies have shown the value of this approach for resolving systematic questions at taxonomic ranks from the genus level and up (e.g., Zhang et al., 2011; Richards et al., 2014; Ormeno-Orrillo et al., 2015; Rahman et al., 2015). Our study thus adds to the growing body of work demonstrating how genome-informed taxonomic decisions represent more robust solutions than those based solely on 16S rRNA or conventional MLSA.

Based on our results, boundaries can for the first time be confidently demarcated for *Burkholderia* sensu stricto, *Caballeronia* and *Paraburkholderia*. These three genera, respectively, represent three of the five distinct lineages recovered among the *Burkholderia* sensu lato species. *Burkholderia* sensu stricto is represented by a large clade that includes the *B. cepacia* complex as well as the *B. pseudomallei* group, and consists primarily of pathogenic species, as suggested previously (Gyaneshwar et al., 2011; Sawana et al., 2014; Estrada-de los Santos et al., 2016). The *Caballeronia* clade includes environmental species that initially formed part of Transition Group 2 of Estrada-de los Santos et al. (2016) and that were transferred to the genus *Caballeronia* by Dobritsa and Samadpour (2016). This clade also includes all 13 of the recently described and validated *Burkholderia glathei*-like species (Oren and Garrity, 2016; Peeters et al., 2016). Based on these findings, we propose the formal inclusion of these species in the genus *Caballeronia* (*sensu* Dobritsa and Samadpour, 2016) (see **Table 3** for details of the proposed new combinations). The inclusion of these taxa into *Caballeronia* raises the number of species to 25. Based on our analyses of their genomes, these species do not encode common *nod* or *nif* and *fix* loci, suggesting that none of the current *Caballeronia* species represent rhizobia or diazotrophs.

The *Paraburkholderia* clade is represented by diverse species, including both free-living and symbiotic diazotrophs, as well as environmental species. Although most of the taxa in this clade have already been formally transferred to *Paraburkholderia* (Sawana et al., 2014) and the novel combinations have been validated (Oren and Garrity, 2015a,b), this genus should also clearly include '*P. acidipaludis*' (Aizawa et al., 2010) isolated from water chestnut as suggested by Sawana et al. (2014). This novel combination, however, still awaits validation. Interestingly, *Paraburkholderia* separates into two fully supported sub-clades, one including at least 23 species (spanning from *P. caledonica* to *P. hospita* in **Figure 2**) and the other including 11 species (*P. kururiensis* to *P. sacchari* in **Figure 2**). Although we could not identify any obvious reason for this split, future studies should explore its possible biological and taxonomic significance.

The two remaining lineages of *Burkholderia* sensu lato is represented by *R. andropogonis* [a pathogen of sorghum (Lopes-Santos et al., 2017)] and *P. rhizoxinica* [a member of Transition Group 1 of Estrada-de los Santos et al. (2016)]. Various previous studies have pointed out that these species should be excluded from *Burkholderia* sensu stricto, *Caballeronia* and/or *Paraburkholderia* (e.g., Estrada-de los Santos et al., 2013, 2016; Dobritsa and Samadpour, 2016). In fact, they have been suggested to represent new genera (Estrada-de los Santos et al., 2013; Dobritsa and Samadpour, 2016). This debate ultimately culminated in the introduction of the new genus

Robbsia to accommodate *R. andropogonis* (Lopes-Santos et al., 2017). Based on our findings, the taxonomy of *P. rhizoxinica* requires similar revision. This species is definitely not a member of *Paraburkholderia* despite having been moved there from *Burkholderia* by Sawana et al. (2014). Both *R. andropogonis* and *P. rhizoxinica* currently represent the only members of their respective lineages for which whole genome sequences are available. Future studies should therefore seek to identify their respective congeneric species [some of which will likely include those in Transition Group 1 (Estrada-de los Santos et al., 2016)] and to understand the biological and evolutionary properties underlying these two lineages.

In addition to allowing unambiguous demarcation of the genera in *Burkholderia* sensu lato, this study also revealed, for the first time, the relationships among these taxa. *Burkholderia* sensu stricto has a well-supported sister group relationship with the clade containing *Caballeronia,* and *Paraburkholderia*. *P. rhizoxinica* is sister to the *Burkholderia* sensu stricto+*Caballeronia*+*Paraburkholderia* clade, while *R. andropogonis* occupies the most basal position in the tree. Knowledge about these relationships could inform hypotheses regarding the biology and evolution of these bacteria, especially in terms of virulence and pathogenicity. For example, *Burkholderia* sensu stricto primarily includes human and animal pathogens, while *P. rhizoxinica* and *Robbsia* are also represented by pathogens (Estrada-de los Santos et al., 2016; Lopes-Santos et al., 2017). Moreover, certain *Caballeronia* and *Paraburkholderia* species have also been isolated from clinical samples [e.g., '*C. consitans*' and '*C. turbans*' (Peeters et al., 2016) and *P. fungorum* (Coenye et al., 2001), and *P. tropica* (Deris et al., 2010), respectively]. The availability of a robust phylogenetic framework for these taxa would thus be invaluable for deciphering the processes and mechanisms involved in the evolution of these species.

DESCRIPTION OF NEW SPECIES COMBINATIONS

Description of *Caballeronia arvi* comb. nov.

Caballeronia arvi (ar'vi. L. gen. n. *arvi* of a field).

Basonym: *Burkholderia arvi* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Analysis of 106 conserved protein-coding sequences have shown that this species is placed in the genus *Caballeronia* with very high support.

The type strain is LMG 29317^T (= CCUG 68412^T = MAN34^T).

Description of *Caballeronia arationis* comb. nov.

Caballeronia arationis (a.ra.ti.o'nis. L. gen. n. *arationis* from a field).

Basonym: *Burkholderia arationis* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Phylogenetic analysis of 106 conserved protein-coding loci clearly showed that there is high support for the placement of this species in *Caballeronia*.

The type strain is LMG 29324^T (=CCUG 68405^T).

Description of *Caballeronia calidae* comb. nov.

Caballeronia calidae (ca'li.dae. L. gen. n. *calidae* from warm water, because this strain was isolated from pond water in a tropical garden).

Basonym: *Burkholderia calidae* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Phylogenetic analysis of 106 conserved protein-coding loci showed (with a high degree of certainty) that this species belongs in the genus *Caballeronia*.

The type strain is LMG 29321^T (=CCUG 68408^T).

Description of *Caballeronia catudaia* comb. nov.

Caballeronia catudaia (ca.tu.da'ia. Gr. adj. *catudaios* subterraneous; N. L. fem. adj. *catudaia*, earth-born).

Basonym: *Burkholderia catudaia* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Our analyses of 106 conserved protein-coding loci clearly indicate that this species has high support for being included in *Caballeronia*.

The type strain is LMG 29318^T (=CCUG 68411^T).

Description of *Caballeronia concitans* comb. nov.

Caballeronia concitans (con.ci'tans. L. fem. part. pres. *concitans* disturbing, upsetting; because the isolation of this bacterium from human sources, including blood, further disturbs the image of this lineage of *Burkholderia* species as benign bacteria).

Basonym: *Burkholderia concitans* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Analysis of 106 conserved protein-coding loci showed that this species has high support for belonging to the genus *Caballeronia*.

The type strain is LMG 29315^T (=CCUG 68414^T = AU12121^T).

Description of *Caballeronia fortuita* comb. nov.

Caballeronia fortuita (for.tu.i'ta. L. fem. adj. *fortuita* accidental, unpremeditated; referring to its fortuitous isolation when searching for *Burkholderia caledonica* endophytes).

Basonym: *Burkholderia fortuita* Peeters et al., 2016.

The description is as described in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci clearly show this species is included in the genus *Caballeronia*.

The type strain is LMG 29320^T (=CCUG 68409^T).

Description of *Caballeronia glebae* comb. nov.

Caballeronia glebae (gle'bae. L. gen. n. *glebae* from a lump or clod of earth, soil).

Basonym: *Burkholderia glebae* Peeters et al., 2016.

The description appears in Peeters et al. (2016). Analysis of 106 conserved protein-coding loci shows high support for the placement of this species in the genus *Caballeronia*.

The type strain is LMG 29325^T (=CCUG 68404^T).

Description of *Caballeronia hypogeia* comb. nov.

Caballeronia hypogeia (hy.po.ge'ia. Gr. adj. *hypogeios* subterraneous; N. L. fem. adj. *hypogeia*, subterraneous, earth-born).

Basonym: *Burkholderia hypogeia* Peeters et al., 2016.

The description appears in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci supports the inclusion of this species into the genus *Caballeronia*.

The type strain is LMG 29322^T (=CCUG 68407^T).

Description of *Caballeronia pedi* comb. nov.

Caballeronia pedi (pe'di. Gr. n. *pedon* soil, earth; N. L. gen. n. *pedi*, from soil).

Basonym: *Burkholderia pedi* Peeters et al., 2016.

The description is listed in Peeters et al. (2016). The analysis of 106 conserved protein-coding loci, showed that this species is placed in *Caballeronia*.

The type strain is LMG 29323^T (=CCUG 68406^T).

Description of *Caballeronia peredens* comb. nov.

Caballeronia peredens (per.e'dens. L. fem. part. pres. *peredens* consuming, devouring; referring to the capacity of this bacterium to degrade fenitrothion).

Basonym: *Burkholderia peredens* Peeters et al., 2016.

The description is as discussed in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci clearly shows that this species should be included in the genus *Caballeronia*.

The type strain is LMG 29314^T (=CCUG 68415^T = NF100^T).

Description of *Caballeronia ptereochthonis* comb. nov.

Caballeronia ptereochthonis (pte.re.o.chtho'nis Gr. n. *pteris* fern; Gr. n. *chthon* soil; N. L. gen. n. *ptereochthonis*, from fern soil).

Basonym: *Burkholderia ptereochthonis* Peeters et al., 2016.

The description appears in Peeters et al. (2016). The analysis of 106 conserved protein-coding loci clearly shows that this species should be included in *Caballeronia*.

The type strain is LMG 29326^T (=CCUG 68403^T).

Description of *Caballeronia temeraria* comb. nov.

Caballeronia temeraria (te.me.ra'ri.a. L. fem. adj. *temeraria* accidental, inconsiderate; referring to its accidental isolation when searching for *Burkholderia caledonica* endophytes).

Basonym: *Burkholderia temeraria* Peeters et al., 2016.

The description of this species appears in Peeters et al. (2016). The analysis of 106 conserved protein-coding loci here, shows that this species is included in *Caballeronia* with high support.

The type strain is LMG 29319^T (=CCUG 68410^T).

Description of *Caballeronia turbans* comb. nov.

Caballeronia turbans (tur'bans. L. fem. part. pres. *turbans* disturbing, agitating, because the isolation of this bacterium from human pleural fluid further disturbs the image of this lineage of *Burkholderia* species as benign bacteria).

Basonym: *Burkholderia turbans* Peeters et al., 2016.

The original species description appears in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci shows that this species forms part of *Caballeronia*.

The type strain is LMG 29316^T (=CCUG 68413^T = HI4065^T).

AUTHOR CONTRIBUTIONS

CB, MPa, PM, SV, and ES: Original concept; analyses; interpretation of results, writing and proofreading. WC, JA, and EvZ: Analyses, writing and proofreading. MH, AC, MPi, KP, NV, NM, DS, TR, CD, NS, VM, NI, NK, and TW: Genome analyses; interpretation of results; proofreading. JB and WW Genome analyses; interpretation of results; proofreading.

FUNDING

We thank the South African National Research Foundation and the Department of Science and Technology for the funding received via the Centre of Excellence programme. The work conducted by the United States Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the United States Department of Energy under Contract No. DE-AC02-05CH11231.

ACKNOWLEDGMENT

We also acknowledge the Bioinformatics and Computational Biology Unit of the University of Pretoria for access to their computational infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01154/full#supplementary-material

REFERENCES

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105. doi: 10.1093/bioinformatics/bti263

Aizawa, T., Ve, N. B., Vijarnsorn, P., Nakajima, M., and Sunairi, M. (2010). *Burkholderia acidipaludis* sp. nov., aluminium-tolerant bacteria isolated from Chinese water chestnut *(Eleocharis dulcis)* growing in highly acidic swamps in South-East Asia. *Int. J. Syst. Evol. Microbiol.* 60, 2036–2041. doi: 10.1099/ijs.0.018283-0

Andam, C. P., and Gogarten, J. P. (2011). Biased gene transfer in microbial evolution. *Nat. Rev. Microbiol.* 9, 543–555. doi: 10.1038/nrmicro2593

Bennet, J. S., Jolley, K. A., Earle, S. G., Corton, C., Bentley, S. D., Parkhill, J., et al. (2012). A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology* 158, 1570–1580. doi: 10.1099/mic.0.056077-0

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44, W22–W28. doi: 10.1093/nar/gkw255

Bushnell, B. (2017). *BBTools Software Package*. Available at: http://sourceforge.net/projects/bbmap

Chan, J. Z.-M., Halachev, M. R., Loman, N. J., Constantinidou, C., and Pallen, M. J. (2012). Defining bacterial species in the genomic era: insights from the genus Acinetobacter. *BMC Microbiology* 12:302. doi: 10.1186/1471-2180-12-302

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P., and Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol. Rev.* 29, 147–167. doi: 10.1016/j.fmrre.2004.11.004

Coenye, T., Laevens, S., Willems, A., Ohlén, M., Hannant, W., Govan, J. R. W., et al. (2001). *Burkholderia fungorum* sp. nov. and *Burkholderia caledonica* sp. nov., two new species isolated from the environment, animals and human clinical samples. *Int. J. Syst. Evol. Microbiol.* 51, 1099–1107. doi: 10.1099/00207713-51-3-1099

Cohen, O., Gophna, U., and Pupko, T. (2011). The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* 28, 1481–1489. doi: 10.1093/molbev/msq333

Daubin, V., Gouy, M., and Perrière, G. (2002). A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 12, 1080–1090. doi: 10.1101/gr.187002

Depoorter, E., Bull, M. J., Peeters, C., Coenye, T., Vandamme, P., and Mahenthiralingam, E. (2016). *Burkholderia*: and update on taxonomy and biotechnological potential as antibiotic producers. *Appl. Microbiol. Biotechnol.* 100, 5215. doi: 10.1007/s00253-016-7520-x

Deris, Z. Z., Van Rostenberghe, H., Habsah, H., Noraida, R., Tan, G. C., Chan, Y. Y., et al. (2010). First isolation of *Burkholderia tropica* from a neonatal patient successfully treated with imipenem. *Int. J. Infect. Dis.* 14, e73–e74. doi: 10.1016/j.ijid.2009.03.005

Desper, R., and Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the Minimum-Evolution Principle. *J. Comput. Biol.* 9, 687–705. doi: 10.1089/106652702761034136

Dobritsa, A. P., and Samadpour, M. (2016). Transfer of eleven *Burkholderia* species to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov., a new genus to accommodate twelve species of *Burkholderia* and *Paraburkholderia*. *Int. J. Syst. Evol. Microbiol.* 66, 2836–2846. doi: 10.1099/ijsem.0.001065

Eberl, L., and Vandamme, P. (2016). Members of the genus *Burkholderia*: good and bad guys. *F1000Res.* 5:1007. doi: 10.12688/f1000research.8221.1

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

Estrada-de los Santos, P., Rojas-Rojas, F. U., Tapia-García, E. Y., Vásquez-Murrieta, M. S., and Hirsch, A. M. (2016). To split or not to split: an opinion on dividing the genus *Burkholderia*. *Ann. Microbiol.* 66, 1303–1314. doi: 10.1007/s13213-015-1183-1

Estrada-de los Santos, P., Vinuesa, P., Martínez-Aguilar, L., Hirsch, A. M., and Caballero-Mellado, J. (2013). Phylogenetic analysis of *Burkholderia* species by multilocus sequence analysis. *Curr. Microbiol.* 67, 51–60. doi: 10.1007/s00284-013-0330-9

Galtier, N., and Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses. *Philos. Tans. R. Soc. B* 363, 4023–4029. doi: 10.1098/rstb.2008.0144

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739. doi: 10.1038/nrmicro1236

Gillis, M., Van Van, T., Bardin, R., Goor, M., Hebbar, P., Willems, A., et al. (1995). Polyphasic taxonomy in the genus *Burkholderia* leading to an emended description of the genus and proposition of *Burkholderia vietnamiensis* sp. nov. for N2-fixing isolates from rice in Vietnam. *Int. J. Syst. Bacteriol.* 45, 274–289. doi: 10.1099/00207713-45-2-274

Gnerre, S., and MacCallum, I. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1513–1518. doi: 10.1073/pnas.1017351108

Goris, J., Dejonghe, W., Falsen, E., De Clerck, E., Geeraerts, B., Willems, A., et al. (2002). Diversity of transconjugants that acquired plasmid pJP4 or pEMT1 after inoculation of a donor strain in the A- and B-horizon of an agricultural soil and description of *Burkholderia* hospita sp. nov. and *Burkholderia terricola* sp. nov. *Syst. Appl. Microbiol.* 25, 340–352. doi: 10.1078/0723-2020-00134

Gyaneshwar, P., Hirsch, A. M., Moulin, L., Chen, W.-M., Elliott, G. N., Bontemps, C., et al. (2011). Legume-nodulating Betaproteobacteria: diversity, host range, and future prospects. *Mol. Plant Microbe Interact.* 24, 1276–1288. doi: 10.1094/MPMI-06-11-0172

Hall, T. (2011). BioEdit: an important software for molecular biology. *GERF Bull Biosci.* 2, 60–61.

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

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Lar-imer, 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

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

Jeffroy, O., Brinkmann, H., Delsuc, F. D. R., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231. doi: 10.1016/j.tig.2006.02.003

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

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

Koonin, E. V. (2005). Orthologs, paralogs and evolutionary genomics. *Annu. Rev. Genet.* 39, 309–338. doi: 10.1146/annurev.genet.39.073003.114725

Kuck, P., and Longo, G. (2014). FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front. Zool.* 11:81. doi: 10.1186/s12983-014-0081-x

Lopes-Santos, L., Castro, D. B. A., Ferreira-Tonin, M., Corrêa, D. B. A., Weir, B. S., Park, D., et al. (2017). Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis* gen. nov., comb. nov. *Antonie Van Leeuwenhoek* 110, 727–736. doi: 10.1007/s10482-017-0842-6

Markowitz, V. M., Chen, I. M., 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

Martínez-Aguilar, L., Salazar-Salazar, C., Méndez, R. D., Caballero-Mellado, J., Hirsch, A. M., Vásquez-Murrieta, M. S., et al. (2013). *Burkholderia caballeronis* sp. nov., a nitrogen fixing species isolated from tomato (*Lycopersicon esculentum*) with the ability to effectively nodulate *Phaseolus vulgaris*. *Antonie Van Leeuwenhoek* 104, 1063–1071. doi: 10.1007/s10482-013-0028-9

Mavengere, N. R., Ellis, A. G., and Le Roux, J. J. (2014). *Burkholderia aspalathi* sp. nov., isolated from root nodules of the South African legume *Aspalathus abietina* Thunb. *Int. J. Syst. Evol. Microbiol.* 64, 1906–1912. doi: 10.1099/ijs.0.057067-0

Oren, A., and Garrity, G. M. (2015a). List of new names and new combinations previously effectively, but not validly, published. Validation List no. 164. *Int. J. Syst. Evol. Microbiol.* 65, 2017–2025. doi: 10.1099/ijs.0.000317

Oren, A., and Garrity, G. M. (2015b). List of new names and new combinations previously effectively, but not validly, published. Validation List no. 165. *Int. J. Syst. Evol. Microbiol.* 66, 2777–2783. doi: 10.1099/ijsem.0.000464

Oren, A., and Garrity, G. M. (2016). List of new names and new combinations previously effectively, but not validly, published. Validation List No. 171. *Int. J. Syst. Evol. Microbiol.* 66, 3761–3764. doi: 10.1099/ijsem.0.001321

Oren, A., and Garrity, G. M. (2017). List of new names and new combinations previously effectively, but not validly, published. Validation List No. 173. *Int. J. Syst. Evol. Microbiol.* 67, 1–3. doi: 10.1099/ijsem.0.001733

Ormeno-Orrillo, E., Servín-Garcidueñas, L. E., Rogel, M. A., González, V., Peralta, H., Mora, J., et al. (2015). Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. *Syst. Appl. Microbiol.* 38, 287–291. doi: 10.1016/j.syapm.2014.12.002

Palmer, M., Steenkamp, E. T., Coetzee, M. P. A., Chan, W.-Y., van Zyl, E., De Maayer, P., et al. (2017). Phylogenomic resolution of the bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella*. *Antonie Van Leeuwenhoek* doi: 10.1007/s10482-017-0852-4

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S., and Vandamme, P. (2016). Phylogenomic study of *Burkholderia glathei*-like

organisms, proposal of 13 novel *Burkholderia* species and emended descriptions of *Burkholderia sordidicola*, *Burkholderia zhejiangensis*, and *Burkholderia grimmiae*. *Front. Microbiol.* 7:877. doi: 10.3389/fmicb.2016.00877

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., et al. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:e1000602. doi: 10.1371/journal.pbio.1000602

Philippe, H., and Forterre, P. (1999). The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* 49, 509–523. doi: 10.1007/PL00006573

Rahman, N. A., Parks, D. H., Vanwonterghem, I., Morrison, M., Tyson, G. W., and Hugenholtz, P. (2015). A phylogenomic analysis of the bacterial Phylum Fibrobacteres. *Front. Microbiol.* 6:1469. doi: 10.3389/fmicb.2015.01469

Reis, V. M., Estrada-de los Santos, P., Tenorio-Salgado, S., Vogel, J., Stoffels, M., Guyon, S., et al. (2004). *Burkholderia tropica* sp. nov., a novel nitrogen-fixing, plant-associated bacterium. *Int. J. Syst. Evol. Microbiol.* 54, 2155–2162. doi: 10.1099/ijs.0.02879-0

Richards, V. P., Palmer, S. R., Bitar, P. D. P., Qin, X., Weinstock, G. M., Highlander, S. K., et al. (2014). Phylogenomics and the dynamic genome evolution of the genus Streptococcus. *Genome Boil. Evol.* 6, 741–753. doi: 10.1093/gbe/evu048

Sawana, A., Adeolu, M., and Gupta, R. S. (2014). Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harbouring environmental species. *Front. Genet.* 5:429. doi: 10.3389/fgene.2014.00429

Sheu, S.-Y., Chou, J.-H., Bontemps, C., Elliott, G. N., Gross, E., dos Reis Junior, F. B., et al. (2013). *Burkholderia diazotrophica* sp. nov., isolated from root nodules of *Mimosa* spp. *Int. J. Syst. Evol. Microbiol.* 63, 435–441. doi: 10.1099/ijs.0.039859-0

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

Steenkamp, E. T., van Zyl, E., Beukes, C. W., Avontuur, J. R., Chan, W. Y., Palmer, M., et al. (2015). *Burkholderia kirstenboschensis* sp. nov. nodulates papilionoid legumes indigenous to South Africa. *Syst. Appl. Microbiol.* 38, 545–554. doi: 10.1016/j.syapm.2015.09.003

Suárez-Moreno, Z. R., Caballero-Mellado, J., Coutinho, B. G., Mendonça-Previato, L., James, E. K., and Venturi, V. (2012). Common features of environmental and potentially beneficial plant-associated *Burkholderia*. *Microb. Ecol.* 63, 249–266. doi: 10.1007/s00248-011-9929-1

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

Tavaré, S. (1986). "Some probabilistic and statistical problems in the analysis of DNA sequences," in *Lectures on Mathematics in the Life Sciences*, Vol. 17ed. R. M. Miura (Providence: American Mathematical Society), 57–86.

Vanlaere, E., van der Meer, J. R., Falsen, E., Salles, J. F., de Brandt, E., and Vandamme, P. (2008). *Burkholderia sartisoli* sp. nov., isolated from a polycyclic aromatic hydrocarbon-contaminated soil. *Int. J. Syst. Evol. Microbiol.* 58, 420–423. doi: 10.1099/ijs.0.65451-0

Viallard, V., Poirier, I., Cournoyer, B., Haurat, J., Wiebkin, S., Ophel-Keller, K., et al. (1998). *Burkholderia graminis* sp. nov., a rhizospheric *Burkholderia* species, and reassessment of [*Pseudomonas*] *phenazinium*, [*Pseudomonas*] *pyrrocinia* and [*Pseudomonas*] *glathei* as *Burkholderia*. *Int. J. Syst. Bacteriol.* 48, 549–563. doi: 10.1099/00207713-48-2-549

Whitman, W. B., Woyke, T., Klenk, H.-P., Zhou, Y., Lilburn, T. G., Beck, B. J., et al. (2015). Genomic encyclopedia of bacterial and archaeal type strains, phase III: the genomes of soil and plant-associated and newly described type strains. *Stand. Genomic Sci.* 10:26. doi: 10.1186/s40793-015-0017-x

Xia, X., and Xie, Z. (2001). DAMBE: Software package for data analysis in molecular biology and evolution. *J. Hered.* 92, 371–373. doi: 10.1093/jhered/92.4.371

Xia, X., Xie, Z., Salemi, M., Chen, L., and Wang, Y. (2003). An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* 26, 1–7. doi: 10.1016/S1055-7903(02)00326-3

Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., et al. (1992). Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiol. Immunol.* 36, 1251–1275. doi: 10.1111/j.1348-0421.1992.tb02129.x

Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H., and Nishiuchi, Y. (1995). Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. nov.: Proposal of *Ralstonia pickettii (*Ralston, Palleroni and Doudoroff 1973) comb. nov., *Ralstonia solanacearum* (Smith 1896) comb. nov. and *Ralstonia eutropha* (Davis 1969) comb. nov. *Microbiol. Immunol.* 39, 897–904. doi: 10.111/j.1348-0421.1995.tb03275.x

Zerbino, D., 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

Zhang, Z. G., Ye, Z. Q., Yu, L., and Shi, P. (2011). Phylogenomic reconstruction of lactic acid bacteria: an update. *BMC Evol. Biol.* 11:1. doi: 10.1186/1471-2148-11-1

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 Beukes, Palmer, Manyaka, Chan, Avontuur, van Zyl, Huntemann, Clum, Pillay, Palaniappan, Varghese, Mikhailova, Stamatis, Reddy, Daum, Shapiro, Markowitz, Ivanova, Kyrpides, Woyke, Blom, Whitman, Venter and Steenkamp. 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

CORRECTION
published: 02 March 2018
doi: 10.3389/fmicb.2018.00373

Corrigendum: Genome Data Provides High Support for Generic Boundaries in *Burkholderia* Sensu Lato

Chrizelle W. Beukes[1], Marike Palmer[1], Puseletso Manyaka[1], Wai Y. Chan[1], Juanita R. Avontuur[1], Elritha van Zyl[1], Marcel Huntemann[2], Alicia Clum[2], Manoj Pillay[2], Krishnaveni Palaniappan[2], Neha Varghese[2], Natalia Mikhailova[2], Dimitrios Stamatis[2], T. B. K. Reddy[2], Chris Daum[2], Nicole Shapiro[2], Victor Markowitz[2], Natalia Ivanova[2], Nikos Kyrpides[2], Tanja Woyke[2], Jochen Blom[3], William B. Whitman[4], Stephanus N. Venter[1] and Emma T. Steenkamp[1]*

[1] Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa, [2] DOE Joint Genome Institute, Walnut Creek, CA, United States, [3] Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany, [4] Department of Microbiology, University of Georgia, Athens, GA, United States

OPEN ACCESS

Edited and reviewed by:
Frontiers in Microbiology Editorial Office, Frontiers, Switzerland

***Correspondence:**
Stephanus N. Venter
fanus.venter@up.ac.za

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

Received: 13 February 2018
Accepted: 19 February 2018
Published: 02 March 2018

Citation:
Beukes CW, Palmer M, Manyaka P, Chan WY, Avontuur JR, van Zyl E, Huntemann M, Clum A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Stamatis D, Reddy TBK, Daum C, Shapiro N, Markowitz V, Ivanova N, Kyrpides N, Woyke T, Blom J, Whitman WB, Venter SN and Steenkamp ET (2018) Corrigendum: Genome Data Provides High Support for Generic Boundaries in Burkholderia Sensu Lato. Front. Microbiol. 9:373. doi: 10.3389/fmicb.2018.00373

Keywords: *Burkholderia*, *Paraburkholderia*, *Caballeronia*, phylogenomics, *Robbsia andropogonis*, *Burkholderia rhizoxinica*

A corrigendum on

Genome Data Provides High Support for Generic Boundaries in *Burkholderia* Sensu Lato
by Beukes, C. W., Palmer, M., Manyaka, P., Chan, W. Y., Avontuur, J. R., van Zyl, E., et al. (2017) Front. Microbiol. 8:1154. doi: 10.3389/fmicb.2017.01154

In the original article, there was a spelling mistake in a species name contained in **Table 3** as published. As the information in this table is required for validation of the novel species combinations, this mistake has been corrected from "*Caballeronia ptereocthonis* comb. nov." to "*Caballeronia ptereochthonis* comb. nov." and "*Burkholderia ptereocthonis*" to "*Burkholderia ptereochthonis*". The corrected **Table 3** appears below.

Also, an additional section titled "Description of New Species Combinations" was added to the original article. The section contains a short protologue for each proposed novel combination to conform to the rules of the Bacterial Code of Nomenclature. The protologues appear below.

The authors apologize for this error. The original article has been updated.

DESCRIPTION OF NEW SPECIES COMBINATIONS

Description of *Caballeronia arvi* comb. nov.

Caballeronia arvi (ar'vi. L. gen. n. *arvi* of a field).

Basonym: *Burkholderia arvi* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Analysis of 106 conserved protein-coding sequences have shown that this species is placed in the genus *Caballeronia* with very high support.

The type strain is LMG 29317^T (= CCUG 68412^T = MAN34^T).

TABLE 3 | Summary of the novel combinations proposed for 13 species of *Caballeronia*.

New Combination	Basonym	Type Strain[a]	Reference
Caballeronia arvi comb. nov.	*Burkholderia arvi*	LMG 29317, CCUG68412, MAN34	Peeters et al., 2016
Caballeronia arationis comb. nov.	*Burkholderia arationis*	LMG 29324, CCUG 68405	Peeters et al., 2016
Caballeronia calidae comb. nov.	*Burkholderia calidae*	LMG 29321, CCUG 68408	Peeters et al., 2016
Caballeronia catudaia comb. nov.	*Burkholderia catudaia*	LMG 29318, CCUG 68411	Peeters et al., 2016
Caballeronia concitans comb. nov.	*Burkholderia concitans*	LMG 29315, CCUG 68414, AU12121	Peeters et al., 2016
Caballeronia fortuita comb. nov.	*Burkholderia fortuita*	LMG 29320, CCUG 68409	Peeters et al., 2016
Caballeronia glebae comb. nov.	*Burkholderia glebae*	LMG 29325, CCUG 68404	Peeters et al., 2016
Caballeronia hypogeia comb. nov.	*Burkholderia hypogeia*	LMG 29322, CCUG 68407	Peeters et al., 2016
Caballeronia pedi comb. nov.	*Burkholderia pedi*	LMG 29323, CCUG 68406	Peeters et al., 2016
Caballeronia peredens comb. nov.	*Burkholderia peredens*	LMG 29314, CCUG 68415, NF100	Peeters et al., 2016
Caballeronia ptereochthonis comb. nov.	*Burkholderia ptereochthonis*	LMG 29326, CCUG 68403	Peeters et al., 2016
Caballeronia temeraria comb. nov.	*Burkholderia temeraria*	LMG 29319, CCUG 68410	Peeters et al., 2016
Caballeronia turbans comb. nov.	*Burkholderia turbans*	LMG 29316, CCUG 68413, HI4065	Peeters et al., 2016

[a] *LMG = Belgian Coordinated Collections of Microorganisms, Laboratorium voor Microbiologie, Universiteit Gent; CCUG = Culture Collection, University of Göteborg, Department of Clinical Bacteriology, Institute of Clinical Bacteriology, Immunology, and Virology, University of Göteborg; The strain numbers starting with the abbreviations 'MAN,' 'AU,' 'NF,' and 'HI' are not part of international culture collections.*

Description of *Caballeronia arationis* comb. nov.

Caballeronia arationis (a.ra.ti.o'nis. L. gen. n. *arationis* from a field).

Basonym: *Burkholderia arationis* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Phylogenetic analysis of 106 conserved protein-coding loci clearly showed that there is high support for the placement of this species in *Caballeronia*.

The type strain is LMG 29324^T (=CCUG 68405^T).

Description of *Caballeronia calidae* comb. nov.

Caballeronia calidae (ca'li.dae. L. gen. n. *calidae* from warm water, because this strain was isolated from pond water in a tropical garden).

Basonym: *Burkholderia calidae* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Phylogenetic analysis of 106 conserved protein-coding loci showed (with a high degree of certainty) that this species belongs in the genus *Caballeronia*.

The type strain is LMG 29321^T (=CCUG 68408^T).

Description of *Caballeronia catudaia* comb. nov.

Caballeronia catudaia (ca.tu.da'ia. Gr. adj. *catudaios* subterraneous; N. L. fem. adj. *catudaia*, earth-born).

Basonym: *Burkholderia catudaia* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Our analyses of 106 conserved protein-coding loci clearly indicate that this species has high support for being included in *Caballeronia*.

The type strain is LMG 29318^T (=CCUG 68411^T).

Description of *Caballeronia concitans* comb. nov.

Caballeronia concitans (con.ci'tans. L. fem. part. pres. *concitans* disturbing, upsetting; because the isolation of this bacterium from human sources, including blood, further disturbs the image of this lineage of *Burkholderia* species as benign bacteria).

Basonym: *Burkholderia concitans* Peeters et al., 2016.

The description is as provided in Peeters et al. (2016). Analysis of 106 conserved protein-coding loci showed that this species has high support for belonging to the genus *Caballeronia*.

The type strain is LMG 29315^T (=CCUG 68414^T = AU12121^T).

Description of *Caballeronia fortuita* comb. nov.

Caballeronia fortuita (for.tu.i'ta. L. fem. adj. *fortuita* accidental, unpremeditated; referring to its fortuitous isolation when searching for *Burkholderia caledonica* endophytes).

Basonym: *Burkholderia fortuita* Peeters et al., 2016.

The description is as described in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci clearly show this species is included in the genus *Caballeronia*.

The type strain is LMG 29320^T (=CCUG 68409^T).

Description of *Caballeronia glebae* comb. nov.

Caballeronia glebae (gle'bae. L. gen. n. *glebae* from a lump or clod of earth, soil).

Basonym: *Burkholderia glebae* Peeters et al., 2016.

The description appears in Peeters et al. (2016). Analysis of 106 conserved protein-coding loci shows high support for the placement of this species in the genus *Caballeronia*.

The type strain is LMG 29325^T (=CCUG 68404^T).

Description of *Caballeronia hypogeia* comb. nov.

Caballeronia hypogeia (hy.po.ge'ia. Gr. adj. *hypogeios* subterraneous; N. L. fem. adj. *hypogeia*, subterraneous, earth-born).

Basonym: *Burkholderia hypogeia* Peeters et al., 2016.

The description appears in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci supports the inclusion of this species into the genus *Caballeronia*.

The type strain is LMG 29322^T (=CCUG 68407^T).

Description of *Caballeronia pedi* comb. nov.

Caballeronia pedi (pe'di. Gr. n. *pedon* soil, earth; N. L. gen. n. *pedi*, from soil).

Basonym: *Burkholderia pedi* Peeters et al. 2016.

The description is listed in Peeters et al. (2016). The analysis of 106 conserved protein-coding loci, showed that this species is placed in *Caballeronia*.

The type strain is LMG 29323^T (=CCUG 68406^T).

Description of *Caballeronia peredens* comb. nov.

Caballeronia peredens (per.e'dens. L. fem. part. pres. *peredens* consuming, devouring; referring to the capacity of this bacterium to degrade fenitrothion).

Basonym: *Burkholderia peredens* Peeters et al., 2016.

The description is as discussed in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci clearly shows that this species should be included in the genus *Caballeronia*.

The type strain is LMG 29314^T (=CCUG 68415^T = NF100^T).

Description of *Caballeronia ptereochthonis* comb. nov.

Caballeronia ptereochthonis (pte.re.o.chtho'nis Gr. n. *pteris* fern; Gr. n. *chthon* soil; N. L. gen. n. *ptereochthonis*, from fern soil).

Basonym: *Burkholderia ptereochthonis* Peeters et al., 2016.

The description appears in Peeters et al. (2016). The analysis of 106 conserved protein-coding loci clearly shows that this species should be included in *Caballeronia*.

The type strain is LMG 29326^T (=CCUG 68403^T).

Description of *Caballeronia temeraria* comb. nov.

Caballeronia temeraria (te.me.ra'ri.a. L. fem. adj. *temeraria* accidental, inconsiderate; referring to its accidental isolation when searching for *Burkholderia caledonica* endophytes).

Basonym: *Burkholderia temeraria* Peeters et al., 2016.

The description of this species appears in Peeters et al. (2016). The analysis of 106 conserved protein-coding loci here, shows that this species is included in *Caballeronia* with high support.

The type strain is LMG 29319^T (=CCUG 68410^T).

Description of *Caballeronia turbans* comb. nov.

Caballeronia turbans (tur'bans. L. fem. part. pres. *turbans* disturbing, agitating, because the isolation of this bacterium from human pleural fluid further disturbs the image of this lineage of *Burkholderia* species as benign bacteria).

Basonym: *Burkholderia turbans* Peeters et al., 2016.

The original species description appears in Peeters et al. (2016). Our analysis of 106 conserved protein-coding loci shows that this species forms part of *Caballeronia*.

The type strain is LMG 29316^T (=CCUG 68413^T = HI4065^T).

REFERENCES

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S., and Vandamme, P. (2016). Phylogenomic study of Burkholderia glathei-like organisms, proposal of 13 novel *Burkholderia* species and emended descriptions of *Burkholderia sordidicola, Burkholderia zhejiangensis,* and *Burkholderia grimmiae. Front. Microbiol.* 7:877. doi: 10.3389/fmicb.2016.00877

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 Beukes, Palmer, Manyaka, Chan, Avontuur, van Zyl, Huntemann, Clum, Pillay, Palaniappan, Varghese, Mikhailova, Stamatis, Reddy, Daum, Shapiro, Markowitz, Ivanova, Kyrpides, Woyke, Blom, Whitman, Venter and Steenkamp. 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: 12 July 2017
doi: 10.3389/fmicb.2017.01296

In-silico Taxonomic Classification of 373 Genomes Reveals Species Misidentification and New Genospecies within the Genus *Pseudomonas*

Phuong N. Tran[1,2], *Michael A. Savka*[3] *and Han Ming Gan*[1,2,4]*

[1] Genomics Facility, Tropical and Medicine Biology Platform, Monash University Malaysia, Bandar Sunway, Malaysia, [2] School of Science, Monash University Malaysia, Bandar Sunway, Malaysia, [3] Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY, United States, [4] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Geelong, VIC, Australia

OPEN ACCESS

Edited by:
Jesus L. Romalde,
Universidade de Santiago de Compostela, Spain

Reviewed by:
Siomar De Castro Soares,
Universidade Federal do Triângulo Mineiro, Brazil
Se-Ran Jun,
University of Arkansas for Medical Sciences, United States

***Correspondence:**
Han Ming Gan
han.gan@deakin.edu.au

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

Received: 26 April 2017
Accepted: 27 June 2017
Published: 12 July 2017

Citation:
Tran PN, Savka MA and Gan HM (2017) In-silico Taxonomic Classification of 373 Genomes Reveals Species Misidentification and New Genospecies within the Genus Pseudomonas. Front. Microbiol. 8:1296. doi: 10.3389/fmicb.2017.01296

The genus *Pseudomonas* has one of the largest diversity of species within the Bacteria kingdom. To date, its taxonomy is still being revised and updated. Due to the non-standardized procedure and ambiguous thresholds at species level, largely based on 16S rRNA gene or conventional biochemical assay, species identification of publicly available *Pseudomonas* genomes remains questionable. In this study, we performed a large-scale analysis of all *Pseudomonas* genomes with species designation (excluding the well-defined *P. aeruginosa*) and re-evaluated their taxonomic assignment via *in silico* genome-genome hybridization and/or genetic comparison with valid type species. Three-hundred and seventy-three pseudomonad genomes were analyzed and subsequently clustered into 145 distinct genospecies. We detected 207 erroneous labels and corrected 43 to the proper species based on Average Nucleotide Identity Multilocus Sequence Typing (MLST) sequence similarity to the type strain. Surprisingly, more than half of the genomes initially designated as *Pseudomonas syringae* and *Pseudomonas fluorescens* should be classified either to a previously described species or to a new genospecies. Notably, high pairwise average nucleotide identity (>95%) indicating species-level similarity was observed between *P. synxantha-P. libanensis*, *P. psychrotolerans–P. oryzihabitans,* and *P. kilonensis- P. brassicacearum*, that were previously differentiated based on conventional biochemical tests and/or genome-genome hybridization techniques.

Keywords: *Pseudomonas*, taxonomy, phylogenomics, *in-silico* genome-genome hybridization, comparative genomics, average nucleotide identity, genospecies

INTRODUCTION

The genus *Pseudomonas* is considerably diverse and consists of more than 100 characterized species to date (Mulet et al., 2013). Some species of *Pseudomonas* are well-known and characterized such as *P. aeruginosa* (Oliver et al., 2000). Some are plant-associated such as *Pseudomonas fluorescens* and *Pseudomonas syringae* and define or are present in numerous plant-microbe system

(Pieterse et al., 2014). *Pseudomonas fluorescens* in general, is known for its ability to colonize plant rhizospheres and produce antimicrobial compounds targeting pathogens; thus, protecting plants from diseases (Hol et al., 2013). On the other hand, *P. syringae* is commonly associated with plant disease and has been known to invade a variety of plant species with its extracellular flagella and pili appendages. Infection with *P. syringae* can also predispose the host plant to environmental stresses such as frost damage (Maki et al., 1974).

The genetic threshold level for bacterial species definition has seen various changes in the past 20 years due to the development and availability of new bioinformatics tools and genetic resources. For example, Hagström et al. (2002) reported a threshold of 97% or lower in homology of the 16S rRNA DNA sequence to be sufficient to characterize two bacteria as different species. Later this value was raised and is now more widely accepted at 99% (Buckley and Roberts, 2007). However, 16S rRNA has also been previously suggested to be efficient at delineating bacterial strains to a genera but not for species identification (Moore et al., 1996; Anzai et al., 2000; Yamamoto et al., 2000).

Based on pairwise comparisons of amino acid identity (AAI) with cutoff of 95% such that members within one genomic cluster have AAI >95%, Jun et al. (2016) reported nine major groups corresponding to the major *Pseudomonas* species groups including seven well-described species *P. aeruginosa, P. fluorescens, P. syringae, P. putida, P. stuzeri, P. fragi*, *P. fusovaginae* and two mixtures of unidentified species. *P. aeruginosa* contributes to more than half of the 1,073 genomes used in the study and forms a single well delineated genomic cluster suggesting that it is well-characterized. On the contrary, 29 genomes deposited as *P. fluorescens* in public database were found in 20 different genomic clusters indicating potential mislabeling of these genomes. In addition, many suggestions for redefinition of some *Pseudomonas* species such as *P. avellanae* and *P. putida* have been raised (Jun et al., 2016).

In this study, we inferred the phylogenomic placement of 373 *Pseudomonas* genomes identified to the species level representing 79 unique species and evaluated their species validity based on *in silico* genome-genome hybridization. The re-classification of *Pseudomonas* strains based on whole genome sequences will assist future comparative genomics analysis study and more importantly highlights the need for a more robust classification of *Pseudomonas* strains in the future especially with the availability of new genomic resources.

MATERIALS AND METHODS

Datasets

Whole genome sequences of *Pseudomonas* strains with species designation excluding those of *P. aeruginosa* were downloaded from the NCBI FTP server (ftp.ncbi.nih.gov) in February, 2016. All genomes were filtered for assemblies with contig number no greater than 300 to avoid the inclusion of overly fragmented genome into the analysis. In addition, ten *Acinetobacter*, four *Cellvibrio* species and two *Azotobacter* species whole genomes were also included as the outgroup for phylogenomic analysis.

Phylogenomic Inference

Whole proteome was predicted using Prodigal V2.6.2 (Hyatt et al., 2010) and piped into PhyloPhlAn 0.99 (Segata et al., 2013) to construct a maximum likelihood tree using FastTree2 (–bionj –slownni –mlacc 2 –pseudo –spr 4 setting) (Price et al., 2010) based on the identification, alignment and concatenation of 400 universally conserved proteins (Edgar, 2004, 2010). Local branch support values were computed by FastTree2 using the Shimodaira-Hasegawa test. The tree was subsequently visualized and manually collapsed based on genospecies (**Supplemental Table 1**) using MEGA6 (Tamura et al., 2013).

Average Nucleotide Identity (ANI) Calculation

Pairwise ANIm was calculated with Jspecies v1.2.1 using the standard MUMmer algorithm (Richter and Rosselló-Móra, 2009). To reduce computational calculation, we separated the ANIm calculation into 4 groups based on phylogenomic clustering e.g., Clades 1, 2, 3 and strains that are not part of the labeled clade (**Figure 1**). The matrix obtained from Jspecies for each major group was clustered and visualized using Rstudio 0.99.902, pheatmap package.

Housekeeping Gene Similarity Calculation

Nucleotide similarity search on housekeeping genes (*rpoB, recA, rpoD,* and/or *gyrB*) was performed with basic local alignment search tool (BLASTN) using an *e*-value cutoff of $1 \times e^{-10}$ (Altschul et al., 1990).

RESULTS AND DISCUSSION

High Genomic Diversity among *Pseudomonas* Strains

A total of 373 *Pseudomonas* genome sequences were selected for phylogenomic tree reconstruction. The genome size among *Pseudomonas* strains is highly variable, ranging from 3,022,325 bp (*P. caeni* DSM 24390) to 8,608,769 bp (*P. bauzanesis* W13Z2). Similarly, the GC content of their genomes ranges from 56.51% (*P. endophytica* BSTT44) to 67.43% (*P. syringae pv tomato* T1). Based on current classification, the most commonly deposited non-*P. aeruginosa* pseudomonad species is *P. syringae* ($N = 67$), followed by *P. fluorescens* ($N = 62$), *P. putida* ($N = 28$), *P. stuzeri* ($N = 17$), *P. amygdali* ($N = 15$), *P. psychrotolerans* ($N = 14$), *P. savastanoi* ($N = 13$), and *P. denitrificans* ($N = 11$). Out of a total of 79 species outcomes, surprisingly 43 species have only one whole genome sequence representative in the database.

Maximum likelihood tree as inferred from 400 universal conserved proteins clustered the *Pseudomonas* strains into several major clades with maximal SH-like branch support (**Figure 1** and **Supplemental Figure 1**). Support values of some inner clades are low suggesting that shallow relationships between strains within the same genospecies are not resolved, presumably due to the lack of divergence at the amino acid level among closely related strains. A wgMLST analysis restricted to strains from closely related *Pseudomonas* species will likely identify more conserved loci which can then be used to refine

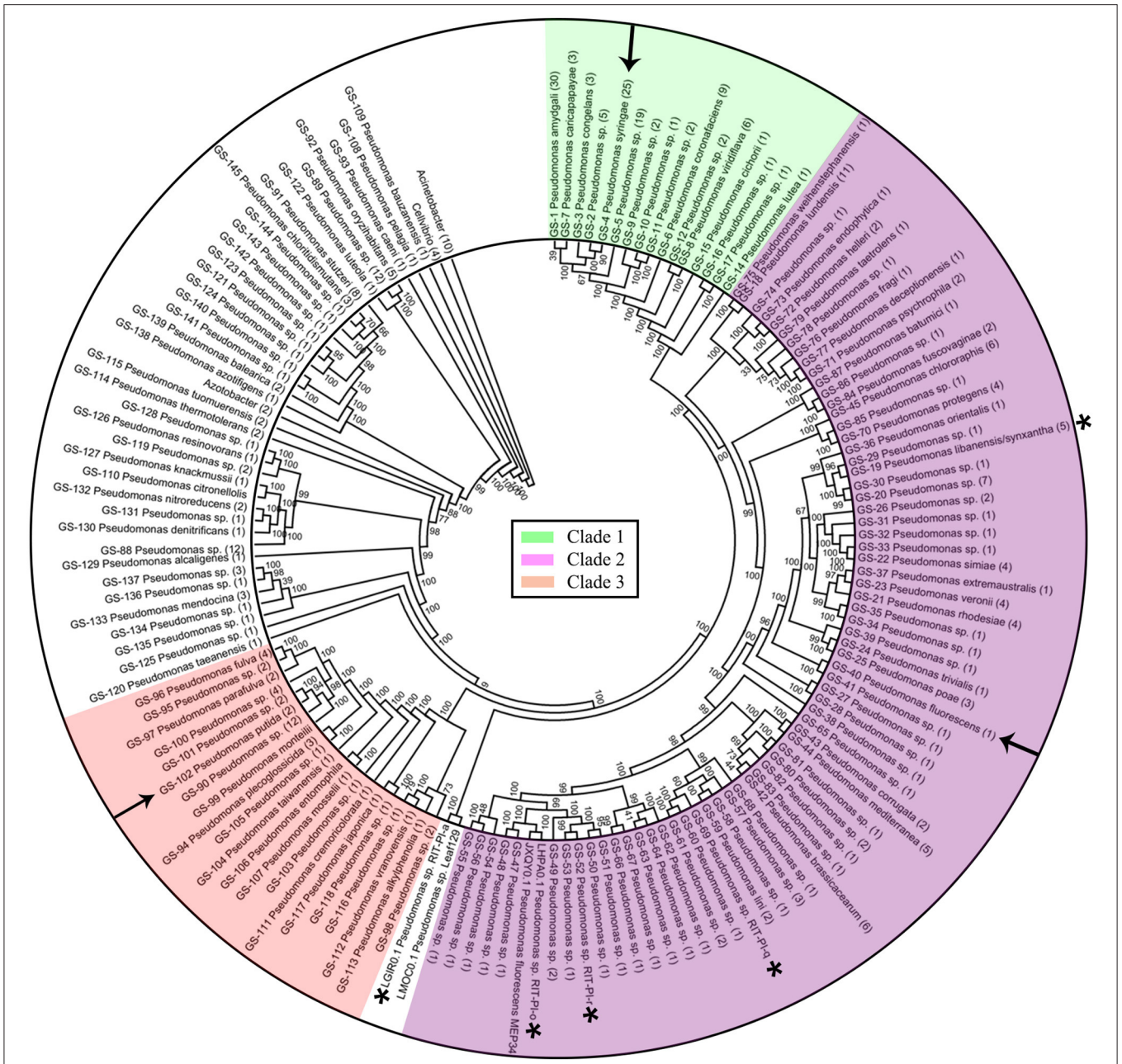


FIGURE 1 | Phylogenomic tree of 373 *Pseudomonas* strains with corrected taxonomic assignments. Three major clades were colored accordingly with arrows indicating the type strains of 3 common *Pseudomonas* species e.g., *P. putida*, *P. fluorescens*, and *P. syringae*. Node labels indicate Shimodaira-Hasegawa(SH)-like bootstrap values in percentage. Branches classified as the same genospecies (GS) were collapsed with the number in bracket at the end of each lineage representing the number of whole genome sequences in the corresponding GS. Whole genome sequence of *P. syringae*, *P. fluorescens*, and *P. putida* type strains are indicated with arrows while asterisks-labeled genomes are the five *Pseudomonas* species isolated from poison ivy vine tissue.

their evolutionary relationships which unfortunately is beyond the scope of this study. Such wgMLST analysis can be performed using existing bioinformatic tools such as BIGSdb, Roary and PGAdb-builder (Jolley and Maiden, 2010; Page et al., 2015; Liu et al., 2016). Strikingly, several whole genome sequences labeled as the same *Pseudomonas* species were located in different major clades or distantly apart taxa (**Supplemental Figure 1**). For example, while *P. protegens* CHA0^T clustered within Clade 2, *P. protegens* 231 PPRO formed a monophyletic group with other sequences labeled as *P. denitrificans* in a distant clade. Similarly, *P. marginalis* ICMP 11289 belongs to Clade 1 and clusters with other *P. viridiflava* whereas *P. marginalis* ICMP 9505 belongs to Clade 2 (**Supplemental Figure 1**). The lack of taxonomic congruence as reflected by the inconsistent

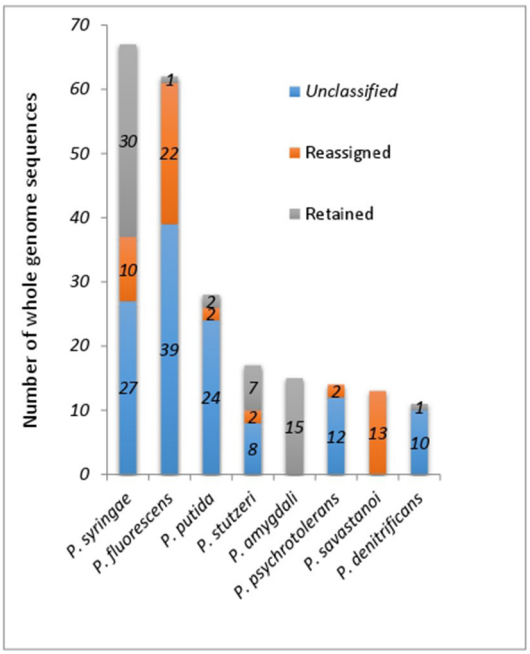


FIGURE 2 | Proportion of retained, reassigned and unclassified whole genome sequences within 8 most common *Pseudomonas* species excluding *P. aeruginosa*.

phylogenetic clustering among *Pseudomonas* strains with the same species designation raises suspicion about the species validity of *Pseudomonas* genomes submitted to public database.

Disentangling the Taxonomy of *Pseudomonas* Using *In silico* Genome-Genome Hybridization

Using 95% ANIm as the cutoff point for species delineation, a total of 145 genomic clusters were formed and when possible, assigned to a valid *Pseudomonas* species (**Figure 1** and **Supplemental Table 1**). Two-hundred and seven genomes were wrongly assigned at the species level as evidence by the lack of genomic clustering with their expected type strain genome and/or low nucleotide similarity (<95%) to the type strain Multilocus Sequence Typing (MLST) sequences.

All of the major species contain at least one member with incorrect species classification except for *P. amygdali* (**Figure 2**). Surprisingly, all 13 *P. savastanoi* genomes, majority of which originated from a single study (Mott et al., 2016) may have been misclassified as they were clustered into GS-1 containing *P. amygdali* ICMP3918^T (**Figure 2** and **Supplemental Table 1**). ANIm matrix of 113 whole genome sequences within Clade 1 (**Figure 1**) containing a majority of *P. syringae* identified at least 14 different genomic clusters (**Supplemental Figure 2**). The presence of whole genome sequence for the *P. syringae* type strain KTCC 12500^T = DSM 10604^T in GS-4 suggests that only 23 out of 67 *P. syringae* genomes were correctly assigned (**Supplemental Figure 2**), indicating that more than 50% of the currently deposited *P. syringae* genomes should be taxonomically revised. At least 10 *P. syringae* genomes have to be assigned to other valid *Pseudomonas* species (**Supplemental Table 1**). For example, *P. syringae* CC1513 should be re-classified as *P. coronafaciens* given its high pairwise ANIm of 99.2% to *P. coronafaciens* LMG5060^T.

Prior to the genome availability of *P. fluorescens* DSM 50090^T, *P. fluorescens* strain WH6 has served as the reference genome for the taxonomic assignment of *P. fluorescens* (Duan et al., 2013; Feng et al., 2015). Unfortunately, ANIm calculation shows that strain WH6 belongs to GS-35 with a pairwise ANIm of only 87.94% to the recently available *P. fluorescens* DSM 50090^T genome (**Supplemental Table 1** and **Supplemental Figure 3**). In another study, strain WH6 was also shown to be more similar to *P. azotofomans* LMG 21611^T than the representative species *P. fluorescens* DSM 50090^T using the MLSA method (Gomila et al., 2015). This finding along with our results here, suggest that the designation of *P. fluorescens* WH6 genome as the reference genome might have led to the subsequent mislabeling of newly-sequenced *Pseudomonas* genomes. As expected, out of a total of 62 whole genome sequences that were assigned as *P. fluorescens*, none of them belongs to the same genospecies as *P. fluorescens* DSM 50090^T. By re-evaluating the species designation based on ANIm clustering against other type strains of *Pseudomonas*, 22 of the 61 wrongly identified *P. fluorescens* strains were successfully re-assigned to the correct valid species name leaving 39 pseudomonad genospecies in Clade 2 (**Figure 1**) unassigned at the species level.

Only 1 strain of the total 28 *P. putida* strains belongs to the same genospecies with the *P. putida* type strain NBRC 14164^T (GS-102) which was similarly supported by their monophyletic clustering in the phylogenomic tree (Clade 3 in **Supplemental Figure 1**). Some of the mislabeled *P. putida* strains might be novel species given its independent formation of genospecies e.g., GS-103 in *P. putida* MTCC5279 (**Supplemental Table 1** and **Supplemental Figure 4**).

Species Validity of Closely Related *Pseudomonas* Species

Pseudomonas libanensis and *Pseudomonas synxantha*

Pseudomonas libanensis DSM 17149^T was isolated from Lebanon spring water whereas *P. synxantha* DSM 18928^T was isolated from milk cream in Iowa, USA (De Vos et al., 1989). Despite a pairwise 16S rDNA similarity of up to 99.5% between *P. libanensis* DSM 17149^T and *P. synxantha* DSM 18928^T, these two strains were considered as different species on the basis that strikingly high 16S rDNA similarity does not corroborate their reported low relative binding ratio of DNAs (RBR) (Dabboussi et al., 1999). The difference between *P. libanensis* and *P. synxantha* was further substantiated by other phenotypes such as levan formation and assimilation of histidine and erythritol (Dabboussi et al., 1999). Since 2005, both species have been included in the Bergey's Manual (Palleroni, 2005). However, it is worth noting that the difference in biochemical property can be due to single nucleotide polymorphism or presence/absence of horizontally transfer gene(s) that does not

contribute significantly to the overall nucleotide difference at the genomic level (Carnoy and Moseley, 1997; Boddicker et al., 2002; Monk et al., 2017). Furthermore, conventional DNA-DNA hybridization (DDH) procedures are technically demanding yet error-prone and often result in different outcomes (Rosselló-Mora, 2006; Johnson and Whitman, 2007). On the other hand, digital DDH is considered a more robust, pragmatic and an accurate replacement method for conventional DDH procedures (Richter and Rosselló-Móra, 2009; Auch et al., 2010). Contrary to their low RBR result, the whole genome sequences of *P. synxantha* DSM 18928^T and *P. libanensis* DSM 17149^T exhibit a pairwise ANIm of 96.7% clustering placing them both into GS-19 (**Figure 3**). In other words, genomic evidence does not support the classification of *P. synxantha* and *P. libanensis* as two separate species, warranting additional taxonomic investigations.

Pseudomonas brassicacearum and *Pseudomonas kilonensis*

Similarly, *Pseudomonas brassicacearum* and *Pseudomonas kilonensis* were described as two distinct species (Sikorski et al., 2001) on the basis of substantial dissimilarity in metabolic properties and a surprisingly borderline DNA-DNA hybridization similarity for species delineation. *P. kilonensis* was isolated from agricultural soil of northern Germany and described as a distinct species from *P. brassicacearum* by 10-12 different metabolic properties with convential DNA-DNA hybridization similarity of 93% (Sikorski et al., 2001). *Pseudomonas brassicacearum* was firstly described in 2000 when these strains were isolated from the rhizoplane of *Arabidopsis thaliana* and *Brassica napus* growing on different soils (Achouak et al., 2000). Since whole genome sequence of *P. brassicacearum* type strain is not yet available, blastn was conducted against housekeeping genes (*rpoB, rpoD* and *gyrB*) of *P. brassicacearum* CFBP 11706^T = CIP 107059^T and *P. kilonensis* 520-20^T = DSM 13647^T to validate the correct identification of 3 *P. brassicacearum* genomes belonging to GS-42 (**Figure 3**). *P. kilonensis* DSM 13647^T also show strikingly high pairwise ANIm to these three *P. brassicacearum* genomes leading to the clustering of all strains to GS-42, indicating that *P. kilonensis* should be considered as the junior synonym to *P. brassicacearum* as *P. brassicacearum* was described earlier than *P. kilonensis*.

Pseudomonas psychrotolerans and *Pseudomonas oryzihabitans*

Fourteen whole genome sequences labeled as *P. psychrotolerans* were distributed into two separate genospecies: GS-89 and GS-92. The presence of both *P. oryzihabitans* NBRC 102199^T and *P. psychrotolerans* DSM 15758^T in GS-92 implies that the two species are closely-related (ANIm similarity of 98.5%). *P. psychrotolerans* strain NS337 and L19 were correctly

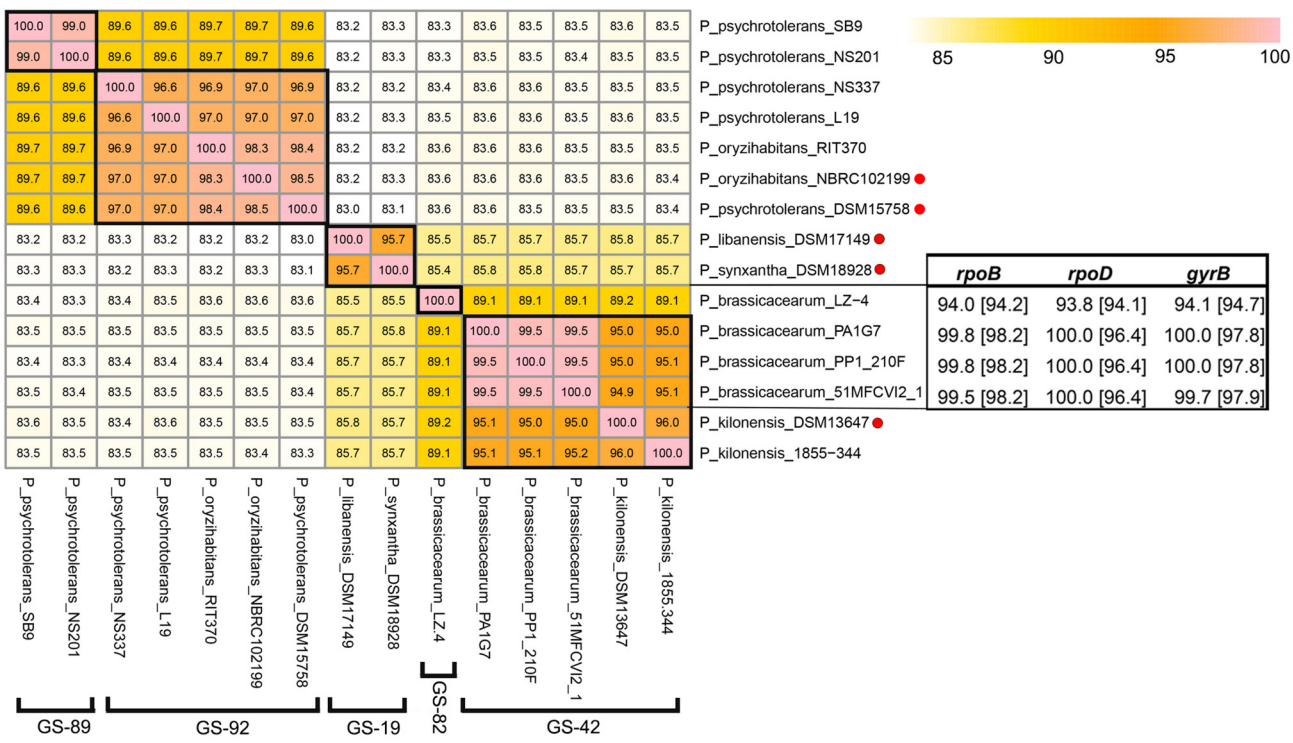


FIGURE 3 | ANIm matrix of 15 *Pseudomonas* genomes from closely-related species including *P. psychrotolerans-P. oryzihanbitans*, *P. brassicacearum-P. kilonensis,* and *P. libanensis-P. synxantha*. Type strains were indicated with red dots and the genospecies corresponding to each cluster was specified in bracket. Blastn results were expressed in percentage with values in and outside the bracket show percentage identity to the housekeeping genes (*rpoB, rpoD,* and *gyrB*) of *P. kilonensis* DSM 13647^T and *P. brassicacearum* CIP 107059^T, respectively. These two genomes along with whole genome of *P. psychrotolerans* DSM 15758^T (downloaded in April 2017) were only used to generate the ANIm matrix and were excluded from other analyses.

labeled but the other 12 sequences in GS-89 including strain SB9 and NS201 were incorrectly identified (**Figure 3**). *P. psychrotolerans* was firstly isolated from small European ungulates (Hauser et al., 2004) whereas *P. oryzihabitans* was isolated from soil in rice paddies (Kodama et al., 1985). It is suggested that *P. psychrotolerans* can be a junior synonym of *P. oryzihabitans* for such high similarity in digital DNA-DNA hybridization.

Taxonomic Re-Evaluation of *Pseudomonas* Strains Isolated from Poison Ivy (*Toxicodendron radicans*) Vine Tissue

With the current availability of defined genospecies within the genus *Pseudomonas*, we re-evaluated the taxonomic status of 5 *Pseudomonas* strains previously isolated from interior vine tissue of poison ivy by our group (Tran et al., 2015). Strain RIT-PI-g could be assigned to the species *Pseudomonas libanensis* forming GS-19 along with the *P. libanensis* DSM 17149^T corroborating our previous species assignment based on similarity to various housekeeping genes. *Pseudomonas.* sp. RIT-PI-q and *P.* sp. RIT-PI-r both belong to a single-member genospecies (GS-69 and GS-52, respectively), further underscoring the high genomic diversity of plant-associated *Pseudomonas* that remains to be explored and described with extensive taxon sampling effort. On the contrary, *Pseudomonas* sp. RIT-PI-o and *P.* sp. RIT-PI-a were placed in the same genomic cluster (GS-13) as the incorrectly labeled *P. fluorescens* MEP34 and *Pseudomonas* sp. leaf129 (**Supplemental Table 1** and **Supplemental Figure 5**). Both *P. fluorescens* MEP34 and *P.s* sp. leaf129 were isolated from leaf of *Arabidopsis thaliana*, a popular flowering plant in North America that has been adopted as the flowering plant genetic model species (Bai et al., 2015; Perisin et al., 2015). Such similar plant origins, e.g., poision ivy vine and Arabidopsis, poses interesting links between bacterial genomic characteristics and plant-hosting capacity.

CONCLUSIONS

The bacterial community has waited many years for the whole genome of *P. fluorescens* type strain to be sequenced (Flury et al., 2016). It is also surprising to note that whole genome sequence of some well-known *Pseudomonas* species such as *P. brassicacearum* is still unavailable despite the dramatic reduction in sequencing cost. Usually, a traditional BLASTn against 16S rRNA database was performed with varying degree of analysis depth to taxonomically assign newly-sequenced bacterial genomes which has potentially resulted in taxonomic misclassifications of several *Pseudomonas* genomes. Our study shows that in the future, a phylogenomic inference coupled with ANIm calculation could be a practical and a more reproducible method for inferring accurate genomic relatedness among *Pseudomonas* strains.

AUTHOR CONTRIBUTIONS

PT and HG conceived the experiments. PT performed the analysis of the data. PT, HG, and MS interpret the results and contributed to the manuscript write-up.

FUNDING

Funding for this study is provided by the Monash University Malaysia Tropical and Medicine Biology Platform. We also thank the Monash University Malaysia Genomics Facility for the provision of computational resources.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01296/full#supplementary-material

Supplemental Figure 1 | Phylogenomic tree of 373 *Pseudomonas* strains with original taxonomic assignments. Three major clades were colored accordingly with arrows indicating the type strains of 3 common *Pseudomonas* species e.g., *P. putida*, *P. fluorescens*, *P. syringae*. Whole genome sequence of *P. syringae*, *P. fluorescens* and *P. putida* type strains are indicated with arrows while asterisks-labeled genomes are the five *Pseudomonas* species isolated from poison ivy vine tissue.

Supplemental Figure 2 | Genomic clustering of Clade 1 using ANIm calculation.

Supplemental Figure 3 | Genomic clustering of Clade 2 using ANIm calculation.

Supplemental Figure 4 | Genomic clustering of Clade 3 using ANIm calculation.

Supplemental Figure 5 | Genomic clustering of other *Pseudomonas* strains using ANIm calculation.

Supplemental Table 1 | Genomic clusters of Clade 1, Clade 2, Clade 3 and other *Pseudomonas* strains with suggested assignation. Mislabeled genomes are indicated with asterisk whereas type strains are labeled "T."

REFERENCES

Achouak, W., Sutra, L., Heulin, T., Meyer, J.-M., Fromin, N., Degraeve, S., et al. (2000). *Pseudomonas brassicacearum* sp. nov. and *Pseudomonas thivervalensis* sp. nov., two root-associated bacteria isolated from *Brassica napus* and *Arabidopsis thaliana*. *Int. J. Syst. Evol. Microbiol.* 50, 9–18. doi: 10.1099/00207713-50-1-9

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

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.* 50(Pt 4), 1563–1589. doi: 10.1099/00207713-50-4-1563

Auch, A. F., Jan, M., Klenk, H.-P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., et al. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature.* 528, 364–369. doi: 10.1038/nature16192

Boddicker, J. D., Ledeboer, N. A., Jagnow, J., Jones, B. D., and Clegg, S. (2002). Differential binding to and biofilm formation on, HEp-2 cells by Salmonella enterica serovar Typhimurium is dependent upon allelic variation in the fimH gene of the fim gene cluster. *Mol. Microbiol.* 45, 1255–1265. doi: 10.1046/j.1365-2958.2002.03121.x

Buckley, M., and Roberts, R. J. (2007). *Reconciling Microbial Systematics and Genomics.* Washington, DC: American Academy of Microbiology.

Carnoy, C., and Moseley, S. L. (1997). Mutational analysis of receptor binding mediated by the Dr family of *Escherichia coli* adhesins. *Mol. Microbiol.* 23, 365–379. doi: 10.1046/j.1365-2958.1997.2231590.x

Dabboussi, F., Hamze, M., Elomari, M., Verhille, S., Baida, N., Izard, D., et al. (1999). Pseudomonas libanensis sp. nov., a new specie isolated from Lebanese spring waters. *Int. J. Syst. Evol. Microbiol.* 49, 1091--1101.

De Vos, P., Van Landschoot, A., Segers, P., Tytgat, R., Gillis, M., Bauwens, M., et al. (1989). Genotypic relationships and taxonomic localization of unclassified Pseudomonas and Pseudomonas-like strains by deoxyribonucleic acid: ribosomal ribonucleic acid hybridizations. *Int. J. Syst. Evol. Microbiol.* 39, 35–49.

Duan, J., Jiang, W., Cheng, Z., Heikkila, J. J., and Glick, B. R. (2013). The complete genome sequence of the plant growth-promoting bacterium Pseudomonas sp. UW4. *PLoS ONE* 8:e58640. doi: 10.1371/journal.pone.0058640

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

Feng, K., Li, R., Chen, Y., Zhao, B., and Yin, T. (2015). Sequencing and analysis of the Pseudomonas fluorescens GcM5-1A genome: a pathogen living in the surface coat of Bursaphelenchus xylophilus. *PLoS ONE* 10:e0141515. doi: 10.1371/journal.pone.0141515

Flury, P., Aellen, N., Ruffner, B., Péchy-Tarr, M., Fataar, S., Metla, Z., et al. (2016). Insect pathogenicity in plant-beneficial pseudomonads: phylogenetic distribution and comparative genomics. *ISME. J.* 10, 2527–2542. doi: 10.1038/ismej.2016.5

Gomila, M., Peña, A., Mulet, M., Lalucat, J., and García-Valdés, E. (2015). Phylogenomics and systematics in Pseudomonas. *Front. Microbiol.* 6:214. doi: 10.3389/fmicb.2015.00214

Hagström, Å., Pommier, T., Rohwer, F., Simu, K., Stolte, W., Svensson, D., et al. (2002). Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl. Environ. Microbiol.* 68, 3628–3633. doi: 10.1128/AEM.68.7.3628-3633.2002

Hauser, E., Kampfer, P., and Busse, H. J. (2004). *Pseudomonas psychrotolerans* sp. nov. *Int. J. Syst. Evol. Microbiol.* 54(Pt.5), 1633–1637. doi: 10.1099/ijs.0.03024-0

Hol, W. H. G., Bezemer, T. M., and Biere, A. (2013). Getting the ecology into interactions between plants and the plant growth-promoting bacterium *Pseudomonas fluorescens*. *Front. Plant Sci.* 4:81. doi: 10.3389/fpls.2013.00081

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

Johnson, J. L., and Whitman, W. B. (2007). "Similarity analysis of DNAs," in *Methods for General and Molecular Microbiology,* 3rd Edn., eds C. A. Reddy, T. J. Beveridge, J. A. Breznak and G. Marzluf (American Society of Microbiology), 624–652.

Jolley, K. A., and Maiden, M. C. J. (2010). BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595

Jun, S.-R., Wassenaar, T. M., Nookaew, I., Hauser, L., Wanchai, V., Land, M., et al. (2016). Diversity of Pseudomonas genomes, including populus-associated isolates, as revealed by comparative genome analysis. *Appl. Environ. Microbiol.* 82, 375–383. doi: 10.1128/AEM.02612-15

Kodama, K., Kimura, N., and Komagata, K. (1985). Two new species of Pseudomonas: *P. oryzihabitans* isolated from rice paddy and clinical specimens and *P. luteola* isolated from clinical specimens. *Int. J. Syst. Evol. Microbiol.* 35, 467–474.

Liu, Y.-Y., Chiou, C.-S., and Chen, C.-C. (2016). PGAdb-builder: a web service tool for creating pan-genome allele database for molecular fine typing. *Sci. Rep.* 6:36213. doi: 10.1038/srep36213

Maki, L. R., Galyan, E. L., Chang-Chien, M. M., and Caldwell, D. R. (1974). Ice nucleation induced by pseudomonas syringae. *Appl. Microbiol.* 28, 456–459.

Monk, I. R., Howden, B. P., Seemann, T., and Stinear, T. P. (2017). Correspondence: Spontaneous secondary mutations confound analysis of the essential two-component system WalKR in *Staphylococcus aureus*. *Nat. Commun.* 8:14403. doi: 10.1038/ncomms14403

Moore, E. R. B., Mau, M., Arnscheidt, A., Böttger, E. C., Hutson, R. A., Collins, M. D., et al. (1996). The determination and comparison of the 16S rRNA gene sequences of species of the genus Pseudomonas (sensu stricto and estimation of the natural intrageneric relationships). *Syst. Appl. Microbiol.* 19, 478–492. doi: 10.1016/S0723-2020(96)80021-X

Mott, G. A., Thakur, S., Smakowska, E., Wang, P. W., Belkhadir, Y., Desveaux, D., et al. (2016). Genomic screens identify a new phytobacterial microbe-associated molecular pattern and the cognate Arabidopsis receptor-like kinase that mediates its immune elicitation. *Genome Biol.* 17:98. doi: 10.1186/s13059-016-0955-7

Mulet, M., García-Valdés, E., and Lalucat, J. (2013). Phylogenetic affiliation of *Pseudomonas putida* biovar A and B strains. *Res. Microbiol.* 164, 351–359. doi: 10.1016/j.resmic.2013.01.009

Oliver, A., Cantón, R., Campo, P., Baquero, F., and Blázquez, J. (2000). High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science* 288, 1251–1253. doi: 10.1126/science.288.5469.1251

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Palleroni, N. J. (2005). "Genus I. Pseudomonas Migula 1894.," in *Bergey's Manual of Systematic Bacteriology,* 2nd Edn., eds D. J. Brenner, N. R. Krieg, and J. T. Staley (New York, NY: Springer), 323–379.

Perisin, M., Vetter, M., Gilbert, J. A., and Bergelson, J. (2015). 16Stimator: statistical estimation of ribosomal gene copy numbers from draft genome assemblies. *ISME. J.* 10, 1020–1024. doi: 10.1038/ismej.2015.16

Pieterse, C. M. J., Zamioudis, C., Berendsen, R. L., Weller, D. M., Van Wees, S. C. M., and Bakker, P. A. H. M. (2014). Induced systemic resistance by beneficial microbes. *Annu. Rev. Phytopathol.* 52, 347–375. doi: 10.1146/annurev-phyto-082712-102340

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Nat. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rosselló-Mora, R. (2006). "DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation," in *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, ed E. Stackebrandt (Berlin: Springer), 23–50.

Segata, N., Bornigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4:2304. doi: 10.1038/ncomms3304

Sikorski, J., Stackebrandt, E., and Wackernagel, W. (2001). Pseudomonas kilonensis sp. nov., a bacterium isolated from agricultural soil. *Int. J. Syst. Evol. Microbiol.* 51, 1549–1555. doi: 10.1099/00207713-51-4-1549

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

Tran, P. N., Tan, N. E. H., Lee, Y. P., Gan, H. M., Polter, S. J., Dailey, L. K., et al. (2015). Whole-genome sequence and classification of 11 endophytic bacteria from poison ivy (*Toxicodendron radicans*). *Genome Announc.* 3, e01319–e01315. doi: 10.1128/genomeA.01319-15

Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., Vivian, A., and Harayama, S. (2000). Phylogeny of the genus Pseudomonas: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. *Microbiology* 146(Pt 10), 2385–2394. doi: 10.1099/00221287-146-10-2385

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 Tran, Savka and Gan. 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 March 2018
doi: 10.3389/fmicb.2018.00512

Genotypic and Lipid Analyses of Strains From the Archaeal Genus *Halorubrum* Reveal Insights Into Their Taxonomy, Divergence, and Population Structure

Rafael R. de la Haba[1†], Paulina Corral[1†], Cristina Sánchez-Porro[1], Carmen Infante-Domínguez[1], Andrea M. Makkay[2], Mohammad A. Amoozegar[3], Antonio Ventosa[1] and R. Thane Papke[2*]*

[1] Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [2] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, United States, [3] Department of Microbiology, Faculty of Biology and Center of Excellence in Phylogeny of Living Organisms, College of Science, University of Tehran, Tehran, Iran

OPEN ACCESS

Edited by:
Haiwei Luo,
The Chinese University of Hong Kong,
Hong Kong

Reviewed by:
Shaoxing Chen,
Anhui Normal University, China
Lin Xu,
Zhejiang Sci-Tech University, China
Heng-Lin Cui,
Jiangsu University, China

***Correspondence:**
Antonio Ventosa
ventosa@us.es
R. Thane Papke
thane@uconn.edu

[†] These authors have contributed equally to this work.

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

Received: 11 December 2017
Accepted: 06 March 2018
Published: 29 March 2018

Citation:
de la Haba RR, Corral P, Sánchez-Porro C, Infante-Domínguez C, Makkay AM, Amoozegar MA, Ventosa A and Papke RT (2018) Genotypic and Lipid Analyses of Strains From the Archaeal Genus Halorubrum Reveal Insights Into Their Taxonomy, Divergence, and Population Structure. Front. Microbiol. 9:512. doi: 10.3389/fmicb.2018.00512

To gain a better understanding of how divergence occurs, and how taxonomy can benefit from studying natural populations, we isolated and examined 25 closely related *Halorubrum* strains obtained from different hypersaline communities and compared them to validly named species and other reference strains using five taxonomic study approaches: phylogenetic analysis using the 16S rRNA gene and multilocus sequencing analysis (MLSA), polar lipid profiles (PLP), average nucleotide identity (ANI) and DNA-DNA hybridization (DDH). 16S rRNA gene sequence could not differentiate the newly isolated strains from described species, while MLSA grouped strains into three major clusters. Two of those MLSA clusters distinguished candidates for new species. The third cluster with concatenated sequence identity equal to or greater than 97.5% was comprised of strains from Aran-Bidgol Lake (Iran) and solar salterns in Namibia and Spain, and two previously described species isolated from Mexico and Algeria. PLP and DDH analyses showed that Aran-Bidgol strains formed uniform populations, and that strains isolated from other geographic locations were heterogeneous and divergent, indicating that they may constitute different species. Therefore, applying only sequencing approaches and similarity cutoffs for circumscribing species may be too conservative, lumping concealed diversity into a single taxon. Further, our data support the interpretation that local populations experience unique evolutionary homogenization pressures, and once relieved of insular constraints (e.g., through migration) are free to diverge.

Keywords: *Halorubrum*, taxonomy, divergence, MLSA, polar lipid, ANI, DNA–DNA hybridization

INTRODUCTION

The genus *Halorubrum* was proposed in order to accommodate the species *Halobacterium saccharovorum*, *Halobacterium sodomense*, *Halobacterium trapanicum*, and *Halobacterium lacusprofundi* (McGenity and Grant, 1995). Currently, it is the largest genus in the class *Halobacteria*, and was recently assigned to the newly described family *Halorubraceae* within the

order *Haloferacales* (Domain Archaea) (Gupta et al., 2016), and at present contains 36 species with validly published names (Parte, 2018). Members of this genus are phenotypically diverse but all are metabolically aerobic, chemoorganotrophic, and obligately halophilic with growth occurring in media containing 1.0–5.2 M NaCl. They have been isolated from marine salterns, salt lakes, coastal sabkhas, hypersaline soda lakes, saline soils and salt-fermented seafood (Oren et al., 2009; Amoozegar et al., 2017) and they are frequently the numerically dominant microorganisms present in hypersaline environments, as revealed through both culture-dependent and culture-independent techniques (Ghai et al., 2011; Makhdoumi-Kakhki et al., 2012; Ma and Gong, 2013; Fernández et al., 2014a,b; Ventosa et al., 2015).

The 16S rRNA gene sequence is a universal phylogenetic marker within the prokaryotes, and therefore considered essential in taxonomic studies of haloarchaea, including the genus *Halorubrum*. However, over the years it has been demonstrated that the 16S rRNA gene has many disadvantages to its utilization as a phylogenetic and taxonomic marker for the class *Halobacteria* (also known as the haloarchaea). Its highly conserved nature does not allow relevant discernable differentiation among closely related species (for example, 99.4% sequence similarity between *Halorubrum californiense* and *Halorubrum chaoviator* [Pesenti et al., 2008; Mancinelli et al., 2009]); the rRNA operons experience intragenic recombination (Boucher et al., 2004), resulting in reticulated evolutionary histories. It was the most frequently transferred gene among closely related but otherwise distinct lineages (Papke, 2009), making haloarchaeal phylogeny and taxonomy difficult to interpret. Additionally, many haloarchaeal genera have multiple divergent copies of rRNA genes with greater than 6% sequence dissimilarity -the divergence between copies in a single cell can be equal to that seen between genera- have been described within the class *Halobacteria* (Boucher et al., 2004; Sun et al., 2013). To overcome some of these limitations, protein-encoding housekeeping genes have been suggested as alternative phylogenetic markers within the class *Halobacteria* such as *radA* or *rpoB*' genes (Sandler et al., 1999; Walsh et al., 2004; Enache et al., 2007; Minegishi et al., 2010). However, because haloarchaea are known for their frequent recombination across great genetic distances (Nelson-Sathi et al., 2012; Williams et al., 2012; DeMaere et al., 2013) reliance on a single genetic marker cannot provide enough useful phylogenetic information and may lead to taxonomic confusion. Therefore, a multilocus sequence analysis (MLSA) approach that excludes using the 16S rRNA gene was proposed as a preferred methodology for classification and evolutionary studies of the *Halobacteria* (Papke et al., 2011). More recently, a set of five housekeeping genes, i.e., *atpB*, *EF-2*, *glnA*, *ppsA*, and *rpoB*', have been suggested as recommended markers for the genus *Halorubrum* (Fullmer et al., 2014; Ram Mohan et al., 2014). The use of this MLSA approach instead of the employment of 16S rRNA gene sequence analysis has been endorsed by the ICSP-Subcommittee on the taxonomy of *Halobacteria* (Oren and Ventosa, 2013, 2016).

The protocol for species taxonomy of prokaryotes ultimately relies on DNA-DNA hybridization (DDH) as the "gold standard" for defining bacterial species (Wayne et al., 1987; Stackebrandt et al., 2002). As with 16S rRNA gene sequencing, it too has many associated issues discussed elsewhere (e.g., Johnson, 1994; Zeigler, 2003; Hanage et al., 2006). The use of MLSA as an alternative method for species demarcation in prokaryotes has been successfully applied to several bacterial groups, e.g., lactic acid bacteria (Naser et al., 2006), *Borrelia* (Richter et al., 2006), mycobacteria (Mignard and Flandrois, 2008), pseudomonads and relatives (Young and Park, 2007), *Ensifer* (Martens et al., 2008), *Vibrionaceae* (Pascual et al., 2010; López-Hermoso et al., 2017), and *Aeromonas* (Martinez-Murcia et al., 2011; Roger et al., 2012). Though clearly useful for phylogeny, the degree of congruence between the MLSA and DNA-DNA reassociation data has not been established for taxonomic purposes within the haloarchaea. Therefore, one focus of this study was to compare the MLSA results with DDH and other polyphasic analyses for the genus *Halorubrum* specifically, but also the haloarchaea in general. Polar lipid analysis, a powerful taxonomic marker in haloarchaea (Torreblanca et al., 1986; Oren et al., 2009), and average nucleotide identity (ANI), which has advanced the understanding of prokaryotic taxonomy in other taxa (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Chun and Rainey, 2014; Rosselló-Móra and Amann, 2015; Chun et al., 2018) were compared in this study to assess their usefulness in capturing haloarchaeal taxonomy, divergence and population structure.

MATERIALS AND METHODS

Strains and Culture Conditions

Most of the *Halorubrum* strains used in this study, and other strains from members of the class *Halobacteria* used for comparative analysis, were obtained from the respective culture collection (Table S1) and cultivated following the media and growth conditions recommended by the culture collections. Several strains were isolated in this study from sediment samples of the hypersaline lake Aran-Bidgol, Iran (Table S1), using the plate dilution technique on YPC (Yeast extract, Peptone, Casamino acids) medium, which contained a mixture of 20% (w/v) salts (15.0% NaCl, 2.3% $MgSO_4$, 2% $MgCl_2$, 0.6% KCl, 0.01% $MnCl_2$), 10 mM Tris-HCl (pH 8), 0.5% yeast extract, 1% peptone, and 1% casamino acids, after incubation at 37°C for 15 days. The remaining strains isolated in this study were recovered from hypersaline water samples of a solar saltern in the Namibia desert and salterns in Huelva, Spain (Table S1), again using the plate dilution technique on Halophilic Medium (HM) (Ventosa et al., 1982) with ∼20% (w/v) total salts (17.8% NaCl, 0.1% $MgSO_4$, 0.036% $CaCl_2$, 0.2% KCl, 0.006% $NaHCO_3$, 0.023% NaBr, traces of $FeCl_3$), 1% yeast extract (Difco), 0.5% proteose-peptone no. 3 (Difco), and 0.1% glucose. The pH was adjusted to 7.2 with 1 M KOH. For routine growth the strains were cultivated in modified SW20 (SeaWater 20%) medium (Rodríguez-Valera et al., 1980) with 20% (w/v) total salts (16.2% NaCl, 1.4% $MgCl_2$, 1.92% $MgSO_4$, 0.072% $CaCl_2$, 0.4% KCl, 0.012% $NaHCO_3$, 0.0052% NaBr), 0.5% yeast extract (Difco), and 0.5% casamino acids. The pH was adjusted to 7.2 with 1 M KOH. When necessary, solid media were prepared by adding 2.0% (w/v) Bacto-agar (Difco). These cultures were maintained at−80°C as

suspensions (prepared with 15%, v/v glycerol) in modified SW20 medium.

DNA Preparation

Genomic DNA from each culture was isolated and purified using standard methods (Qiagen kit), quantified and checked for quality using a Nanodrop spectrophotometer ND-1000 at 260/280 nm and diluted with TE (10 mM Tris, pH 8.0, 1 mM EDTA) to 20 ng μl^{-1} for subsequent PCR analysis.

Amplification and Sequencing

From each strain, the following five genes were amplified and sequenced: *atpB* (ATP synthase subunit B), *EF-2* (elongation factor 2), *glnA* (glutamine synthetase), *ppsA* (phosphoenolpyruvate synthase), and *rpoB*' (RNA polymerase subunit B'). These genes were chosen for analysis because they are single copy protein-encoding genes previously investigated with success in haloarchaea (Fullmer et al., 2014; Ram Mohan et al., 2014) and recently recommended for MLSA scheme by the ICSP-Subcommittee on the taxonomy of *Halobacteria* (Oren and Ventosa, 2013, 2016). Primers used for amplification and sequencing (Papke et al., 2011; Fullmer et al., 2014) annealed the respective locus across the *Halobacteria*, and therefore one to three degenerated positions were included into the primers, with the exception of rpoB_962F_M13 primer. Additionally, to enhance the sequencing success rate primers containing M13 sequence were used as reported elsewhere (Fullmer et al., 2014; Ram Mohan et al., 2014) (Table S2). The 16S rRNA gene was amplified and sequenced for those strains isolated from the lake Aran-Bidgol and from the salterns in Namibia and Spain, using previously described universal primers (Arahal et al., 1996; López-García et al., 2001) (Table S2).

PCR amplification was performed in a 50 μl reaction mixture composed of 5.0 μl 10× PCR buffer, 1.5 μl $MgCl_2$ (50 mM stock), 1.0 μl dNTPs (10 mM each), 2.0 μl each forward and reverse primers (10 μM), 1.0 μl *Taq* polymerase (5 U μl^{-1}; Invitrogen *Taq* DNA Polymerase Native or Roche Fast Start Universal SYBR Green Master [Rox]), 1.0 μl template DNA (20 ng μl^{-1}), and ddH_2O to a final volume of 50 μl. All reactions were performed in an Eppendorf Mastercycler Ep gradient thermocycler (Eppendorf). The PCR cycling conditions included an initial denaturation step (1 min, 94°C) followed by 30 cycles of denaturation (1 min, 94°C), annealing (1 min) and extension (1 min, 72°C) and a final extension period (5 min, 72°C). The annealing temperature for the thermal profile was optimized for each primer set and is shown in Table S2.

The PCR amplicons were examined by agarose gel electrophoresis (1%) and stained with ethidium bromide. Purification of the amplicons was carried out by using standard procedures and sequenced in both directions by the dideoxynucleotide chain termination method using the BigDye chemistry on an ABI 3130XL DNA Analyzer or an ABI 3730XL DNA Analyzer (Applied Biosystems), according to the manufacturer's instructions. Sequences belonging to the same locus were assembled using the software package Geneious (http://www.geneious.com/) and edited manually to resolve ambiguous positions. For several strains with sequenced genomes, housekeeping gene sequences were retrieved from GenBank/EMBL/DDBJ databases. In the case of *Halorubrum californiense* DSM 19288^T, it was not possible to obtain the *atpB* gene sequence probably because its chromosome has not been completely assembled: the possibility that this strain does not require *atpB* for survival in nature is exceedingly remote. Several 16S rRNA gene sequences were also retrieved from the GenBank/EMBL/DDBJ databases (Table S1).

In summary, five protein-encoding genes from 55 strains (30 type strains and 25 new isolates) within the genus *Halorubrum* were obtained and analyzed. Unfortunately, for *glnA* and *ppsA* genes there were one and two isolated strains, respectively, from which the sequence could not be obtained, no matter how many attempts were made or conditions were tested (Table S1). Gene sequences of *Haloarcula vallismortis* ATCC 29715^T, *Halobacterium salinarum* R1^T, and *Haloferax volcanii* DS2^T/NCIMB 2287^T were also included for phylogenetic analysis as outgroups.

Multiple Sequence Alignments

DNA sequences for each housekeeping gene were aligned using Muscle version 3.6 (Edgar, 2004a,b) taking into account the corresponding amino acid alignments for protein-coding genes. Alignments were edited using Mesquite version 2.75 (Maddison and Maddison, 2011). Individual gene alignments were concatenated in the following order: *atpB*, *EF-2*, *glnA*, *ppsA* and *rpoB*'. For the 16S rRNA gene, the obtained sequences were aligned using the ARB software (Ludwig et al., 2004).

The analyzed lengths of sequence data determined from the multiple alignments were: 496 bp for the *atpB* gene, 507 bp for the *EF-2* gene, 526 bp for the *glnA* gene, 514 bp for the *ppsA* gene, and 522 bp for the *rpoB*' gene (Table S3). Multitaxon alignments for the *EF-2* and *rpoB*' loci did not contain gaps, whereas several gaps were present within the *atpB*, *glnA*, and *ppsA* gene alignments. None of the positions in the alignments were omitted for the analysis.

Phylogenetic Tree Reconstructions

Phylogenies were calculated for the 16S rRNA gene as well as for the individual housekeeping gene alignments and for the five concatenated loci. Optimal models of evolution were estimated from the nucleotide data using jModelTest version 2.1 (Darriba et al., 2012) considering 11 nucleotide substitution schemes, including models with equal/unequal base frequencies, with/without a proportion of invariable sites and with/without four rate variation among sites, and selecting the best model according to the Akaike (AIC) criterion (Akaike, 1974). The models proposed for the nucleotide data included TVM, TIM1, TIM2 (Posada, 2003), and GTR (Tavare, 1986). These models all consider unequal base frequencies, but vary in the number of transition and transversion rates deemed necessary to model evolution. All of the models proposed included the gamma shape parameter and considered invariable sites (Table S4).

The subsequent sequence analyses were performed using the PAUP* version 4.0b10 phylogenetic software (Swofford, 2003) for the neighbor-joining (NJ) (Saitou and Nei, 1987) and maximum-parsimony (MP) (Kluge and Farris, 1969) methods

and PhyML (Guindon and Gascuel, 2003) for the maximum-likelihood (ML) (Felsenstein, 1981) method. Support for NJ, MP, and ML phylogenies was determined through bootstrap analysis with 1,000 replications. Only bootstrap values equal or greater than 70% are shown on the trees. Topology congruence tests among individual and concatenated gene phylogenies were performed using Concaterpillar v. 1.8a software (Leigh et al., 2008) setting the *P*-value cutoff to 0.05.

A supertree, considering all previously obtained individual trees, was constructed on the basis of the five individual gene phylogenies by means of Matrix Representation using Parsimony (MRP) method (Loomis and Smith, 1990; Baum, 1992; Ragan, 1992) as implemented in Clann 3.2.3 (Creevey and McInerney, 2005) setting parameters by default.

Lipid Profile

Total lipids were extracted with chloroform/methanol using the extraction method described by Bligh and Dyer (1959), as modified for extreme halophiles (Corcelli et al., 2000). The extracts were carefully dried in a SpeedVac Thermo Savan SPD111V. The stock was prepared dissolving the dried extracts in chloroform to a final concentration 10 mg·ml-1. From this stock, 10 μl equivalent to 100 μg of total lipid extract were applied and analyzed by one dimensional High-Performance Thin Layer Chromatography (HPTLC) on Merck silica gel plates with solvent system A (chloroform: methanol: 90% acetic acid, 65: 4: 35, v/v) (Angelini et al., 2012). Staining of the lipids present in the HPTLC bands was carried out by spraying the plates with two different stains followed by brief heating at 160°C: (a) sulfuric acid 5% (v/v), a universal developer for visualizing all lipids; (b) molybdenum blue, specific for phospholipids.

DNA–DNA Hybridization, ANI Calculation, and Correlation Studies

The strains used for DDH experiments included those isolated from the lake Aran-Bidgol and from the solar salterns in Namibia and Spain belonging to the MLSA defined groups 1, 2, and 3, and the species of the genus *Halorubrum* with validly published names that shared equal to or more than 97% 16S rRNA gene sequence similarity. Additionally, the type species of the genus *Halorubrum*, *Hrr. saccharovorum* DSM 1137^T was also included in the study as a reference. For DNA-DNA hybridizations strain Fb21 from group 1, strain Ib24 from group 2, and strain Cb34 from group 3 were randomly selected as representatives of each group and were used as reference for these studies.

DDH studies were conducted according to the competition procedure of the membrane filter method (Johnson, 1994), as previously reported for haloarchaea studies (Pesenti et al., 2008; Mancinelli et al., 2009; Corral et al., 2015). The hybridization temperature was 61.4°C, which was within the limit of validity for the filter method (De Ley and Tijtgat, 1970) and the percentage of hybridization was calculated according to Johnson (1994). The experiments were performed in triplicate. The interpretation is according to Wayne et al. (1987) where it has been established that strains belonging to the same species have values of DNA-DNA hybridization at or above 70% and ΔTm equal or less than 5°C. Two strains having high values of DNA-DNA hybridization are phylogenetically related.

The *in silico* determined average nucleotide identity (ANI) between two genomes has been widely accepted by taxonomist as the substitute for DDH species delineation, with a cutoff value of 95–96% (Goris et al., 2007; Richter and Rosselló-Móra, 2009; Rosselló-Móra and Amann, 2015). ANI similarity index between pairs of genomes was calculated using BLAST (ANI_b) as implemented in JSpecies software (Konstantinidis and Tiedje, 2005; Richter and Rosselló-Móra, 2009). The genome sequences of *Halorubrum* strains used for ANI_b calculation were retrieved from GenBank/EMBL/DDBJ databases or obtained in this study (Table S1).

DNA–DNA hybridization data obtained in this study were compared to the distance matrix data for the 16S rRNA gene, the distance matrix for each gene individually and the distance matrix corresponding to the 5-gene concatenated sequences, as well as to the ANI_b values. Correlation between values was calculated using Pearson's product–moment correlation coefficient.

RESULTS AND DISCUSSION

16S rRNA Gene Sequence Analyses

Using samples obtained from the hypersaline lake Aran-Bidgol (Iran) and solar salterns in Namibia and Spain we were able to isolate 21, two, and two strains respectively (Table S1) that were affiliated with the genus *Halorubrum* according to their 16S rRNA and housekeeping gene sequences. In comparison with other *Halorubrum* type strains, the 16S rRNA gene phylogenetic trees showed that they formed four different phylogenetic clusters: group 1, comprising strains Fb21, C191, G37, SD683, and Ga66, and the species *Halorubrum chaoviator* Halo-G*^{T}, *Halorubrum californiense* SF3-213^T, and *Halorubrum sodomense* ATCC 33755^T; group 2, including the Aran-Bidgol strains Fa5, Ga2p, Fc2, ASP57, SD612, Ec15, Ga36, and Ec5, as well the species *Halorubrum ezzemoulense* CECT 7099^T; group 3, containing the Aran-Bidgol strains Ib24, Eb13, Ib25, Ea1, Ea10, Hd13, Ib43, Ea8, and Ea4p; and group 4, clustering the Aran-Bidgol strains Cb34 and C170. Strain ARQ123 fell in the boundaries of group 2, but did not belong to any of the above-identified phylogroups. Only groups 3 and 4 showed a strong bootstrap support (**Figure 1**). The 16S rRNA gene sequence similarities within each group were 100–99.7% for group 1, 100–98.0% for group 2, 100–98.3% for group 3, and 99.9% for group 4. Groups 1 and 2 were very closely related, sharing between 100–98.8% sequence similarities. The use of EzBioCloud database (Yoon et al., 2017) confirmed that for each group the most closely related taxa with validly published species names were: *Halorubrum chaoviator* DSM 19316^T for group 1 and group 2 (100% 16S rRNA gene sequence similarity in both cases); *Halorubrum kocurii* JCM 14978^T for group 3 (98.8%); and *Halorubrum cibi* B31^T for group 4 (98.9%). Despite the proximity of strain ARQ123 to group 2 observed in **Figure 1**, the EzBioCloud server indicated that the closest relative to strain ARQ123 was also the species *Halorubrum chaoviator* DSM 19316^T, sharing 99.3% 16S rRNA gene sequence similarity. Applying the typically used 97% 16S rRNA gene species cutoff

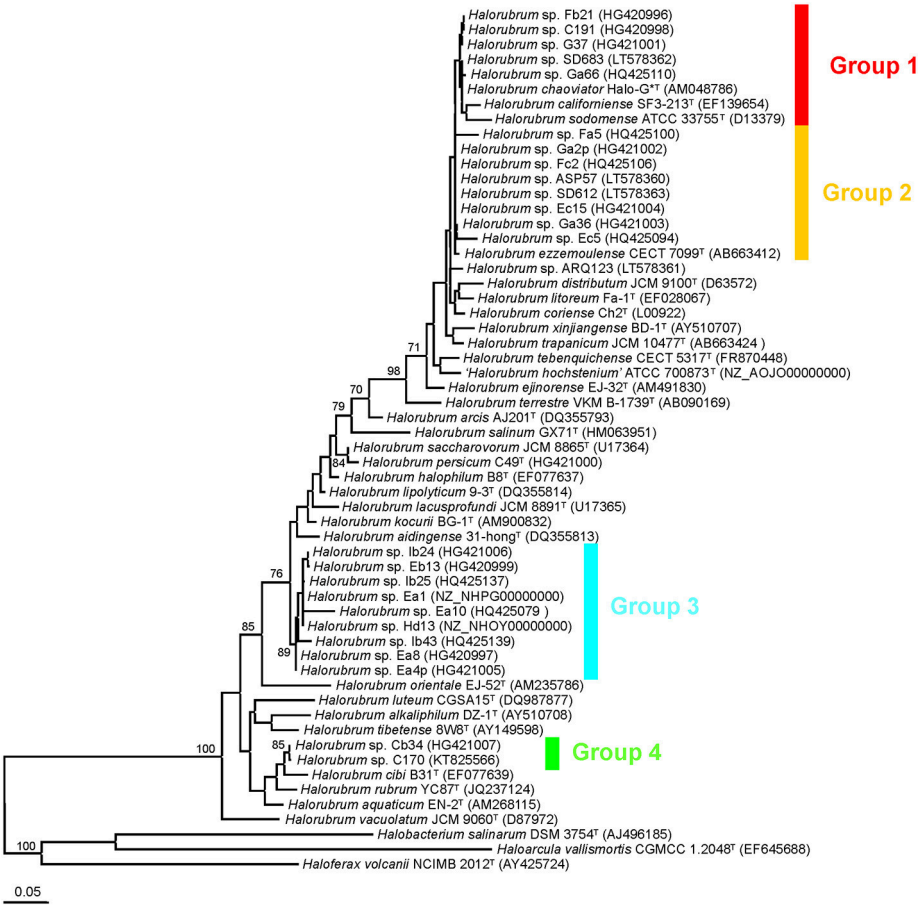


FIGURE 1 | Maximum-likelihood tree based on 16S rRNA gene sequence showing the phylogenetic relationship between members of the genus *Halorubrum* and strains isolated in this study. The accession numbers of the sequences used are shown in parentheses after the strain designation. Bootstrap values >70% are indicated. The species *Haloarcula vallismortis*, *Haloferax volcanii,* and *Halobacterium salinarum* were used as outgroups. The scale bar represents 0.05 substitutions per nucleotide position. Different phylogroups have been marked with different colors.

concept (Stackebrandt et al., 2002), the sequence data would suggest that all new isolates in this study should be considered as strains of previously described species that would not merit further analysis for the description of those groups as new species within the *Halobacteria*. However, 16S rRNA gene sequence conservation and transfer frequency conceals diversity in the class *Halobacteria* (Papke et al., 2004, 2007, 2011; Papke, 2009), suggesting the possibility of cryptic species in these *Halorubrum* strains. Therefore, we proceeded to investigate these *Halorubrum* strains with MLSA.

MLSA Analyses

For each locus phylogenetic trees with 1,000 bootstrap pseudoreplicates were constructed based on ML, MP and NJ methods, and according to the best evolutionary model calculated using jModelTest program (Table S4). Although all five alignments possessed a similar length (between 496 and 526 bp), *ppsA* had the highest proportion of parsimony-informative sites (39%), followed by *EF-2* (30%), *glnA* (27%), *rpoB'* (25%) and *atpB* (21%) genes (Table S3). The average pairwise sequence similarity values across the genus *Halorubrum* for *atpB*, *EF-2*, *glnA*, *ppsA*, and *rpoB'* were 95.0, 92.5, 95.5, 89.9, and 94.1%, respectively, suggesting that *ppsA* is the most resolving phylogenetic marker. Concatenation of the five loci produced an alignment of 2,565 bp, containing 28% parsimony-informative sites (Table S3), with an average pairwise sequence similarity of 93.7%.

Phylogenetic trees constructed from individual protein coding genes (Figure S1), concatenated genes (**Figure 2**), and the supertree analysis (**Figure 3**) produced different overall topologies in comparison to each other, but they all support the inclusion of the same strains for groups 1, 2, and 3, with only two exceptions: in group 1, the *EF-2* and *rpoB'* phylogenies placed strains ARQ123 and ASP57, respectively, far from the other strains of group 1, which can be explained by gene transfer events. Groups 2 and 3 were consistently composed regardless of the analysis. The biggest difference between the protein coding and the 16S rRNA gene-based phylogenetic analyses was groups 1 and 2 in the 16S rRNA gene tree were collapsed into group 1 in the protein coding gene-based trees. *Halorubrum chaoviator*

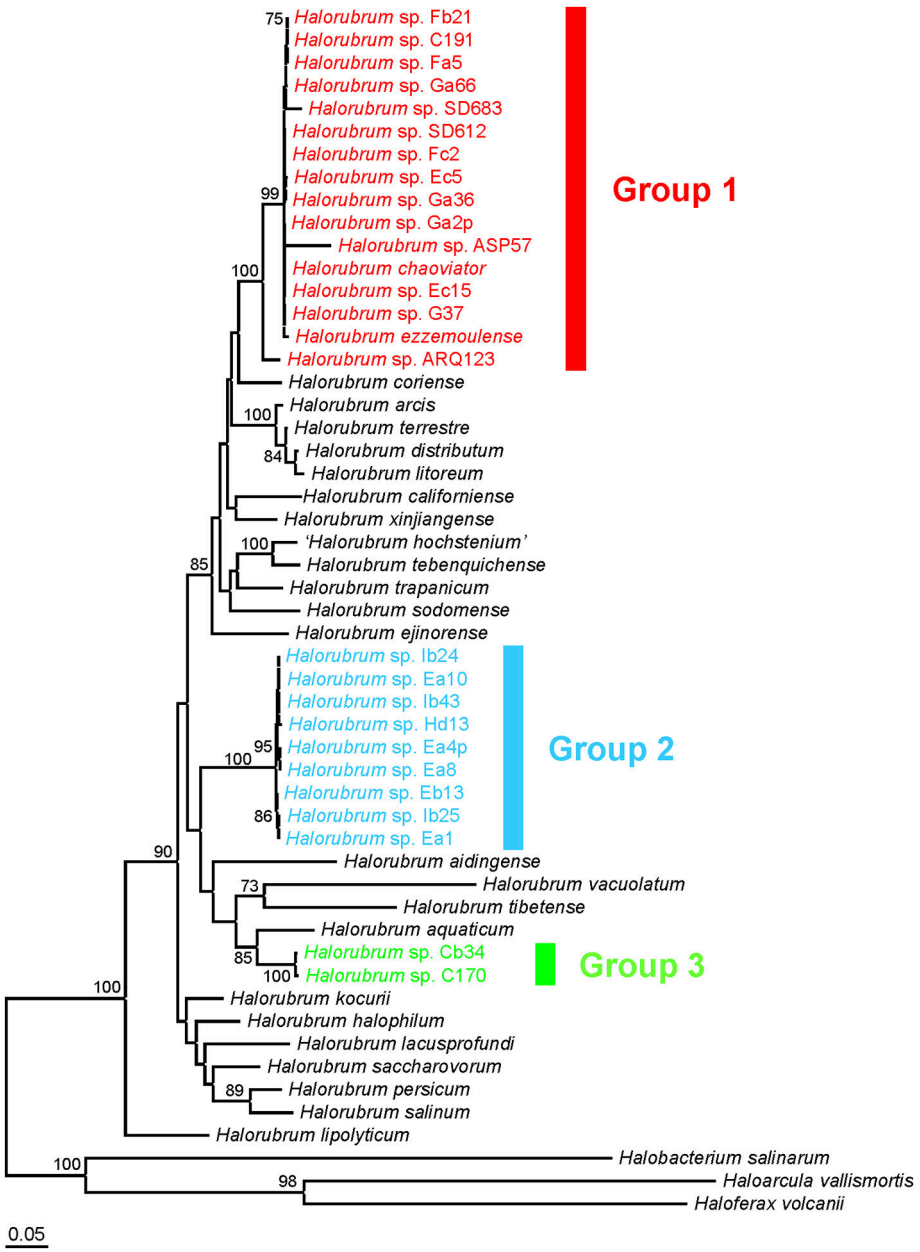


FIGURE 2 | Maximum-likelihood tree based on the five-gene concatenated sequence showing the phylogenetic relationship between members of the genus *Halorubrum* and strains isolated in this study. The accession numbers of the sequences used are shown in Table S1. Bootstrap values >70% are indicated. The species *Haloarcula vallismortis*, *Haloferax volcanii,* and *Halobacterium salinarum* were used as outgroups. The scale bar represents 0.05 substitutions per nucleotide position. Different phylogroups have been marked with different colors.

and *Halorubrum ezzemoulense* from groups 1 and 2 respectively, in the 16S rRNA gene tree, collapsed into a single large cluster called group 1 in the MLSA tree. *Halorubrum californiense* and *Halorubrum sodomense,* which fell into group 1 the rRNA gene tree were excluded from any of the new MLSA defined groups. Strain ARQ123, which was not part of groups 1 or 2 in the 16S rRNA gene tree found a home in group 1 in the concatenated and supertree analyses (**Figures 2**, **3**).

Five-gene concatenated sequence similarities ranged between 100–95.8% for strains of group 1 (this range decreased to 100–98.8% when we excluded strains ARQ123 and ASP57 which have highly divergent *EF-2* and *rpoB'* genes respectively), 100–99.6% for group 2, and 99.7% for group 3 (**Table 1**). The most closely related type strains to group 1 were *Hrr. chaoviator* Halo-G^{*T}/DSM 19316^{T} and *Hrr. ezzemoulense* DSM 17463^{T} (99.8% sequence similarity for both of them), to group 2 it was *Hrr. salinum* JCM 17093^{T} (94.4%), and to group 3 it was *Hrr. aquaticum* JCM 14031^{T}/CGMCC 1.6377^{T} (95.0%) (**Table 1**); however, phylogenetic analyses indicated that the species *Hrr. salinum* JCM 17093^{T} is quite distantly related

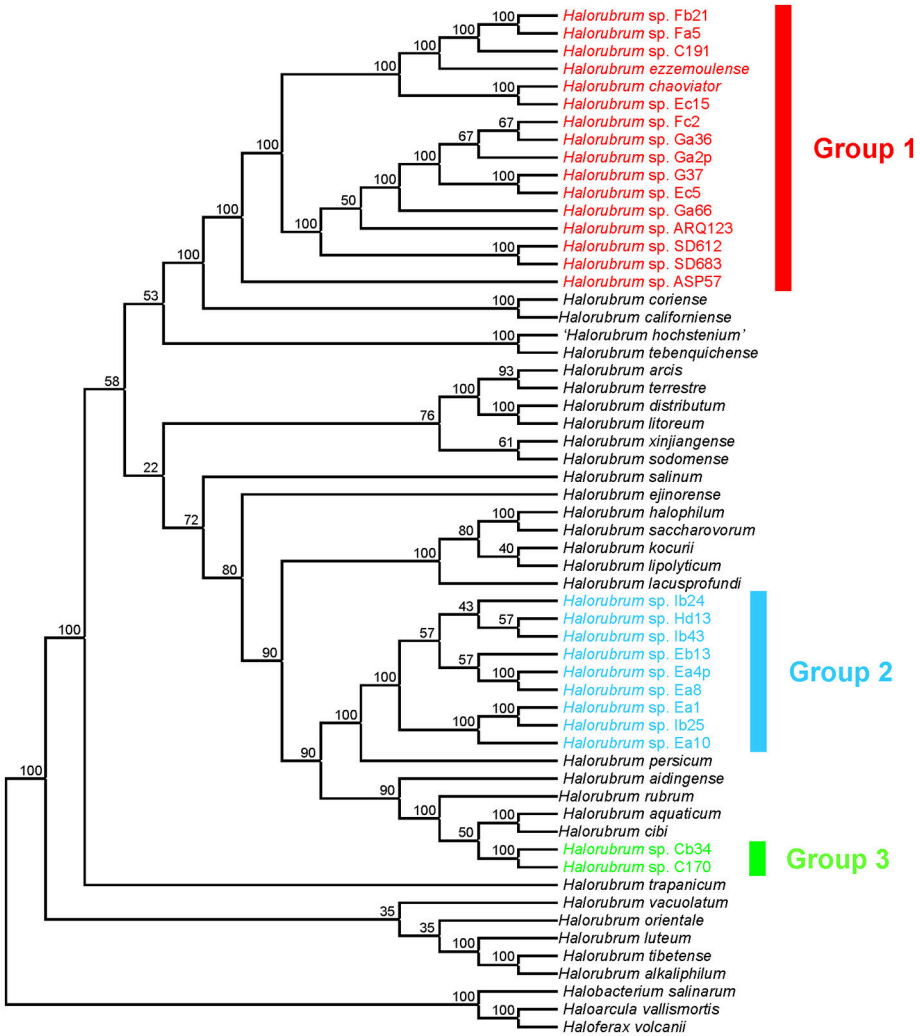


FIGURE 3 | Consensus supertree constructed on the basis of the five individual gene phylogenies by means of Matrix Representation using Parsimony (MRP) method showing the phylogenetic relationship between members of the genus *Halorubrum* and strains isolated in this study. The numbers represent the proportion of universally distributed input supertrees that contained that particular branch in the consensus supertree. The species *Haloarcula vallismortis*, *Haloferax volcanii,* and *Halobacterium salinarum* were used as outgroups. Different phylogroups have been marked with different colors.

to group 2 strains in contrast to other *Halorubrum* species (**Figures 2**, **3**). In comparison to the 16S rRNA gene sequence analysis, the MLSA approach provided clearer distinction between strains isolated in this study and previously described type strains, suggesting that groups 2 and 3 might each constitute new species. Because the MLSA analysis formed a monophyletic cluster inclusive of our group 1 new strains with two previously characterized and validly named species, *Hrr. chaoviator* Halo-G*^{T}/DSM 19316^{T} and *Hrr. ezzemoulense* DSM 17463^{T}, they could all comprise a single widely distributed species. The three strains used initially to characterize *Hrr. chaoviator* were isolated from Mexico, Australia and Greece (Mancinelli et al., 2009), and *Hrr. ezzemoulense* was cultivated from Algeria (Kharroub et al., 2006), which supports that conjecture.

Although the concatenated (**Figure 2**) and the individual gene (Figure S1) phylogenies consistently clustered strains from this study into three groups, it appeared that their relationships with respect to the other *Halorubrum* strains (type and reference strains) might be different. Therefore, we compared the topology of each phylogenetic tree using the freely available program Concaterpillar (Leigh et al., 2008), and showed that there was substantial disagreement between trees (**Table 2**), which is a common observation caused by HGT (Martens et al., 2008; Pascual et al., 2010; Papke et al., 2011; de la Haba et al., 2012). In our phylogeny of *Halorubrum* isolates, the bootstrap analysis found support primarily for groups 1, 2, 3, and a few other shallow nodes, but very little support was found for many of the deeper nodes indicating an unknown relationship between species of *Halorubrum*. This poor resolution of relationships could be caused by many processes including rampant homoplasy and saturation of homologous sites, however it is well demonstrated that *Halorubrum* specifically (Papke et al.,

TABLE 1 | Similarity values for strains within groups 1, 2, and 3 based on the concatenated and individual housekeeping gene sequence and their most closely related taxa with validly published names.

Strains		Similarity values within group (%)	Closest relative (% similarity)
Concatenated	Group 1	100–95.8	*Hrr. chaoviator/Hrr. ezzemoulense* (99.8)
	Group 2	100–99.6	*Hrr. salinum* (94.4)
	Group 3	99.7	*Hrr. aquaticum* (95.0)
atpB	Group 1	100–94.5	*Hrr. chaoviator/Hrr. ezzemoulense* (100)
	Group 2	100–99.6	*Hrr. persicum* (97.1)
	Group 3	99.4	*Hrr. salinum* (94.9)
EF-2	Group 1	100–94.7	*Hrr. chaoviator/Hrr. ezzemoulense* (100)
	Group 2	100–99.0	*Hrr. kocurii/Hrr. lipolyticum* (94.9)
	Group 3	99.4	*Hrr. aquaticum* (95.7)
glnA	Group 1	100–99.2	*Hrr. chaoviator* (100)
	Group 2	100–99.4	*Hrr. distributum* (96.7)
	Group 3	100	*Hrr. aquaticum* (96.7)
ppsA	Group 1	100–99.2	*Hrr. ezzemoulense* (99.8)
	Group 2	100–99.4	*Hrr. chaoviator* (89.7)
	Group 3	99.8	*Hrr. aquaticum* (91.4)
rpoB'	Group 1	100–90.8	*Hrr. chaoviator/Hrr. ezzemoulense* (100)
	Group 2	100–99.8	*Hrr. xinjiangense* (96.2)
	Group 3	100	*Hrr. aquaticum/Hrr. rubrum* (96.7)

TABLE 2 | Phylogenetic topology congruence analysis.

Dataset	Concatenated	*atpB*	*EF-2*	*glnA*	*ppsA*	*rpoB'*
Concatenated	1					
atpB	0.124	1				
EF-2	**0.016**	**0.001**	0.990			
glnA	**0.026**	**0.006**	**0.000**	1		
ppsA	0.455	**0.022**	**0.000**	**0.000**	0.999	
rpoB'	0.543	**0.015**	**0.000**	**0.016**	0.247	0.834

Tree comparisons with a P-value < 0.05 are incongruent and bold highlighted.

2004, 2007) and all haloarchaea in general (Sharma et al., 2007; Papke et al., 2011; Nelson-Sathi et al., 2012; DeMaere et al., 2013) undergo tremendous amounts of homologous recombination and the more closely related two cells are the more frequent the transfer is between them (Naor et al., 2012; Williams et al., 2012). Therefore we suggest that the differences in tree topology reflect gene transfer rather than other evolutionary processes.

Although concatenated sequence alignments have proven to be a very useful method of reconstructing orthologous gene phylogenies, there are limitations to this approach, e.g., assumes that the same process of evolution has been acting on all the genes in the same manner. Therefore, we combined the phylogenetic relationships from the individual protein coding gene trees into an overall consensus supertree (Sanderson et al., 1998) and obtained a single phylogeny (**Figure 3**) by means of the MRP method. This strategy allowed us to analyse the same data through a very different set of assumptions and algorithms and yet the same groups were reconstructed, including that *Hrr. chaoviator* Halo-G^{*T}/DSM 19316^{T} and *Hrr. ezzemoulense* DSM 17463^{T} clustered together with MLSA defined group 1 (**Figure 3**). Therefore, we have added evidence and confidence that each group is real, and that groups 2 and 3 may constitute new *Halorubrum* species. Interestingly, the supertree analysis distinguished subclusters that the MLSA tree did not. Consensus supertree group 1 strains forms two distinct clades each with 100% branch support. Within each, the Aran-Bidgol group 1 strains, with the lone exception of Ec15, are more closely related to each other than those from other locations, with 100% branch support.

Lipid Profiles

In the domain Archaea, polar lipid content was demonstrated to differentiate among taxa at the genus level, and sometimes at the species level (Torreblanca et al., 1986; Kates, 1993; Oren et al., 2009). In this study, we analyzed the lipid profile for strains isolated from the lake Aran-Bidgol and from the solar salterns in Namibia and Spain and their closest relatives and reference strains (*Halobacterium salinarum*, *Hrr. saccharovorum*, *Hrr. chaoviator*, *Hrr. ezzemoulense*, *Hrr. tibetense*, *Hrr. kocurii*, *Hrr. cibi* and *Natronococcus amylolyticus*) in order to further study, and possibly corroborate the obtained MLSA results. Since the lipid pattern may vary according to the culture conditions, a rigorous standardization of those conditions as well as of the starting quantity was applied to our analyses (see Methods).

HPTLC results showed that strains from the lake Aran-Bidgol and from the solar salterns in Namibia and Spain belong to the genus *Halorubrum*, showing the characteristic lipids of this genus growing optimally at neutral pH values (McGenity and Grant, 2001; Oren et al., 2009). All strains belonging to group 1 possess $C_{20}C_{20}$ and $C_{20}C_{25}$ derivates of phosphatidylglycerolphosphate methyl esther (PGP-Me), phosphatidylglycerolsulphate (PGS), $C_{20}C_{20}$ derivates of phosphatidylglycerol (PG) and biphosphatidylglycerol (BPG) as the main polar lipids. A sulphated glycolipid similar to sulphated mannosyl glycosyl diether (S-DGD) was also detected in group 1 strains and a minor co-migratory band with BPG was also present in some of these strains (**Figure 4A** and Figure S2A). Two haloalkaliphilic species, *Hrr. tibetense* JCM 11889^{T} and *Natronococcus amylolyticus* DSM 10524^{T} were included in this study with the aim of comparing the presence of the double chain length $C_{20}C_{20}$ and $C_{20}C_{25}$ derivates of PG not typically found in neutrophilic species of *Halorubrum*. The species *Hrr. chaoviator* Halo-G^{*T} and *Hrr. ezzemoulense* DSM 17463^{T}, showed lipid profiles similar to our isolated strains of group 1, but were not identical. The type strain *Hrr. chaoviator* Halo-G^{*T} presented GL1 and GL2, which is absent in all the other strains of group 1 (**Figure 4A** and Figure S2A). Additionally, all group 1 Aran-Bidgol strains lacked a minor spot at the bottom of the HPTLC

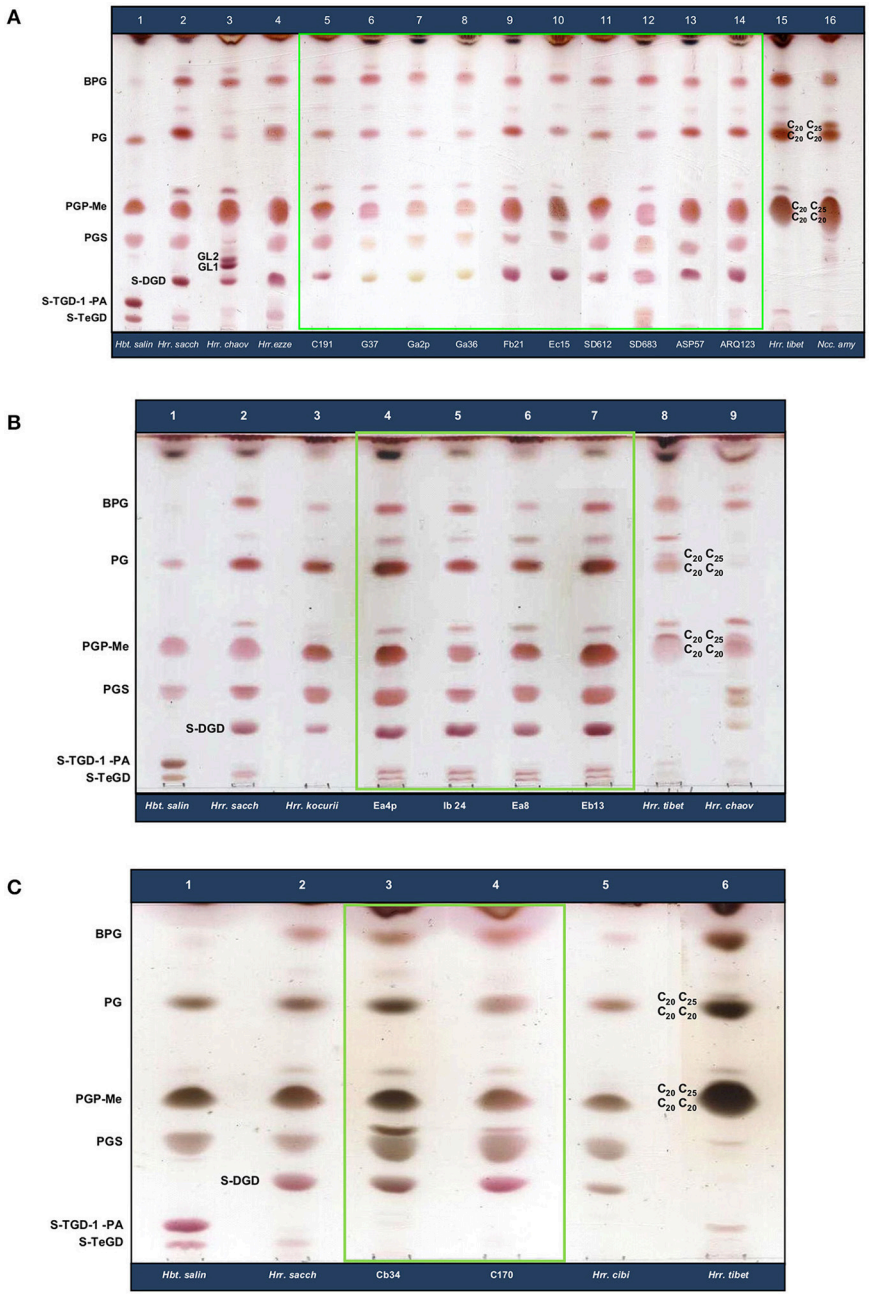


FIGURE 4 | HPTLC stained with 5% (v/v) sulfuric acid showing the polar lipid profiles for *Halorubrum* strains belonging to group 1 and related taxa **(A)**, to group 2 and related taxa **(B),** and to group 3 and related taxa **(C)**. *Hbt. salin*, *Halobacterium salinarum* DSM 3754^T; *Hrr. sacch*, *Halorubrum saccharovorum* DSM 1137^T; *Hrr. chaov*, *Halorubrum chaoviator* Halo-G*^{T}; *Hrr. ezze*, *Halorubrum ezzemoulense* DSM 17463^T; *Hrr. tibet*, *Halorubrum tibetense* JCM 11889^T; *Ncc. amy*, *Natronococcus amylolyticus* DSM 10524^T; *Hrr. kocurii*, *Halorubrum kocurii* CECT 7322^T; *Hrr. cibi*, *Halorubrum cibi* JCM 15757^T. BPG, biphosphatidylglycerol; PG, phosphatidylglycerol; PGP-Me, phosphatidylglycerolphosphate methyl ester; PGS, phosphatidylglycerolsulfate; GL, unidentified glycolipid; S-DGD, sulfated mannosyl glucosyl diether; S-TGD-1-PA, sulfated galactosyl mannosy glucosyl dietherphosphatidic acid; S-TeGD, sulfated tetraglycosyl diether.

plate that was present in the Namibian and Spanish strains as well as in the species *Hrr. chaoviator* Halo-G*^{T} and *Hrr. ezzemoulense* DSM 17463^T. These differences in lipid profiles demonstrate well the phenotypic and likely genotypic plasticity within group 1. Further, given that the group 1 Aran-Bidgol strains profiles are more similar to each other than they are to the rest of group 1, we propose that the Aran-Bidgol group 1 strains are evolving in concert, and separately from the other group 1 strains.

Lipid analysis of Aran-Bidgol strains that form group 2, and its closest relative *Hrr. kocurii* CECT 7322^T also demonstrated a clearly differentiated lipid pattern. All Aran-Bidgol group 2 strains had the same profile: a minor phosphoglycolipid below

the S-DGD spot, a minor phospholipid near the PGP-Me spot, and a minor glycolipid close to the PG spot, which are all absent in the *Hrr. kocurii* CECT 7322^T profile (**Figure 4B** and Figure S2B). Therefore, the MLSA differences observed between group 2 and its closest known validly named species *Hrr. kocurii* are further corroborated by the lipid analysis providing additional evidence that these strains likely constitute a new *Halorubrum* lineage. Nevertheless, more work needs to be done to corroborate the validity of this statement.

Both strains included in group 3 showed the same lipid pattern, which was different in comparison to their most closely related species, *Hrr. cibi* JCM 15757^T. The main difference is the presence of minor phospholipids as co-migratory bands above the PGP-Me and PGS spots, respectively, in contrast with *Hrr. cibi* JCM 15757^T, where these minor lipids cannot be observed (**Figure 4C** and Figure S2C). Again, the lipid profile agrees with the MLSA data and supports the placement of Aran-Bidgol strains forming group 3 as a different lineage than that of *Hrr. cibi* JCM 15757^T, although a more extensive analysis including the polar lipids of all the species in the genus *Halorubrum* would be required prior to become widespread the use of the lipid profile as a discriminating approach for species delineation within this genus.

DNA–DNA Hybridization and ANI

Results of DDH indicated that group 1 Aran-Bidgol strains Fb21, C191, Ec15, G37, Ga2p, and Ga36 are homogeneous with respect to their reassociation values: they are all equal to or higher than 70% in comparison to the group 1 reference strain Fb21. This result indicates they belong to the same taxonomic species (**Table 3**). The ANI values between those strains were all 96.6% or higher (**Table 3**), which is above the 95–96% cutoff limit for species delineation (Goris et al., 2007; Richter and Rosselló-Móra, 2009; Rosselló-Móra and Amann, 2015). Therefore, the DDH and ANI data are in agreement that these Aran-Bidgol strains should be classified as belonging to the same species. On the other hand, the DDH analysis does not support Aran-Bidgol group 1 strains as belonging to the previously described and closely related *Hrr. ezzemoulense* (22%) and *Hrr. chaoviator* (54%), which could not be distinguished as different species based on MLSA or ANI (**Table 3**). In order to confirm these results, the strain *Hrr. ezzemoulense* DSM 17463^T was used as reference for DDH experiments within group 1, showing 91 and 79% reassociation values with the Spanish strains ASP57 and ARQ123, 89 and 86% with the Namibian strains SD683 and SD612, and 79% with the named species *Hrr. chaoviator* HaloG*^{T}, but 26, 10 and 1% with the Aran-Bidgol strains Ga36, Ga2p and G37, respectively. The value differences between ANI and DDH likely reflect accessory genome content evolution, which is known to change quickly between closely related *Halorubrum* strains (Ram Mohan et al., 2014) and others (Welch et al., 2002; Thompson et al., 2005). Although MLSA and ANI group Aran-Bidgol, Spanish and Namibian strains, as well as *Hrr. chaoviator* Halo-G*^{T}/DSM 19316^T and *Hrr. ezzemoulense* DSM 17463^T into a single species, it is clear that the DDH results are in agreement with the polar lipid profiles that indicated Aran-Bidgol group 1 strains are homogeneous, but clearly different from *Hrr. chaoviator* Halo-G*^{T}/DSM 19316^T and *Hrr. ezzemoulense* DSM 17463^T strains.

Similar DDH results were obtained in analyzing the other strains from Aran-Bidgol: strains within the phylogenetic clusters 2 and 3 were homogeneous with respect to each other, having reassociation values higher than 70% and being unmistakably separated from closely related validly named species with values below the 70% threshold. ANI values were in agreement with the MLSA and DDH analysis for groups 2 and 3 being differentiated from all known validly named species, indicating these likely constitute new species.

Another widely accepted alternative to experimental DDH and ANI for prokaryotic species circumscription is the so-called *in silico* DDH (Auch et al., 2010). However, since ANI_b values $> 75\%$ show similar results/interpretations as *in silico* DDH (Li et al., 2015), it was not taken into consideration in the present study.

Correlation Studies

To gain a broader understanding for the different molecular techniques applied in this study, we compared the DDH data against the 16S rRNA gene sequence similarities (**Figure 5**) and against MLSA data (**Figure 6**) and calculated the correlation values. As observed in **Figure 5**, the traditional threshold value of 97% for 16S rRNA gene sequences is not useful for delineating species. All but one *Halorubrum* species having less than 70% DDH reassociation values had greater than 97% 16S rRNA gene sequence similarity, the majority had greater than 98% similarity and four had 99% or more (**Table 3**). Of particular interest was that strains exhibiting nearly 100% DDH reassociation values did not have 100% 16S rRNA gene sequence similarity (suggesting evidence for 16S rRNA gene HGT) (**Table 3**). On the other hand, comparison of DDH values and MLSA gene concatenation showed that having less than 96% sequence similarity indicates strains are different species. A stricter cutoff of 99% might be feasible based on the clustering of points above 99% sequence similarity, and absence of data between 96 and 99% for MLSA defined species (**Figure 6**). However, of special consideration is that the threshold values only demarcates unambiguously that two strains belong to separate species, which is not equivalent to saying that strains with greater than 96% (or 99%) belong to the same species.

The Pearson's product-moment correlation coefficient was calculated between the DDH relatedness matrix and the corresponding similarity matrices of the 16S rRNA gene, the five individual genes, the concatenation of the five genes and the ANI values of the sequenced genomes. Each comparison correlated between 33 and 27 pairs of values. The coefficients obtained were 0.58 for the 16S rRNA gene, and 0.57 for *ppsA*, 0.63 for *atpB*, 0.64 for *rpoB'*, 0.73 for *glnA*, and 0.74 for *EF-2*. The corresponding coefficient for the five concatenated genes was 0.70 and for the ANI value it was 0.65. Therefore, an acceptable correlation between DDH and evolutionary distances is observed, similar to those observed for *Streptomyces griseus* (Rong and Huang, 2010), *Vibrio* (Pascual et al., 2010) and *Salinivibrio* (López-Hermoso et al., 2017). However, though the correlation is acceptable, it is not as high as in those studies. One reason could be the low DDH values obtained between Fb21 and *Hrr. chaoviator* DSM 19316^T and *Hrr. ezzemoulense* DSM 17463^T. Reanalysis that excluded comparison of the two type strains with Fb21 demonstrated a rise

TABLE 3 | DNA–DNA hybridization data (%) among representative strains of groups 1 (Fb21), 2 (Ib24), and 3 (Cb34) and its closest relatives.

GROUP 1				
Competitor DNA	**DDH with strain Fb21**	**16S rRNA gene similarity (%) with Fb21**	**Five concatenated gene similarity (%) with Fb21**	**ANI_b-values (%) with Fb21**
Halorubrum sp. Fb21	100	100	100	100
Halorubrum sp. C191	71	100	100	98.3
Halorubrum sp. G37	99	99.9	99.8	97.0
Halorubrum sp. SD683	ND	99.9	98.9	97.2
Halorubrum sp. SD612	ND	99.8	99.7	97.3
Halorubrum chaoviator DSM 19316^T	54	99.8	99.8	98.5
Halorubrum sp. Ec15	75	99.6	99.7	96.8
Halorubrum sp. Ga2p	99	99.6	99.8	96.6
Halorubrum sp. Ga36	98	99.6	99.7	96.7
Halorubrum sp. ARQ123	ND	99.1	98.0	ND
Halorubrum sp. ASP57	ND	99.1	97.5	ND
Halorubrum ezzemoulense DSM 17463^T	22	99.1[a]	99.7	98.3
Halorubrum californiense CECT 7256^T/SF3 213^T/DSM 19288^T	43	99.1	94.3	87.3
Halorubrum trapanicum NRC 34021^T	3	99.0	94.4	87.2
Halorubrum coriense JCM 9275^T/Ch2^T/DSM 10284^T	21	98.9	95.4	88.0
Halorubrum sodomense JCM 8880^T/DSM 3755^T/RD 26^T	3	98.9	93.7	89.5
Halorubrum xinjiangense JCM 12388^T/CGMCC 1.3527^T	15	98.6	94.7	90.1
Halorubrum litoreum JCM 13561^T	31	98.5	95.2	87.7
Halorubrum tebenquichense CECT 5317^T/DSM 14210^T	18	98.3	94.5	85.5
Halorubrum distributum JCM 10118T/JCM 9100^T	49	98.2	95.2	87.8
Halorubrum ejinorense CECT 7194^T/EJ-32^T	38	98.1	93.5	ND
Halorubrum terrestre JCM 10247^T/VKM B-1739^T	9	97.2	95.1	87.6
Halorubrum saccharovorum DSM 1137^T/JCM 8865^T	7	95.7	92.9	81.9
GROUP 2				
Competitor DNA	**DDH with strain Ib24**	**16S rRNA gene similarity (%) with Ib24**	**Five concatenated gene similarity (%) with Ib24**	**ANI_b-values (%) with Ib24**
Halorubrum sp. Ib24	100	100	100	100
Halorubrum sp. Eb13	76	99.9	99.8	98.7
Halorubrum sp. Ea8	83	99.5	99.8	98.3
Halorubrum sp. Ea4p	73	99.5	99.8	ND
Halorubrum kocurii CECT 7322^T/JCM 14978^T	37	98.3	93.5	85.5
Halorubrum lipolyticum JCM 13559^T/9-3^T/DSM 21995^T	7	98.2	91.4	85.7
Halorubrum aidingense JCM 13560^T/31-hong^T	9	98.1	91.1	83.9
Halorubrum halophilum B8^T	ND	97.8	93.2	84.1
Halorubrum lacusprofundi JCM 8891^T/ATCC 49239^T	14	97.6	91.6	84.9
Halorubrum saccharovorum DSM 1137^T/JCM 8865^T	1	97.5	93.4	85.5
GROUP 3				
Competitor DNA	**DDH with strain Cb34**	**16S rRNA gene similarity (%) with Cb34**	**Five concatenated gene similarity (%) with Cb34**	**ANI_b-values (%) with Cb34**
Halorubrum sp. Cb34	100	100	100	100
Halorubrum sp. C170	100	99.9	99.7	ND
Halorubrum cibi JCM 15757^T/B31^T	42	98.9	ND	ND
Halorubrum aquaticum EN-2^T/CGMCC 1.6377^T	59	98.5	95.0	87.8

(Continued)

TABLE 3 | Continued

	GROUP 3			
Competitor DNA	**DDH with strain Cb34**	**16S rRNA gene similarity (%) with Cb34**	**Five concatenated gene similarity (%) with Cb34**	**ANI_b-values (%) with Cb34**
Halorubrum alkaliphilum JCM 12358^T/DZ-1^T	51	97.3	ND	ND
Halorubrum tibetense JCM 11889^T/8W8^T	56	97.3	91.3	ND
Halorubrum kocurii CECT 7322^T/JCM 14978^T	57	97.2	93.6	81.5
Halorubrum lipolyticum JCM 13559^T/9-3^T/DSM 21995^T	50	97.1	90.4	81.8

Halorubrum species for DDH experiments were selected according to their 16S rRNA gene sequence similarity (calculated using EzBioCloud) respect to the representative strain of groups 1, 2, and 3. Five concatenated gene sequence pairwise comparison values and ANI_b-values are also shown. ND, not determined.

[a] *16S rRNA gene sequence similarity calculated using BLAST and the sequence AB663412 for Hrr. ezzemoulense CECT 7099^T, since the sequence available on EzBioCloud database (accession number DQ118426) was a low quality one.*

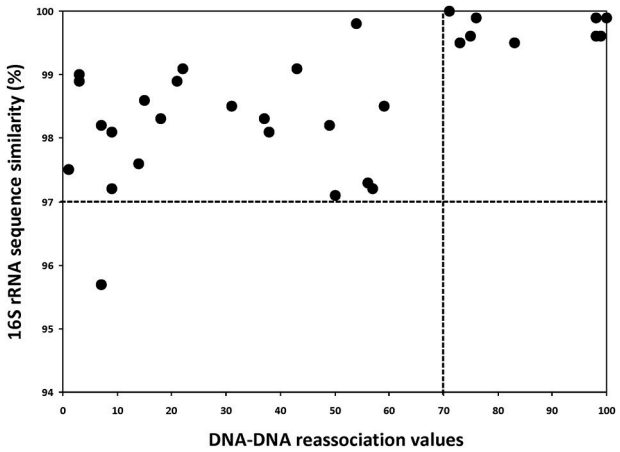


FIGURE 5 | Graph of the 16S rRNA gene sequence similarity vs. DDH relatedness values for the genus *Halorubrum*. The vertical dashed line indicates the 70% DDH threshold, while the horizontal dashed line indicates 97% 16S rRNA gene sequence similarity.

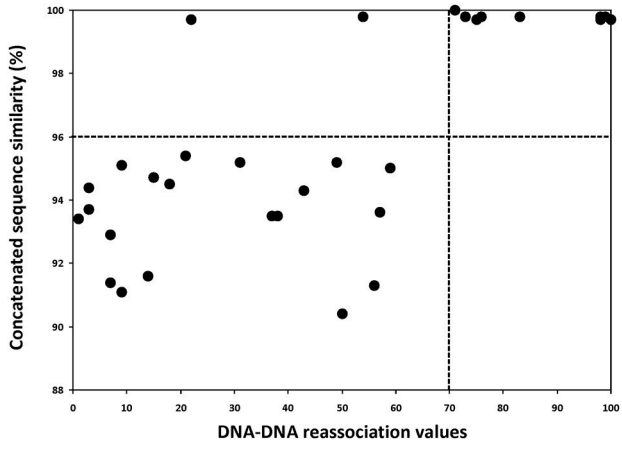


FIGURE 6 | Graph of the five concatenated gene sequence similarity vs. DDH relatedness values for the genus *Halorubrum*. The vertical dashed line indicates the 70% DDH threshold, while the horizontal dashed line indicates the proposed genetic similarity threshold for distinguishing genomic species. The two points above the 96% limit and below 70% DDH are pairwise comparisons of *Hrr. chaoviator* Halo-G*^{T}/DSM 19316^T vs. *Halorubrum* sp. Fb21 and *Hrr. ezzemoulense* DSM 17463^T vs. *Halorubrum* sp. Fb21.

in the correlation coefficient for all genes or genomes: 0.60 for 16S rRNA gene, 0.60 for *ppsA*, 0.69 for *atpB*, 0.70 for *rpoB'*, 0.80 for *glnA*, 0.83 for *EF-2*, and 0.77 and 0.73 for the concatenated sequences and the ANI values, respectively.

CONCLUSIONS

In this study we used an MLSA approach to infer phylogenetic relationship among the species of the genus *Halorubrum* and we have applied this scheme to analyze many new isolates. The MLSA results were complemented with 16S rRNA gene sequencing, lipid profiling, DDH, and ANI analyses. While it is clear that all data presented are in agreement regarding groups 2 and 3 as belonging to new species, they were tentative for group 1. On the basis of MLSA and ANI results, it is demonstrated that all isolated strains of group 1 (from Aran-Bidgol lake and Namibian and Spanish solar salterns, and the previously named species *Hrr. chaoviator* Halo-G*^{T}/DSM 19316^T and *Hrr. ezzemoulense* DSM 17463^T) constitute a single species. A lack of clarity develops from the observed DDH values and lipid profiling which indicated that group 1 Aran-Bidgol strains are all more similar to each other, yet different from the other group 1 strains, and could belong to separate species if DDH and PLP analyses have more weight than sequence data. Historically and contemporaneously, DDH is the gold standard for species inclusion/exclusion, and polar lipids provide a diagnostic phenotype to separate them, indicating the MLSA defined and ANI validated group 1 is likely composed of more than one taxonomic species.

Uniqueness in taxonomic classification likely hinges on the consideration of evolutionary forces that homogenize diversity within populations, and that generate diversity between them. All of the Aran-Bidgol strains forming group 1 are genetically homogeneous according to their MLSA, ANI, DDH and lipids, while strains from the same MLSA cluster but cultivated from Mexico, Algeria, Namibia and Spain, demonstrate clear differences. This indicates that *Hrr. ezzemoulense*, *Hrr. chaviator* and strains from Namibia (SD612 and SD683) and Spain (ARQ123 and ASP57) are not undergoing the same homogenizing forces as the population from Aran-Bidgol. Based on previous analyses of gene flow in *Halorubrum* spp. and other haloarchaea (Papke et al., 2011; Naor et al., 2012;

Nelson-Sathi et al., 2012; Williams et al., 2012; DeMaere et al., 2013), HGT is a strong candidate for the evolutionary forces homogenizing the Aran-Bidgol groups. Since *Hrr. ezzemoulense*, *Hrr. chaviator*, strains SD612 and SD683 and strains ARQ123 and ASP57 were each cultivated from different locations this may not be surprising, as hypersaline environments are patchily distributed microbial "Galapagos" islands, that the differences observed may reflect "slow" migration between sites relative to the rate of evolutionary change. Previous analyses of the Aran-Bidgol strains support our hypothesis: they are losing and adding variability through gene loss and HGT possibly at rates faster than the accumulation of substitutions at redundant codon positions in MLSA genes (Fullmer et al., 2014; Ram Mohan et al., 2014). Therefore, our data support the interpretation that there is an absence of a worldwide homogenizing force, and separated populations are able to acquire localized variation and diverge, which promotes speciation. Additional data supporting our hypothesis is that the three strains used to characterize *Hrr. chaoviator*, which were cultivated from three different geographic locations, also had different lipid profiles (Mancinelli et al., 2009). Our observations also indicate the capacity for rapid adaptation to a new location once arrived, as our lipids analysis shows high variability among cluster 1 strains that were cultivated from different locations. Limitations to haloarchaeal dispersal and invasion are further validated by recent analyses showing that different locations distributed around the globe, and even ones within the same country, have substantially different community compositions (Oh et al., 2010; Dillon et al., 2013; Zhaxybayeva et al., 2013; Fernández et al., 2014b). Our data are in agreement with those studies, indicating that limitations to dispersal and invasiveness may be primary drivers of haloarchaeal divergence and speciation.

There has long been a debate in the taxonomy and systematics fields over whether or not multiple strains should be required for classifying new species. It is worthwhile pointing out that our conclusions would be dampened immensely if we had not analyzed many closely related strains: the homogeneity observed in the Aran-Bidgol group 1 population would have gone undetected, or might have been considered a one-off result. Our population data showing that all the group 1 Aran-Bidgol strains were similar to each other and different from their closely related kin reinforced our conclusions regarding their differences and amplifying the importance of using multiple strains. Including evidence from natural populations to guide taxonomy appears to boost the robustness and accuracy in classification and should be considered as important as the methodological aspects to classification.

Our study establishes the usefulness of an MLSA approach for distinguishing between species within the genus *Halorubrum*. We were able to demonstrate a 4% MLSA sequence divergence cutoff correlates with the 70% DDH gold standard of prokaryotic taxonomy, thus reducing the necessity for performing DDH in future taxonomic studies: in the case of strains obtained from different geographic locations with high sequence similarity it appears DDH and other analyses would still provide useful taxonomic insight. The advantages of using an MLSA approach over DDH are enormous: sequence data can be stored for all downstream discoveries of new species; it is less prone to error; it does not require radioisotopes or other tedious methodologies; and it can provide useful evolutionary relationships. While 16S rRNA gene sequence analysis is storable and capable of phylogenetic reconstruction in the haloarchaea, we have shown conclusively that a follow up DDH analysis is nearly certainly needed to distinguish any new species. This does not appear to be the case for the MLSA. Further, several genera of the haloarchaea (e.g., *Halomicrobium, Haloarcula, Halosimplex*, *Halomicroarcula*, *Haloarchaeobius*) carry multiple highly divergent rRNA operons and are known to be extremely prone to PCR artifacts and thus erroneous identification and classification (Boucher et al., 2004; Zhang and Cui, 2014a,b). Therefore, in agreement with the recommendations of the ICSP-Subcommittee on the taxonomy of *Halobacteria* (Oren and Ventosa, 2013, 2016), we strongly encourage the use of our MLSA approach for descriptions of novel species within this family. For practical matters of classification, we propose the 4% MLSA nucleotide sequence dissimilarity threshold value for unequivocally distinguishing new genomic species for the haloarchaea, with the caveat that strains having less concatenated sequence divergence could also be different species and would need to be tested by DDH or genomic indexes for certainty.

AUTHOR CONTRIBUTIONS

AV, RP, CS-P, and RH conceived and designed the study. PC, CI-D, CS-P, MA, AV, and RP designed and performed the acquisition of environmental isolates. RH, PC, CS-P, CI-D, and AM performed the microbial experiments. RH, PC, CS-P, CI-D, AM, AV, and RP analyzed and interpreted the data. RH, PC, CS-P, CI-D, AM, MA, AV, and RP discussed the paper. RH and PC drafted the paper. RH, PC, CS-P, CI-D, AM, MA, AV, and RP critically revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by grants from National Science Foundation (Grants DEB-0919290 and 0830024), NASA Astrobiology: Exobiology and Evolutionary Biology Program Element (Grant Number NNX12AD70G and NNX15AM09G), Spanish Ministry of Economy and Competitiveness (MINECO) (Projects CGL2013-46941-P and CGL2017-83385-P). FEDER funds also supported this project. RH was recipient of the short-stay grant José Castillejo for young Ph.D.s from the Spanish Ministry of Education, Culture and Sport (CAS14/00289).

ACKNOWLEDGMENTS

We wish to thank the reviewers for their generous and meaningful efforts.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00512/full#supplementary-material

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Amoozegar, M. A., Siroosi, M., Atashgahi, S., Smidt, H., and Ventosa, A. (2017). Systematics of haloarchaea and biotechnological potential of their hydrolytic enzymes. *Microbiology* 163, 623–645. doi: 10.1099/mic.0.000463

Angelini, R., Corral, P., Lopalco, P., Ventosa, A., and Corcelli, A. (2012). Novel ether lipid cardiolipins in archaeal membranes of extreme haloalkaliphiles. *Biochim. Biophys. Acta* 1818, 1365–1373. doi: 10.1016/j.bbamem.2012.02.014

Arahal, D. R., Dewhirst, F. E., Paster, B. J., Volcani, B. E., and Ventosa, A. (1996). Phylogenetic analyses of some extremely halophilic archaea isolated from Dead Sea water, determined on the basis of their 16S rRNA sequences. *Appl. Environ. Microbiol.* 62, 3779–3786.

Auch, A. F., von Jan, M., Klenk, H. P., and Göker, M. (2010). Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Baum, B. R. (1992). Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41, 3–10. doi: 10.2307/1222480

Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37, 911–917. doi: 10.1139/y59-099

Boucher, Y., Douady, C. J., Sharma, A. K., Kamekura, M., and Doolittle, W. F. (2004). Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol.* 186, 3980–3990. doi: 10.1128/JB.186.12.3980-3990.2004

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijs.0.054171-0

Corcelli, A., Colella, M., Mascolo, G., Fanizzi, F. P., and Kates, M. (2000). A novel glycolipid and phospholipid in the purple membrane. *Biochemistry* 39, 3318–3326. doi: 10.1021/bi992462z

Corral, P., de la Haba, R. R., Sánchez-Porro, C., Amoozegar, M. A., Papke, R. T., and Ventosa, A. (2015). *Halorubrum persicum* sp. nov., an extremely halophilic archaeon isolated from sediment of a hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 65, 1770–1778. doi: 10.1099/ijs.0.000175

Creevey, C. J., and McInerney, J. O. (2005). Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* 21, 390–392. doi: 10.1093/bioinformatics/bti020

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109

de la Haba, R. R., Márquez, M. C., Papke, R. T., and Ventosa, A. (2012). Multilocus sequence analysis of the family *Halomonadaceae*. *Int. J. Syst. Evol. Microbiol.* 62, 520–538. doi: 10.1099/ijs.0.032938-0

De Ley, J., and Tijtgat, R. (1970). Evaluation of membrane filter methods for DNA-DNA hybridization. *Antonie van Leeuwenhoek* 36, 461–474. doi: 10.1007/BF02069048

DeMaere, M. Z., Williams, T. J., Allen, M. A., Brown, M. V., Gibson, J. A., Rich, J., et al. (2013). High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16939–16944. doi: 10.1073/pnas.1307090110

Dillon, J. G., Carlin, M., Gutierrez, A., Nguyen, V., and McLain, N. (2013). Patterns of microbial diversity along a salinity gradient in the Guerrero Negro solar saltern, Baja CA Sur, Mexico. *Front. Microbiol.* 4:399. doi: 10.3389/fmicb.2013.00399

Edgar, R. C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Edgar, R. C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Enache, M., Itoh, T., Fukushima, T., Usami, R., Dumitru, L., and Kamekura, M. (2007). Phylogenetic relationships within the family *Halobacteriaceae* inferred from *rpoB*' gene and protein sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2289–2295. doi: 10.1099/ijs.0.65190-0

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359

Fernández, A. B., Ghai, R., Martin-Cuadrado, A. B., Sánchez-Porro, C., Rodriguez-Valera, F., and Ventosa, A. (2014a). Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* 88, 623–635. doi: 10.1111/1574-6941.12329

Fernández, A. B., Vera-Gargallo, B., Sánchez-Porro, C., Ghai, R., Papke, R T., Rodríguez-Valera, F., et, al. (2014b). Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front. Microbiol.* 5:196. doi: 10.3389/fmicb.2014.00196

Fullmer, M. S., Soucy, S. M., Swithers, K. S., Makkay, A. M., Wheeler, R., Ventosa, A., et al. (2014). Population and genomic analysis of the genus *Halorubrum*. *Front. Microbiol.* 5:140. doi: 10.3389/fmicb.2014.00140

Ghai, R., Pašić, L., Fernández, A. B., Martin-Cuadrado, A. B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Gupta, R. S., Naushad, S., Fabros, R., and Adeolu, M. (2016). A phylogenomic reappraisal of family-level divisions within the class *Halobacteria*: proposal to divide the order *Halobacteriales* into the families *Halobacteriaceae*, *Haloarculaceae* fam. nov., and *Halococcaceae* fam. nov., and the order *Haloferacales* into the families, *Haloferacaceae* and *Halorubraceae* fam nov. *Antonie van Leeuwenhoek* 109, 565–587. doi: 10.1007/s10482-016-0660-2

Hanage, W. P., Fraser, C., and Spratt, B. G. (2006). Sequences, sequence clusters and bacterial species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1917–1927. doi: 10.1098/rstb.2006.1917

Johnson, J. L. (1994). "Similarity analysis of rRNAs," in *Methods for General and Molecular Bacteriology*, eds. P. Gerhardh, R. G. E. Murray, W. A. Wood, and N. R. Krieg (Washington DC, USA: American Society for Microbiology), 683–700.

Kates, M. (1993). Biology of halophilic bacteria, part II. *Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance. Experientia* 49, 1027–1036. doi: 10.1007/BF01929909

Kharroub, K., Quesada, T., Ferrer, R., Fuentes, S., Aguilera, M., Boulahrouf, A., et al. (2006). *Halorubrum ezzemoulense* sp. nov., a halophilic archaeon isolated from Ezzemoul sabkha, Algeria. *Int. J. Syst. Evol. Microbiol.* 56, 1583–1588. doi: 10.1099/ijs.0.64272-0

Kluge, A. G., and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18, 1–32. doi: 10.2307/2412407

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Leigh, J. W., Susko, E., Baumgartner, M., and Roger, A. J. (2008). Testing congruence in phylogenomic analysis. *Syst. Biol.* 57, 104–115. doi: 10.1080/10635150801910436

Li, X., Huang, Y., and Whitman, W. B. (2015). The relationship of the whole genome sequence identity to DNA hybridization varies between genera of prokaryotes. *Antonie van Leeuwenhoek* 107, 241–249. doi: 10.1007/s10482-014-0322-1

Loomis, W. F., and Smith, D. W. (1990). Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9093–9097. doi: 10.1073/pnas.87.23.9093

López-García, P., López-López, A., Moreira, D., and Rodríguez-Valera, F. (2001). Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol. Ecol.* 36, 193–202. doi: 10.1016/S0168-6496(01)00133-7

López-Hermoso, C., de la Haba, R. R., Sánchez-Porro, C., Papke, R. T., and Ventosa, A. (2017). Assessment of MultiLocus Sequence Analysis as a valuable tool for the classification of the genus *Salinivibrio*. *Front. Microbiol.* 8:1107. doi: 10.3389/fmicb.2017.01107

Ludwig, W., Strunk, O., Westram, R., Richter, L., and Meier, H., Yadhukumar, et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

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

Maddison, W. P., and Maddison, D. R. (2011). *Mesquite: A Modular System for Evolutionary Analysis. Version 2.75*. Available online at: http://mesquiteproject.org.

Makhdoumi-Kakhki, A., Amoozegar, M. A., Kazemi, B., Pašić, L., and Ventosa, A. (2012). Prokaryotic diversity in Aran-Bidgol salt lake, the largest hypersaline playa in Iran. *Microbes Environ.* 27, 87–93. doi: 10.1264/jsme2.ME11267

Mancinelli, R. L., Landheim, R., Sánchez-Porro, C., Dornmayr-Pfaffenhuemer, M., Gruber, C., Legat, A., et al. (2009). *Halorubrum chaoviator* sp. nov., a haloarchaeon isolated from sea salt in Baja California, Mexico, Western Australia and Naxos, Greece. *Int. J. Syst. Evol. Microbiol.* 59, 1908–1913. doi: 10.1099/ijs.0.000463-0

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P., and Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *Int. J. Syst. Evol. Microbiol.* 58, 200–214. doi: 10.1099/ijs.0.65392-0

Martinez-Murcia, A. J., Monera, A., Saavedra, M. J., Oncina, R., Lopez-Alvarez, M., Lara, E., et al. (2011). Multilocus phylogenetic analysis of the genus *Aeromonas*. *Syst. Appl. Microbiol.* 34, 189–199. doi: 10.1016/j.syapm.2010.11.014

McGenity, T. J., and Grant, W. D. (1995). Transfer of *Halobacterium saccharovorum, Halobacterium sodomense, Halobacterium trapanicum* NRC 34041 and *Halobacterium lacusproundi* to the genus *Halorubrum* gen. nov., as *Halorubrum saccharovorum* comb. nov., *Halorubrum sodomense* comb. nov., *Halorubrum trapanicum* comb. nov., and *Halorubrum lacusprofundi* comb. nov. *Syst. Appl. Microbiol.* 18, 237–324. doi: 10.1016/S0723-2020(11)80394-2

McGenity, T. J., and Grant, W. D. (2001). "Genus VII. *Halorubrum*," in *Bergey's Manual of Systematic Bacteriology, 2nd Edn.*, eds. D. R. Boone, R. W. Castenholz, and G. M. Garrity (New York, NY: Springer), 320–324.

Mignard, S., and Flandrois, J. P. (2008). A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. *Int. J. Syst. Evol. Microbiol.* 58, 1432–1441. doi: 10.1099/ijs.0.65658-0

Minegishi, H., Kamekura, M., Itoh, T., Echigo, A., Usami, R., and Hashimoto, T. (2010). Further refinement of the phylogeny of the *Halobacteriaceae* based on the full-length RNA polymerase subunit B' (*rpoB'*) gene. *Int. J. Syst. Evol. Microbiol.* 60, 2398–2408. doi: 10.1099/ijs.0.017160-0

Naor, A., Lapierre, P., Mevarech, M., Papke, R. T., and Gophna, U. (2012). Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22, 1444–1448. doi: 10.1016/j.cub.2012.05.056

Naser, S. M., Hagen, K. E., Vancanneyt, M., Cleenwerck, I., Swings, J., and Tompkins, T. A. (2006). *Lactobacillus suntoryeus* Cachat and Priest 2005 is a later synonym of *Lactobacillus helveticus* (Orla-Jensen 1919) Bergey *et al.* 1925 (Approved Lists 1980). *Int. J. Syst. Evol. Microbiol.* 56, 355–360. doi: 10.1099/ijs.0.64001-0

Nelson-Sathi, S., Dagan, T., Landan, G., Janssen, A., Steel, M., McInerney, J. O., et al. (2012). Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20537–20542. doi: 10.1073/pnas.1209119109

Oh, D., Porter, K., Russ, B., Burns, D., and Dyall-Smith, M. (2010). Diversity of *Haloquadratum* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* 14, 161–169. doi: 10.1007/s00792-009-0295-6

Oren, A., Arahal, D. R., and Ventosa, A. (2009). Emended descriptions of genera of the family *Halobacteriaceae*. *Int. J. Syst. Evol. Microbiol.* 59, 637–642. doi: 10.1099/ijs.0.008904-0

Oren, A, and Ventosa, A. (2013). Subcommittee on the taxonomy of *Halobacteriaceae* and Subcommittee on the taxonomy of *Halomonadaceae*: minutes of the joint open meeting, 24 June 2013, Storrs, Connecticut, USA. *Int. J. Syst. Evol. Microbiol.* 63, 3540–3544. doi: 10.1099/ijs.0.055988-0

Oren, A., and Ventosa, A. (2016). International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of *Halobacteriaceae* and Subcommittee on the taxonomy of *Halomonadaceae*. Minutes of the joint open meeting, 23 May 2016, San Juan, Puerto Rico. *Int. J. Syst. Evol. Microbiol.* 66, 4291–4295. doi: 10.1099/ijsem.0.001282

Papke, R. T. (2009). A critique of prokaryotic species concepts. *Methods Mol. Biol.* 532, 379–395. doi: 10.1007/978-1-60327-853-9_22

Papke, R. T., Koenig, J. E., Rodríguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929. doi: 10.1126/science.1103289

Papke, R. T., White, E., Reddy, P., Weigel, G., Kamekura, M., Minegishi, H., et al. (2011). A multilocus sequence analysis (MLSA) approach to *Halobacteriales* phylogeny and taxonomy. *Int. J. Syst. Evol. Microbiol.* 61, 2984–2995. doi: 10.1099/ijs.0.029298-0

Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., and Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14092–14097. doi: 10.1073/pnas.0706358104

Parte, A. C. (2018). *List of Prokaryotic Names with Standing in Nomenclature*. Available online at: http://www.bacterio.net.

Pascual, J., Macián, M. C., Arahal, D. R., Garay, E., and Pujalte, M. J. (2010). Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR* genes. *Int. J. Syst. Evol. Microbiol.* 60, 154–165. doi: 10.1099/ijs.0.010702-0

Pesenti, P. T., Sikaroodi, M., Gillevet, P. M., Sánchez-Porro, C., Ventosa, A., and Litchfield, C. D. (2008). *Halorubrum californiense* sp. nov., an extreme archaeal halophile isolated from a crystallizer pond at a solar salt plant in California, USA. *Int. J. Syst. Evol. Microbiol.* 58, 2710–2715. doi: 10.1099/ijs.0.2008/002410-0

Posada, D. (2003). Using MODELTEST and PAUP* to select a model of nucleotide substitution. *Curr. Protoc. Bioinformatics* Chapter 6:Unit 6.5. doi: 10.1002/0471250953.bi0605s00

Ragan, M. A. (1992). Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* 1, 53–58. doi: 10.1016/1055-7903(92)90035-F

Ram Mohan, N., Fullmer, M. S., Makkay, A. M., Wheeler, R., Ventosa, A., Naor, A., et al. (2014). Evidence from phylogenetic and genome fingerprinting analyses suggests rapidly changing variation in *Halorubrum* and *Haloarcula* populations. *Front. Microbiol.* 5:143. doi: 10.3389/fmicb.2014.00143

Richter, D., Postic, D., Sertour, N., Livey, I., Matuschka, F. R., and Baranton, G. (2006). Delineation of *Borrelia burgdorferi sensu lato* species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp. *nov. Int. J. Syst. Evol. Microbiol.* 56, 873–881. doi: 10.1099/ijs.0.64050-0

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rodríguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1980). Isolation of extremely halophilic bacteria able to grow in defined inorganic media with single carbon sources. *J. Gen. Microbiol.* 119, 535–538. doi: 10.1099/00221287-119-2-535

Roger, F., Marchandin, H., Jumas-Bilak, E., Kodjo, A., colBVH study group, and Lamy, B. (2012). Multilocus genetics to reconstruct aeromonad evolution. *BMC Microbiol.* 12:62. doi: 10.1186/1471-2180-12-62

Rong, X., and Huang, Y. (2010). Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. *Int. J. Syst. Evol. Microbiol.* 60, 696–703. doi: 10.1099/ijs.0.012419-0

Rosselló-Móra, R., and Amann, R. (2015). Past and future species definitions for *Bacteria* and *Archaea*. *Syst. Appl. Microbiol.* 38, 209–216. doi: 10.1016/j.syapm.2015.02.001

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Sanderson, M. J., Purvis, A., and Henze, C. (1998). Phylogenetic supertrees: assembling the trees of life. *Trends Ecol. Evol. (Amst).* 13, 105–109. doi: 10.1016/S0169-5347(97)01242-1

Sandler, S. J., Hugenholtz, P., Schleper, C., DeLong, E. F., Pace, N. R., and Clark, A. J. (1999). Diversity of *radA* genes from cultured and uncultured archaea: comparative analysis of putative RadA proteins and their use as a phylogenetic marker. J. *Bacteriol.* 181, 907–915.

Sharma, A. K., Walsh, D. A., Bapteste, E., Rodriguez-Valera, F., Doolittle, W. F., and Papke, R. T. (2007). Evolution of rhodopsin ion pumps in haloarchaea. *BMC Evol. Biol.* 7:79. doi: 10.1186/1471-2148-7-79

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A, Kämpfer, P., Maiden, M. C., et al. (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047. doi: 10.1099/ijs.0.02360-0

Sun, D. L., Jiang, X., Wu, Q. L., and Zhou, N. Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl. Environ. Microbiol.* 79, 5962–5969. doi: 10.1128/AEM.01282-13

Swofford, D. L. (2003). *PAUP*: Phylogenetic Analysis Using Parsimony, Version 4.0b10*. Sunderland, MA: Sinauer Associates.

Tavare, S. (1986). "Some probablilistic and statistical problems in the anlysis of DNA sequences," in *Some Mathematical Questions in Biology*, ed. R. M. Miura (Providence, RI: American Mathematical Society), 57–86.

Thompson, J. R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D. E., Benoit, J., et al. (2005). Genotypic diversity within a natural coastal bacterioplankton population. *Science* 307, 1311–1313. doi: 10.1126/science.1106028

Torreblanca, M., Rodriguez-Valera, F., Juez, G., Ventosa, A., Kamekura, M., and Kates, M. (1986). Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid composition, and description of *Haloarcula* gen. nov. and *Haloferax* gen. nov. *Syst. Appl. Microbiol.* 8, 89–99. doi: 10.1016/S0723-2020(86)80155-2

Ventosa, A., de la Haba, R. R., Sánchez-Porro, C., and Papke, R. T. (2015). Microbial diversity of hypersaline environments: a metagenomic approach. *Curr. Opin. Microbiol.* 25, 80–87. doi: 10.1016/j.mib.2015.05.002

Ventosa, A., Quesada, E., Rodriguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1982). Numerical taxonomy of moderately halophilic Gram-negative rods. *J. Gen. Microbiol.* 128, 1959–1968. doi: 10.1099/00221287-128-9-1959

Walsh, D. A., Bapteste, E., Kamekura, M., and Doolittle, W. F. (2004). Evolution of the RNA polymerase B' subunit gene (*rpoB'*) in *Halobacteriales*: a complementary molecular marker to the SSU rRNA gene. *Mol. Biol. Evol.* 21, 2340–2351. doi: 10.1093/molbev/msh248

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, L., et al. (1987). Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37, 463–464. doi: 10.1099/00207713-37-4-463

Welch, R. A., Burland, V., Plunkett, G. III, Redford, P., Roesch, P., Rasko, D., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17020–17024. doi: 10.1073/pnas.252529799

Williams, D., Gogarten, J. P., and Papke, R. T. (2012). Quantifying homologous replacement of loci between haloarchaeal species. *Genome Biol. Evol.* 4, 1223–1244. doi: 10.1093/gbe/evs098

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Young, J. M., and Park, D. C. (2007). Probable synonymy of the nitrogen-fixing genus *Azotobacter* and the genus *Pseudomonas*. *Int. J. Syst. Evol. Microbiol.* 57, 2894–2901. doi: 10.1099/ijs.0.64969-0

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int. J. Syst. Evol. Microbiol.* 53, 1893–1900. doi: 10.1099/ijs.0.02713-0

Zhang, W. J., and Cui, H. L. (2014a). *Haloarchaeobius litoreus* sp. nov., isolated from a marine solar saltern. *Antonie van Leeuwenhoek* 105, 1085–1090. doi: 10.1007/s10482-014-0166-8

Zhang, W. J., and Cui, H. L. (2014b). *Halomicroarcula limicola* sp. nov., isolated from a marine solar saltern, and emended description of the genus *Halomicroarcula*. *Int. J. Syst. Evol. Microbiol.* 64, 1747–1751. doi: 10.1099/ijs.0.062455-0

Zhaxybayeva, O., Stepanauskas, R., Mohan, N. R., and Papke, R. T. (2013). Cell sorting analysis of geographically separated hypersaline environments. *Extremophiles* 17, 265–275. doi: 10.1007/s00792-013-0514-z

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 de la Haba, Corral, Sánchez-Porro, Infante-Domínguez, Makkay, Amoozegar, Ventosa and Papke. 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: 12 January 2018
doi: 10.3389/fmicb.2017.02645

Comparative Genomics of *Thalassobius* Including the Description of *Thalassobius activus* sp. nov., and *Thalassobius autumnalis* sp. nov.

María J. Pujalte, Teresa Lucena, Lidia Rodrigo-Torres and David R. Arahal*

Departamento de Microbiología and Colección Española de Cultivos Tipo, Universitat de València, Valencia, Spain

A taxogenomic study was conducted to describe two new *Thalassobius* species and to analyze the internal consistency of the genus *Thalassobius* along with *Shimia* and *Thalassococcus*. Strains CECT 5113^T, CECT 5114, CECT 5118^T, and CECT 5120 were isolated from coastal Mediterranean seawater, Spain. Cells were Gram-negative, non- motile coccobacilli, aerobic chemoorganotrophs, with an optimum temperature of 26°C and salinity of 3.5–5%. Major cellular fatty acids of strains CECT 5113^T and CECT 5114 were $C_{18:1}$ ω7c/ω6c and $C_{10:0}$ 3OH, G+C content was 54.4–54.5 mol% and were able to utilize propionate, L-threonine, L- arginine, and L-aspartate as carbon sources. They exhibited 98.3% 16S rRNA gene sequence similarity, 75.0–75.1 ANIb and 19.5–20.9 digital DDH to type strain of their closest species, *Thalassobius maritimus*. Based on these data, strains CECT 5113^T and CECT 5114 are recognized as a new species, for which the name *Thalassobius activus* is proposed, with strain CECT 5113^T (=LMG 29900^T) as type strain. Strains CECT 5118^T and CECT 5120 were found to constitute another new species, with major cellular fatty acids $C_{18:1}$ ω7c/ω6c and $C_{18:1}$ ω7c 11-methyl and a G+C content of 59.8 mol%; they were not able to utilize propionate, L-threonine, L- arginine or L-aspartate. Their closest species was *Thalassobius mediterraneus,* with values of 99.6% 16S rRNA gene sequence similarity, 79.1% ANIb and 23.2% digital DDH compared to the type strain, CECT 5383^T. The name *Thalassobius autumnalis* is proposed for this second new species, with strain CECT 5118^T (=LMG 29904^T) as type strain. To better determine the phylogenetic relationship of the two new species, we submitted 12 genomes representing species of *Thalassobius, Shimia,* and *Thalassoccocus,* to a phylogenomic analysis based on 54 single protein-encoding genes (BCG54). The resulting phylogenomic tree did not agree with the current genera classification, as *Thalassobius* was divided in three clades, *Thalassobius* sensu stricto (*T. mediterraneus*, *T. autumnalis* sp. nov., and *T. gelatinovorus*), *Thalassobius aestuarii* plus the three *Shimia* spp (*S. marina, S. haliotis,* and *Shimia* sp. SK013) and finally, *Thalasobius maritimus* plus *T. activus* sp. nov. *Thalassococcus halodurans* remained apart from the two genera. Phenotypic inferences from explored genomes are presented.

Keywords: *Thalassobius*, taxogenomics, phylogenomics, *Rhodobacteraceae*, *Roseobacter* group, *Shimia*, *Thalassococcus*

OPEN ACCESS

Edited by:
Sabela Balboa, Universidade de Santiago de Compostela, Spain

Reviewed by:
Mohammad Ali Amoozegar, University of Tehran, Iran
Maher Gtari, Carthage University, Tunisia

****Correspondence:***
María J. Pujalte
maria.j.pujalte@uv.es

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

Received: *13 October 2017*
Accepted: *19 December 2017*
Published: *12 January 2018*

Citation:
Pujalte MJ, Lucena T, Rodrigo-Torres L and Arahal DR (2018) Comparative Genomics of Thalassobius Including the Description of Thalassobius activus sp. nov., and Thalassobius autumnalis sp. nov. Front. Microbiol. 8:2645. doi: 10.3389/fmicb.2017.02645

INTRODUCTION

The genus *Thalassobius* was established to accommodate the species *Thalassobius mediterraneus* and the reclassified *Thalassobius gelatinovorus* (formerly *Ruegeria gelatinovorans*) by Arahal et al. (2005). It is affiliated to the *Roseobacter* group, in the family *Rhodobacteraceae*, class *Alphaproteobacteria* (Pujalte et al., 2014). Since the description of the first pair of species, five more have been added: *Thalassobius aestuarii* (Yi and Chun, 2006), *Thalassobius maritimus* (Park et al., 2012), *Thalassobius aquaeponti* (Park et al., 2014), *Thalassobius abyssi* (Nogi et al., 2016), and *Thalassobius litorarius* (Park et al., 2016). All species so far characterized are aerobic chemoorganotrophic marine bacteria able to accumulate polyhydroxybutyrate (PHB). They have been isolated from marine environments, particularly surface coastal seawater and tidal flat samples, but one species (*T. abyssi*) was isolated from deep seawater (around 1,000 m depth). Strains identified as *Thalassobius* sp. on the basis of 16S rRNA gene sequence have been isolated and reported from diverse marine samples, including corals, mollusks, sand, microalgal cultures, and mariculture samples of different types, and show a wide geographic distribution (from temperate zones such as Kuwait coast to polar regions, as Antarctica or Norwegian subartic fjords). Most species show complex ionic requirements, as they require seawater based media and are unable to grow in media containing only NaCl or just a combination of NaCl with calcium, magnesium or potassium salts. None of the species produces *BChl a* or synthetizes carotenoid pigments.

The *Roseobacter* group, which comprises a very large number of genera (100 at the time of the writing), expands at a quick rate. The phylogenetic position of the genus *Thalassobius* in the group has been addressed with 16S rRNA gene sequence comparisons: *Thalassobius* species form a clade in the phylogenetic trees, but it frequently incorporates sequences of *Thalassococcus* (LTP128, https://www.arb-silva.de/projects/living-tree/) or *Shimia* species (Pujalte et al., 2014). *Shimia* includes five species, *Shimia marina* (Choi and Cho, 2006), *Shimia isoporae* (Chen et al., 2011), *Shimia haliotis* (Hyun et al., 2013), *Shimia biformata* (Hameed et al., 2013), and *Shimia sagamensis* (Nogi et al., 2015) while *Thalassococcus* includes two, *Thalassococcus halodurans* (Lee et al., 2007) and *Thalassococcus lentus* (Park et al., 2013). The relationships of *Thalassobius* to both genera (*Shimia* and *Thalassococcus*) are thus unclear. The use of whole genome sequences and multiple gene trees would surely add confidence and resolution to the inference of their evolutionary relationships but this approach is, at present, limited by the gap in whole genome sequences from type strains. Unfortunately, recent phylogenomic studies on the *Roseobacter* group or the *Rhodobacteraceae* family have not included any reference genome of the genus *Thalassobius* (Newton et al., 2010; Tang et al., 2010; Luo and Moran, 2014; Simon et al., 2017).

Whole genome drafts of *T. mediterraneus* and *T. gelatinovorus* and *S. marina* type strains have been determined and characterized in the last two years (Rodrigo-Torres et al., 2016a,b, 2017). Also, Nogales and co-workers have explored the phylogenomics of a large collection of the *Roseobacter* group members, including several type strain genomes, by using more than a 100 single copy protein coding genes and found four *Thalassobius* species forming a well-defined lineage, which includes *Shimia* species (unpublished, Nogales et al., 2017, FEMS Meeting of European Microbiologists). Thus, the possible polyphyletic/paraphyletic nature of the genus is an open question.

During a survey aimed to resolve the taxonomic position of unidentified marine strains kept at the Spanish Type Culture Collection (CECT), whole genome sequences were obtained for four *Thalassobius* spp. strains. In this paper, we address the genome-based taxonomy of the genus *Thalassobius*, describe two new species of *Thalassobius* revealed through genome relatedness and present some predicted phenotypic traits of the group.

MATERIALS AND METHODS

Bacterial Strains

The strains used in this study are listed in **Table 1** with indication of their origins and references of previous studies. Marine Agar (MA) and Marine Broth (MB) were used as routine cultivation media and incubations were done at 26°C.

16S rRNA Gene Sequence Analysis

The complete 16S rRNA gene sequence of strains CECT 5113^T, CECT 5114, CECT 5118^T, and CECT 5120 were 1,465, 1,465, 1,469, and 1,469 nucleotides in length, respectively. These sequences extracted from the genome were compared with corresponding sequences of the type strains within the *Roseobacter* group using alignments retrieved from SILVA and LTP (Yarza et al., 2010) latest updates as references. When necessary, additional sequences were retrieved from the GenBank/EMBL/DDBJ databases. Alignments were corrected manually based on secondary structure information. Sequence similarities were calculated in ARB based on sequence similarities without the use of an evolutionary substitution model. Phylogenetic analysis using alternative treeing methods (maximum-parsimony, maximum-likelihood, and distance matrix) and data subsets were performed using the appropriate ARB tools (Ludwig et al., 2004).

Whole Genome Sequencing and Comparison

Genomic DNA was isolated using Real Pure Spin kit (Durviz) following the standard protocol recommended by the manufacturer. The integrity of the extracted DNA was checked by visualization in a 2.0% (w/v) agarose gel electrophoresis. Its purity and quantity was checked by measuring the absorbance at 260 and 280 nm with a spectrophotometer Nanodrop2000c (Thermo Scientific) and calculating the ratio A260/A280. Genome sequencing was achieved at Central Service of Support to Experimental Research (SCSIE) of the University of Valencia (Valencia, Spain) using an Illumina Miseq technology with 2 × 250 paired-end reads. The Illumina reads were analyzed for quality control using FASTQC, a common quality control tool developed by Babraham Bioinformatics to check raw sequencing data, which is wrapped in Galaxy Orione Server (Cuccuru et al., 2014). After filtering, the remaining reads were assembled using

TABLE 1 | Strains used in the study and their origins.

Strain	Equivalent designation	Isolation	Reference
T. activus CECT 5113^T	11SM13^T; LMG 29900^T	Sea water, Vinaroz, Spain. November 1989	Phenon 21, (Ortigosa et al., 1994)
T. activus CECT 5114	11SM18; LMG 29901	Sea water, Vinaroz, Spain November 1989	Phenon 21, (Ortigosa et al., 1994)
T. aestuarii CECT 8650^T	JC2049^T; KCTC 12049^T	Tidal flat sediment Ganghwa Island, Korea	Yi and Chun, 2006
T. autumnalis CECT 5118^T	XSM11^T; LMG 29904^T	Sea water, Vinaroz, Spain October 1989	Phenon 34, (Ortigosa et al., 1994)
T. autumnalis CECT 5120	11SM11; LMG 29905	Sea water, Vinaroz, Spain November 1989	Phenon 35, (Ortigosa et al., 1994)
T. gelatinovorus CECT 4357^T	Ahrens B6^T; LGM 129^T	Sediment, seawater, Kiel Fjord, Baltic Sea, Germany	Rüger and Höfle, 1992; Uchino et al., 1998
T. maritimus CECT 8648^T	GSW-M6^T; KCTC 23347^T	Seawater, Geoje Island, Korea	Park et al., 2012
T. mediterraneus CECT 5383^T	XSM19^T; CIP 108400^T	Seawater, Vinaroz, Spain October 1989	Arahal et al., 2005

several software choices for comparative purposes: (i) Spades 3.0.0 (Bankevich et al., 2012) incorporated as a tool in Galaxy Orione Server, (ii) Seqman Ngen 12.0.1 (DNAstar), and (iii) Velvet 1.0.0 *de novo* assembler (Zerbino and Birney, 2008). After performance evaluation and comparison of metrics, the best assembly for each organism was further processed. The bioinformatic tool CheckM (Parks et al., 2015) was used to assess the genome quality prior to annotation using Prokka v1.4.0 (Seemann, 2014), an open source software tool within Galaxy Orione Server, and RAST v2.0 (Rapid Annotation using Subsystem Technology; Aziz et al., 2008).

The similarity between genomes was assessed using several indices useful for species delineation. Thus, the DNA-DNA hybridization (DDH) was estimated *in silico* with the Genome-to-Genome Distance Calculator (GGDC 2.0) using the BLAST method and recommended formula 2 (Meier-Kolthoff et al., 2013); the average nucleotide identities according to MUMmer (ANIm) and BLAST (ANIb) were determined in JSpeciesWS (Richter et al., 2015); and OrthoANI values were calculated with the standalone Orthologous Average Nucleotide Identity Tool (OAT) (Lee et al., 2016).

The phylogenetic relationship of the genomes was explored with BCG54 using default settings. This software tool is available for download at EzBioCloud (Yoon et al., 2017) and employs a set of bacterial core genes, namely 54, that are single-copy and commonly present in all bacterial genomes.

Phenotypic Characterization

All determinations were performed in duplicate in non-simultaneous assays. Cell morphology was determined on wet mounts prepared from 24 to 48 h MA cultures of the strains, by using phase contrast microscopy in a Leica DMRB fluorescence microscopy. Colony morphology and pigmentation were recorded from 48 h MA cultures. PHB accumulation was determined according to Spiekermann et al. (1999). Ranges of temperature (4, 15, 26, 37, and 40°C) and salinities (3.5–10%) were determined in MA incubated up to 7 days. Marine Agar was supplemented with NaCl to attain total salinities of 4, 5, 6, 7, 8, 9, and 10%. Optimal values were taken from the fastest grown plates. Specific ionic requirements were tested by assessing the growth abilities of the strain on solid media with defined combinations of four sea salts (NaCl, $MgCl_2$, $CaCl_2$, and KCl), according with already reported methods (Macián et al., 2005). Extracellular hydrolytic activities on casein, starch, Tween-80, and DNA were determined after 6 d incubation by using Marine Agar supplemented with 10% (v/v) casein suspension, or 0.2% (w/v) soluble starch, for the first substrates. Tween-80 Agar (Smibert and Krieg, 2007) and DNAse agar (Oxoid) were supplemented with sea salts (Marine Cation Supplement, Farmer and Hickman-Brenner, 2006). Activity on starch was revealed after lugol addition and HCl 1 N was used to show DNAse activity. Oxidase test was performed with Oxoid oxidase discs and catalase was tested with 10 vol. H_2O_2. API 20NE and API ZYM strips were inoculated with cell suspensions prepared in 3.5% Seasalts (Oxoid) and AUX Medium for API 20NE assimilation tubes was supplemented with a concentrated Seasalts solution to give the same salinity used in the cell suspension. Sole carbon and energy sources used for growth were tested on Basal Medium Agar as described by Baumann and Baumann (1981).

Fatty acid methyl esters were extracted from biomass grown for 48 h on MA at 26°C and prepared according to standard protocols as described for the MIDI Microbial Identification System (Sasser, 1990) at the CECT. Cellular fatty acid content was analyzed by gas chromatography with an Agilent 6850 chromatographic unit, with the MIDI Microbial Identification System using the TSBA6 method (MIDI, 2008) and identified using the Microbial Identification Sherlock software package.

RESULTS AND DISCUSSION

16S rRNA Gene Phylogeny

Strains CECT 5113^T, CECT 5114, CECT 5118^T, and CECT 5120 were affiliated to the genus *Thalassobius* based on partial 16S rRNA gene sequence comparison performed among routine identification and authentication procedures at CECT. The strains had been isolated from the same samples and geographic location that rendered the type strain of *T. mediterraneus* (CECT 5383^T = XSM19^T). In fact, they were provisionally identified as *T. mediterraneus* or *Thalassobius* sp. strains in the CECT catalog. The similarities of their respective 16S rRNA gene sequences to *T. mediterraneus* CECT 5383^T were low (only 96.6%) for strains CECT 5113^T and CECT 5114, which showed a closer position to *T. maritimus* (98.3%). On the other hand, strains CECT 5118^T and CECT 5120 showed a 99.6% similarity to *T. mediterraneus* CECT 5383^T.

When a phylogenetic tree based on almost complete (genome derived) 16S rRNA gene sequences was built with the whole set of *Roseobacter* group species plus these four strains (**Figure 1**), a clear affiliation to the genus *Thalassobius* could be seen, as the sequences are included among *Thalassobius* spp. As expected, strains CECT 5113^T and CECT 5114 join *T. maritimus* GSW-M6^T with a high bootstrap support while strains CECT 5118^T and CECT 5120 merge with *T. mediterraneus* CECT 5383^T, also

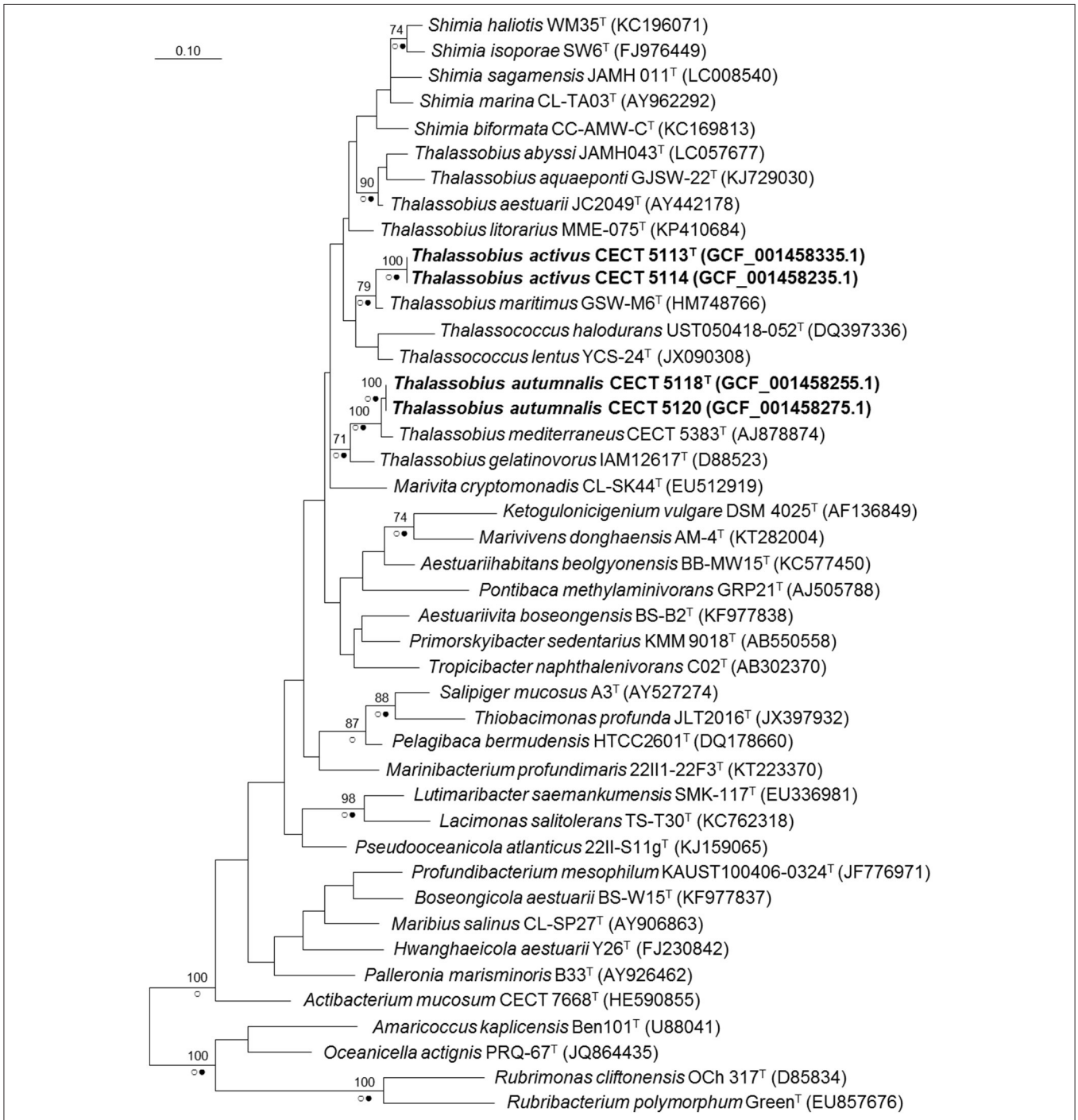


FIGURE 1 | Phylogenetic reconstruction based on the 16S rRNA gene using the maximum likelihood method. Bootstrap values (if >70% and outside subclades) based on 1,000 resamplings are shown as percentages at the branch nodes. Circles indicate that corresponding nodes were recovered in trees generated with the Maximum Parsimony (open circles) or the Neighbor joining (filled circles) methods. Bar. substitutions per nucleotide position. Name of the strains corresponding to new species are indicated in bold. In parentheses, RefSeq assembly accession number.

with high bootstrap, according with their similarity levels toward these species. The *Thalassobius* species, however, do not only include the sequences of these four strains, but also the types of the five *Shimia* species (all forming a monophyletic cluster within the *Thalassobius* group), plus the two species of *Thalassococcus*, which remain as a pair related to *T. maritimus*.

A close revision of 16S rRNA trees presented along the proposals of new species on these three genera reveal large variations in the relative positions of the taxa included in the analysis, depending on the particular set of selected taxa and the treeing method. The instability of the branching is always noticeable: for example, Nogi et al. (2015) used four *Thalassobius* spp. and no *Thalassococcus* to build their Neighbor Joining tree with the five species of *Shimia*: *Thalassobius* spp. appear as closest neighbors, forming a monophyletic cluster with the *Shimia* spp. By contrast, Hameed et al. (2013) presented a Neighbor Joining tree in which neighbors of *Shimia* are not *Thalassobius* spp., but *Nautella italica* and *Lentibacter algarum*. More examples of instability could be seen when comparing other papers describing *Roseobacter* group members. It is clear that 16S rRNA gene sequences alone are unable to resolve the phylogeny of a group displaying such complexity and overburden of genera descriptions. It seems particularly important to include not only taxa showing the highest 16S rRNA sequence similarity to the strains being considered, but a large representative set of members of the group. A more robust phylogeny is likely to be achieved by the use of whole genome information and selection of a large, optimized set of genes allowing more sound phylogenetic resolution of the clade, an approach currently in development. Genomic sequences are also valuable for the circumscription of the isolates at the species level and for providing insights into their biology, as will be reported in the coming subheadings.

Genome Sequence Metrics and Relatedness Indices

Genome length, G+C molar content, protein and rRNA genes are summarized in **Table 2** for all the genomes used. Strains CECT 5113^T and CECT 5114 possess genomes in the lower size and G+C range: <3.5 Mb genome size and <55 mol% G+C molar content. They are similar to *T. maritimus*, the phylogenetically closest species, which is also under 3.5 Mb genome length and was until now the one with lower G+C content (56.3 mol%). Other type strains except for *T. mediterraneus* have genomes larger than 3.9 Mb and G+C higher than 58 mol%. Strains CECT 5118^T and CECT 5120 had larger genomes, 4.3–4.4 Mb, and G+C molar content is among the higher in the genus, 59.8 mol%, surpassed only by the value of the *T. aestuarii* type strain (60.4 mol%).

ANIb, ANIm, OrthoANI, and digital DDH were determined in order to relate the strains CECT 5113^T, CECT 5114, CECT 5118^T, and CECT 5120 to the eight strains of *Thalassobius*, *Shimia,* and *Thalassococcus* sp. with publicly available genomes. Results of these determinations are reported in **Table 3** (ANIb, ANIm, digital DDH) and **Figure 2** (OrthoANI).

The four strains formed two very tight pairs related by almost 100% with all ANI calculations: strains CECT 5113^T and CECT 5114 were 99.8–100% related to each other and showed a maximum relatedness to the genome of the type strain of *T. maritimus* (as expected from 16S rRNA gene sequence similarity), with values of 75.0–75.1% (ANIb), 75.6–75.7% (OrthoANI), and 19.5–20.9% (digital DDH). Values with other species were even lower. According with the currently established boundaries for genomic species definition (95–96 for ANI, 70% for digital DDH) these values qualify both strains as members of a single genomic species, different from any of their closest relatives.

Strains CECT 5118^T and CECT 5120 also show high levels of reciprocal relatedness (99.9% ANI and 98.5% digital DDH). *T. mediterraneus*, the species most related in terms of 16S rRNA gene sequence, is also the closest, with 79.1% ANIb, 79.9% OrthoANI, and 23.2% digital DDH relatedness to them. All other *Thalassobius, Shimia,* and *Thalassococcus* species compared show even lower values. Thus, strains CECT 5118^T and CECT 5120 should also be considered as yet another new species from a genomic point of view.

In order to substantiate the recognition of the two novel species, a wide phenotypic characterization of the strains was

TABLE 2 | Genomic sequences employed with general features and accession numbers in public databases (an asterisk indicates those reported in this study).

Strain	Mb	Contigs	G+C (mol%)	Protein	rRNA	tRNA	Accession
S. haliotis DSM 28453^T	4.0	22	58.0	4020	5	44	FOSZ01
S. marina CECT 7688^T	4.0	45	57.4	3901	3	45	CYPW01
Shimia sp. SK013	4.0	28	57.2	3981	3	44	LAJH01
T. activus CECT 5113^T	3.4	26	54.4	3316	3	41	CYTO01*
T. activus CECT 5114	3.5	26	54.5	3443	5	44	CYUE01*
T. aestuarii DSM 15283^T	4.2	27	60.4	4135	9	51	FOTQ01
T. autumnalis CECT 5118^T	4.4	48	59.8	4163	3	53	CYSB01*
T. autumnalis CECT 5120	4.4	47	59.8	4287	7	53	CYSC01*
T. gelatinovorus CECT 4357^T	3.9	30	58.4	3800	7	47	CYSA01
T. maritimus DSM 28223^T	3.3	14	56.3	3305	12	51	FQWM01
T. mediterraneus CECT 5383^T	3.4	19	58.7	3299	8	47	CYSF01
T. halodurans DSM 26915^T	4.0	12	58.0	3902	9	46	FNUZ01

TABLE 3 | ANIb, ANIm, and estimated DDH values between genomes of *Thalassobius*, *Shimia,* and *Thalassococcus* species (arranged as in **Figure 3**).

	1	2	3	4	5	6	7	8	9	10	11	12
ANIb												
1 *T. maritimus* DSM 28223^T	–											
2 *T. activus* CECT 5113^T	75.0	–										
3 *T. activus* CECT 5114	74.8	**99.8**	–									
4 *T. aestuarii* DSM 15283^T	71.4	70.7	70.6	–								
5 *S. marina* CECT 7688^T	71.4	70.9	70.8	73.2	–							
6 *S. haliotis* DSM 28453^T	71.1	70.7	70.6	73.7	74.3	–						
7 *Shimia* sp. SK013	70.8	70.3	70.2	73.2	73.4	77.8	–					
8 *T. gelatinovorus* CECT 4357^T	70.3	69.9	69.8	71.7	70.5	70.4	70.3	–				
9 *T. autumnalis* CECT 5118^T	71.1	70.6	70.6	71.7	72.2	71.2	70.8	73.8	–			
10 *T. autumnalis* CECT 5120	71.1	70.6	70.6	71.7	72.2	71.2	70.9	73.8	**99.9**	–		
11 *T. mediterraneus* CECT 5383^T	71.2	71.1	71.0	71.8	71.9	71.2	70.9	73.5	79.1	79.1	–	
12 *T. halodurans* DSM 26915^T	70.7	69.9	69.8	71.4	70.3	70.9	70.4	70.9	71.2	71.2	71.4	–
ANIm												
1 *T. maritimus* DSM 28223^T	–											
2 *T. activus* CECT 5113^T	84.1	–										
3 *T. activus* CECT 5114	84.1	**100**	–									
4 *T. aestuarii* DSM 15283^T	84.4	84.1	83.9	–								
5 *S. marina* CECT 7688^T	84.4	84.9	84.8	83.9	–							
6 *S. haliotis* DSM 28453^T	83.9	84.5	84.3	83.5	85.2	–						
7 *Shimia* sp. SK013	83.6	83.9	83.8	83.4	83.9	83.8	–					
8 *T. gelatinovorus* CECT 4357^T	82.2	84.9	85.1	83.0	83.6	82.8	82.6	–				
9 *T. autumnalis* CECT 5118^T	84.1	84.5	84.5	83.9	85.5	83.9	83.4	83.3	–			
10 *T. autumnalis* CECT 5120	84.1	84.6	84.5	84.1	85.5	84.0	83.5	83.3	**99.9**	–		
11 *T. mediterraneus* CECT 5383^T	84.1	85.2	85.2	84.1	84.9	83.9	83.5	83.4	85.7	85.7	–	
12 *T. halodurans* DSM 26915^T	84.4	85.2	84.6	84.2	84.5	84.1	83.8	83.1	84.0	84.2	84.1	–
digital DDH												
1 *T. maritimus* DSM 28223^T	–											
2 *T. activus* CECT 5113^T	19.6	–										
3 *T. activus* CECT 5114	19.5	**99.9**	–									
4 *T. aestuarii* DSM 15283^T	19.8	19.5	19.3	–								
5 *S. marina* CECT 7688^T	21	20.7	20.6	19.6	–							
6 *S. haliotis* DSM 28453^T	20.3	19.6	19.6	19.3	20.8	–						
7 *Shimia* sp. SK013	19.5	19.9	19.7	18.7	19.7	21.2	–					
8 *T. gelatinovorus* CECT 4357^T	19.4	20.6	20.6	18.6	20.2	19.3	18.1	–				
9 *T. autumnalis* CECT 5118^T	20.6	20.9	20.8	20.5	22.3	20.4	19.4	19.9	–			
10 *T. autumnalis* CECT 5120	20.7	21	20.9	20.6	22.2	20.6	19.5	20	**98.5**	–		
11 *T. mediterraneus* CECT 5383^T	21.5	22.1	22	21.1	22.1	20.5	20.8	20	23.1	23.2	–	
12 *T. halodurans* DSM 26915^T	21.5	19.9	19.6	20.2	19.2	19	19.1	18.9	20.1	20.3	20.2	–

Values above the thresholds for species delineation are in bold.

undertaken, including the type strains of *T. mediterraneus, T. gelatinovorus*, *T. aestuarii,* and *T. maritimus*.

Phenotypic Characterization and Discrimination

Strains CECT 5113^T, CECT 5114, CECT 5118^T, and CECT 5120 were Gram-reaction (lyse in 3% KOH) and Gram-staining negative coccobacilli or short rods. Motility was negative in all cases and previous data (Ortigosa et al., 1994) on flagella stain indicate that the strains do not synthetize flagella. All four strains grow well on Marine Agar forming circular, slightly convex colonies with entire margin, which develop in 2 days incubation at 26°C. Strain CECT 5120 was slower and grew to a density lower than the three other strains. Strain CECT 5118^T produced a beige to brown diffusible pigment in prolonged incubations. All strains were able to accumulate polyhydroxyalkanoates (PHA) as revealed by the Nile Red plate assay on Marine Agar plus D-glucose (Spiekermann et al., 1999).

All strains grew optimally at 26°C and with salinities ranging from that in Marine Agar (3.4% aprox.) to 5% (3.5% of mixed salts in MA plus additional 1.6% NaCl). When cultured in media of the same nutritional content as MA (1% Tryptone and 0.3%

yeast extract) but deprived of salts or supplemented with 2% NaCl or 2% KCl none of the strains was able to grow, indicating a strictly halophilic nature. Strains CECT 5118^T and CECT 5120 were also unable to grow when the medium was supplemented with mixtures of 2% NaCl plus 0.9% $MgCl_2$, 0.2% $CaCl_2$, or 0.1% KCl or combinations of the four salts. Strains CECT 5113^T and CECT 5114 did not grow with 2% NaCl or 2% KCl or with 2% NaCl plus $MgCl_2$ or $CaCl_2$ separately, but were able to grow when both divalent cations, calcium, and magnesium chlorides, were added to NaCl.

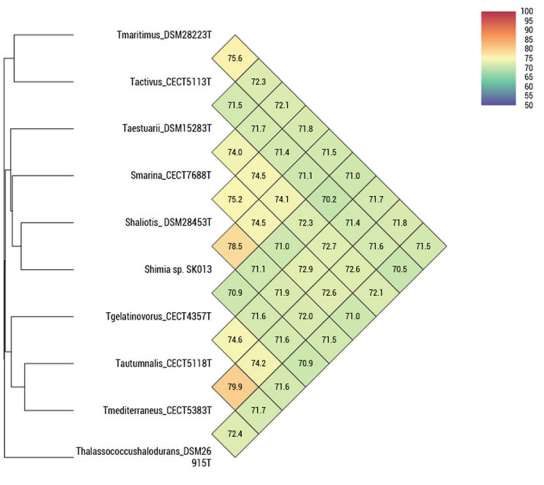


FIGURE 2 | OrthoANI indices between pairs of *Thalassobius*, *Shimia* and *Thalassococcus* strains (according to Lee et al., 2016).

The strains were poorly reactive in API ZYM and API 20NE supplemented with Marine Cations Supplement, which contains the four principal marine cations as sulfate and chloride salts. In order to better fulfill the ionic requirements of the strains, API strips were replicated using Ocean Salts (Oxoid) solution, so the final salinity was 3.5%, but reactivity on the miniaturized tests systems did not improve.

Anaerobic growth with nitrate as alternative electron-acceptor and a mixture of succinate, acetate and lactate as substrates was assayed in Baumann's denitrification medium (Baumann and Baumann, 1981) with negative results for all four strains. *T. gelatinovorus* CECT 4357^T was the only positive strain. Results for nitrate to nitrite reduction in API 20NE strips were in accordance with the results obtained in Baumann's medium. None of the strains was able to ferment D-glucose, a result that, in combination to the inability to live anaerobically with nitrate, usually qualifies a strain as strict aerobe.

A large number of sole carbon and energy sources for growth were tested on Baumann's Basal medium Agar supplemented with 0.2% carbohydrates, or 0.1% organic acids, amino acids, and amines. The list of used substrates is included in the species descriptions. As a general rule, organic acids were the most widely used carbon sources, although most amino acids and some carbohydrates were also used by the strains. This behavior was already described as characteristic for the genus. The spectrum of carbon sources is useful to differentiate species within the group, as indicated in **Table 4**.

Hydrolytic extracellular activities were very scarce among *Thalassobius* strains, with two exceptions: *T. gelatinovorus* CECT 4357^T and *T. aestuarii* CECT 8650^T were able to degrade gelatin, but other strains did not show hydrolysis on any of the substrates tested: gelatin, casein, Tween-80, starch, or DNA.

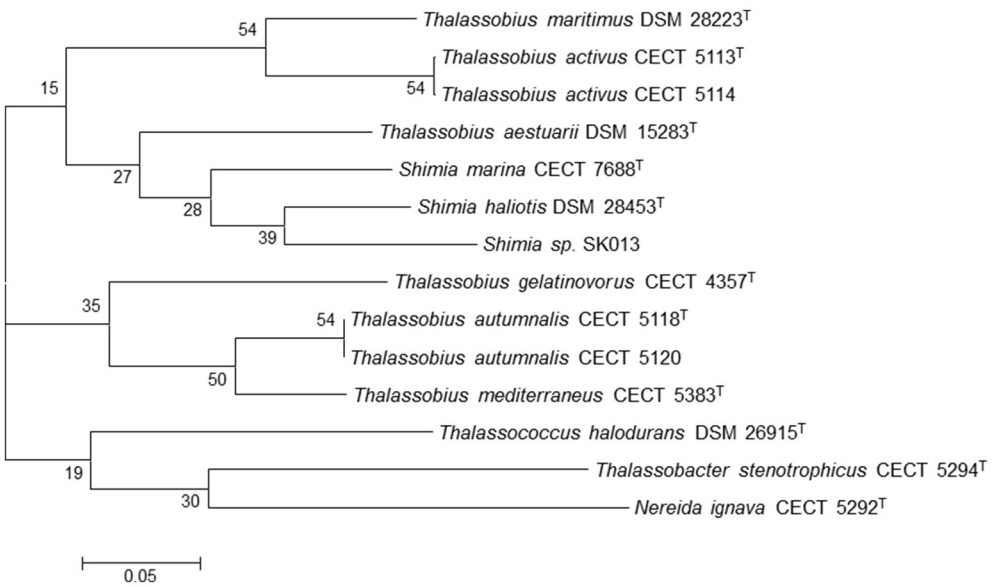


FIGURE 3 | Phylogenetic tree generated with BCG54. *Thalassobacter stenotrophicus* CECT 5294^T (CYRX01) and *Nereida ignava* CECT 5292^T (CVQV01) were included as outgroup organisms. The numbers at the nodes indicate the gene support index (maximal value is 54).

TABLE 4 | Differential characteristics between the type strains of *Thalassobius* species: 1, *T. autumnalis* sp. nov. CECT 5118^T; 2, *T. mediterraneus* CECT 5383^T; 3, *T. gelatinovorus* CECT 4357^T; 4, *T. activus* sp. nov. CECT 5113^T; 5, *T. maritimus* GSW-M6^T (Park et al., 2012); 6, *T. aestuarii* JC2049^T (Yi and Chun, 2006); 7, *T. aquaeponti* GJSW-22^T (Park et al., 2014); 8, *T. abyssi* JAMH043^T (Nogi et al., 2016); 9, *T. litorarius* MME-075^T (Park et al., 2016).

	1	2	3	4	5	6	7	8	9
Motility	–	–	+	–	+	–	–	–	–
NO_3^- reduction	–	–	+	–	+	–	+	+	–
Gelatin hydrolysis	–	–	+	–	–	+	–	–	–
Maximum temperature (°C)	37	37	40	37	30	35	35	30	35
Maximum salinity (%)	6	8	8	6	7	7	6	7	5
Use of carbon sources:									
D-cellobiose	+	–	–	+	+	+	+	+	–
Propionate	–	–	–	+	nd	nd	nd	nd	nd
L-threonine	–	–	–	+	nd	nd	nd	nd	nd
L-arginine	–	+	+	+	nd	+	nd	nd	nd
L-aspartate	–	+	+	+	nd	nd	nd	nd	nd
Fatty acid (%):									
$C_{10:0}$ 3OH	3	2	–	7–10	5	2	2	–	tr
$C_{12:1}$ 3OH	5	4–5	5	–	1	–	–	–	–
$C_{18:1}$ *ω7c* 11 methyl	3–9	1–2	–	3–5	–	1	16	tr	13

Unless otherwise indicated data from this study. +, positive; –, negative; nd, not determined.

TABLE 5 | Fatty acid composition of 1, *T. activus* CECT 5113^T; 2, *T. activus* CECT 5114; 3, *T. autumnalis* CECT 5118^T; 4, *T. autumnalis* CECT 5120; 5, *T. gelatinovorus* CECT 4357^T; 6, *T. mediterraneus* CECT 5383^T; 7, *T. maritimus* CECT 8650^T; 8, *T. aestuarii* CECT 8650^T.

	1	2	3	4	5	6	7	8
HYDROXYLATED								
$C_{10:0}$ 3OH	7.6	9.7	2.5	2.7		2.2	5.5	2.3
$C_{12:1}$ 3OH			5.1	4.9	5.2	4.6	1.2	
$C_{16:0}$ 2OH			tr		1.2			1.5
$C_{17:0}$ iso 3OH				2.6			2.7	2.4
SATURATED								
$C_{12:0}$			tr		tr			
$C_{16:0}$	9.3	9.2	4.1	3.8	4.7	4.7	6.3	3.0
$C_{17:0}$					2.8	tr	tr	tr
$C_{18:0}$	1.2	2.1	1.8	1.8	1.8	2.3	2.6	1.0
UNSATURATED								
Summed Feature 3	tr	tr	tr	tr		1.0		
Summed Feature 8	78.3	73.6	81.0	72.4	80.6	82.2	80.9	88.2
$C_{18:1}$ *ω7c* 11 methyl	3.0	4.8	3.5	9.5		1.4		1.0
$C_{18:1}$ *ω9c*			tr	tr	1.4	tr		
$C_{20:1}$ *ω7c*			tr	1.3				

Summed Feature 3: $C_{16:1}$ ω7c/ω6c; Summed Feature 8: $C_{18:1}$ ω7c/ω6c. All values determined in this study.

Cellular fatty acid composition was determined for the eight strains alongside with the result reflected in **Table 5**. The main fatty acid for all strains is included in the Summed Feature 8, corresponding to $C_{18:1}$ *ω7c*/$C_{18:1}$ *ω6c*, which accounts for 72–88% of the total. It is also the dominant fatty acid in the whole *Rhodobacteraceae* family. The second most abundant, common fatty acid, is $C_{16:0}$ (3–9%) while $C_{18:0}$ is also common to all species but accounts for 1–2.6%. The rest of fatty acids have a differential distribution among species, with $C_{18:1}$ *ω7c* 11methyl, $C_{10:0}$ 3OH, and $C_{12:1}$ 2OH amounting up to 9% in some species.

Differentiation among species of the genus based on experimental phenotypic tests could be achieved as indicated in **Table 4**.

Genome Derived Features of Strains CECT 5113^T, CECT 5114, CECT 5118^T, and CECT 5120

After annotation with RAST and Prokka, using the SEED Viewer (Overbeek et al., 2014) and BIOiPLUG to explore annotated genomes and Metacyc as a resource for metabolic pathway information (Caspi et al., 2016), the following features could be predicted from genome sequences:

1. Chemotaxonomic traits: A set of genes (*cdsA, pgsA, pgpA, pgpB,* and *cls*) coding for polar lipid biosynthetic enzymes for phosphatidylglycerol (EC 2.7.7.41, EC 2.7.8.5, and EC 3.1.3.27) and for phosphatidyl choline (EC 2.7.8.24) production was present but those for phosphatidyl ethanolamine, sulfoquinovosyl diacyl glycerol, diphosphatidyl glycerol (cardiolipin), phosphatidyl serine, or phosphatidyl *myo* inositol biosynthetic enzymes were not found. The presence of the gene coding for enzyme decaprenyl diphosphate synthase (EC 2.5.1.91) suggests ubiquinone 10 as major isoprenoid quinone. Polyamine biosynthetic abilities comprise, at least, production of putrescine, via arginine decarboxylase (*speA*) plus agmatinase (*speB*). Strains CECT 5113^T and CECT 5114 encode carboxy *nor*-spermidine dehydrogenase and carboxy *nor*-spermidine decarboxylase, but genes for these enzymes are absent from strains CECT 5118^T and CECT 5120. Peptidoglycan composition: pentapeptide was predicted to contain *meso*-DAP in third position, as predicted by the presence of *murE*, coding for UDP-N-acetyl muramoyl alanyl D-glutamate- 2,6 diaminopimelate ligase (EC 6.3.2.13).
2. Polyhydroxyalkanoate accumulation and degradation: the four strains present a complete set of genes for this subsystem.
3. Motility: No gene related to flagellar synthesis and movement was annotated in the genomes of strains CECT 5118^T and CECT 5120, in full agreement with their non-motile behavior. Strains CECT 5113^T and CECT 5114, also non motile, contain up to 66 genes in the subsystem, but apparently lack some essential ones (as the flagellin gene), among others, according to RAST annotation. However, in the Prokka-annotated version of these genomes flagellin was detected by BIOiPLUG search.
4. Cell cycle: the absence of the *minCDE* in the cytoskeleton set of genes is notorious; this absence is, to the best of our knowledge, common to the *Roseobacter* group and it might be related to the uneven division behavior found in several *Rhodobacteraceae* taxa. In addition, strains CECT 5113^T and

CECT 5114 lack *ftsQ* and strains CECT 5118^T and CECT 5120 lack *mreD*.

5. DNA metabolism and Horizontal Gene Transfer-related functions. None of the four strains contains a CRISPR/Cas system but defense against foreign DNA in the form of a restriction-modification system type I is present in the four strains. An additional type III R-M system is present in strain CECT 5118. All four strains contain a Gene Transfer Agent (GTA) set of genes. This GTA is universal in the *Rhodobacteraceae* so far explored, but the members of *Thalassobius*, *Shimia,* and *Thalassococcus* genera all lack the major capsid protein gene. Other horizontal transfer mechanisms seem to be present in strains CECT 5118^T and CECT 5120: they have the machinery to produce a type IV secretion system (T4SS) which may mediate protein and nucleoprotein transfer. Strains CECT 5113^T and CECT 5114 lack T4SS. Plasmid related functions were not predicted by RAST search, but all the four strains contain the plasmid replication genes *repA, repB,* and *repC*, characteristic of repABC type plasmids of *Rhodobacteraceae* that may also contain the T4SS (Petersen and Wagner-Döbler, 2017).
6. Cell to cell communication and adherence: Although none of the strains presents identifiable quorum-sensing systems, they all show the two component *chv* regulatory system characteristic of *Alphaproteobacteria*, involved in the regulation of exopolysaccharide synthesis and symbiotic recognition in *Rhizobium*. It is also noticeable the presence in the four strains (and in the rest of *Thalassobius* type strains) of a T2SS related to tight adherence (*tad*) gene cluster, involved in tenacious biofilm formation, as well as several genes encoding for capsular exopolysaccharide synthesis.
7. Stress response: Osmotic stress-related genes code for a complete betaine synthesis from choline. Oxidative stress is counteracted by the presence of catalase, peroxidase and a Fe-containing superoxide dismutase (SOD).
8. Light-related genes: None of the strains contains genes for the synthesis of a photosynthetic apparatus, but *regA*, a part of a two-component system related to the expression of photosynthesis in anaerobiosis, is predicted in all four strains; *hvrA*, a transcriptional activator of *puf* and *puh* genes at dim light intensities, is found only in strains CECT 5118^T and CECT 5120. All four lack genes for rhodopsin synthesis.
9. Anaerobic respiration. Nitrate reduction and denitrification: the strains do not possess respiratory nitrate reductase, in agreement with the negative results obtained in the corresponding test. They lack also the nitrous oxide and nitric oxide reductases needed for denitrification. Dimethyl sulfoxide (DMSO) reductase and thiosulfate or sulfite reductases are also absent but an arsenate reductase gene is detected in all strains.
10. Sulfur metabolism: Sulfur oxidation is predicted from the genomes of strains CECT 5118^T and CECT 5120 (a behavior common to several taxa in the *Rhodobacteraceae*), but not from the genomes of strains CECT 5113^T and CECT 5114, due to the lack of several *sox* genes (*soxA, X, W,* and *H)*. Sulfite oxidase is absent in all four. Another well studied, widely represented activity on sulfur compounds, the degradation of dimethyl sulfoniopropionate (DMSP), is present in the genomes of strains CECT 5118^T and CECT 5120 with a complete demethylation pathway. Strains CECT 5113^T and CECT 5114 present *dmdA* (demethylase) but it is uncertain if the pathway is complete (apparently, they lack the 3-methylmercapto propionyl-CoA ligase). On the other hand, strain CECT 5114 possesses a gene for DMSP lyase, thus being able to degrade DMSP directly to dimethylsulfide (DMS) plus acrylate.
11. Carbon monoxide oxidation: together with sulfur oxidation, this trait is present in several members of the *Roseobacter* group and apparently functions as a complementary energy resource. Aerobic carbon monoxide dehydrogenase is present in the genomes of strains CECT 5118^T and CECT 5120, but not in strains CECT 5113^T or CECT 5114.
12. Membrane transport (solutes): strains CECT 5118^T and CECT 5120 possess the highest number of genes involved in ABC transporters (86) while strains CECT 5113^T and CECT 5114 contain only 35–43 genes in this category. TRAP transporters account for 31–32 genes on strains CECT 5118^T and CECT 5120 and 32–36 in strains CECT 5113^T and CECT 5114.
13. Central metabolism and respiration: the four strains have a complete tricarboxylic acid cycle and aerobic respiration. The ethyl malonyl-CoA pathway for C2 assimilation is also complete in all four strains. Glycolytic abilities, on the other hand, include incomplete glycolysis (no phosphofructokinase is predicted), nor Entner-Doudoroff or Pentose phosphate pathways for strains CECT 5118^T and CECT 5120. Strain CECT 5113^T has a complete pentose phosphate and incomplete glycolytic and KDPG routes while strain CECT 5114 has incomplete versions of Pentose Phosphate and KDPG pathways. In any case, the four strains manage to use and grow with several carbohydrates, including D-ribose, some hexoses, a few disaccharides (as lactose, all present β-galactosidase activity) and some uronic acids. The precise combination of carbohydrates being used by the strains in experimental tests is useful for the species differentiation (**Table 4**).
14. CO_2 fixation is not predicted for any of the four strains.
15. Metabolism of aromatic compounds: the strains do not possess genes encoding for complete pathways for aromatic ring cleavage nor they encode dioxygenases, whereas some members of the genus do.

Selected genome-predicted differences for these two pairs and other *Thalassobius*, *Shimia,* and one *Thalassococcus* spp. are summarized in **Table 6**.

Considering both the experimentally determined features and the presence of genes predicting given phenotypic traits in the genomes, there are enough differences between strains and their closely related species for unequivocally differentiate the genomic groups previously delineated (**Tables 4**, **6**). Genomic, phylogenetic and phenotypic information thus confirm that the four strains represent two novel species. They fulfill the basic characteristics defining the genus *Thalassobius* (Arahal et al., 2005) and have *Thalassobius* species as their closest phylogenetic

TABLE 6 | Selected differential characteristics between the genomes of *Thalassobius, Shimia,* and *Thalassococcus* species: 1, *T. autumnalis* sp. nov. CECT 5118^T and CECT 5120; 2, *T. mediterraneus* CECT 5383^T; 3, *T. gelatinovorus* CECT 4357^T; 4, *T. activus* sp. nov. CECT 5113^T and CECT 5114; 5, *T. maritimus* DSM 28223^T; 6, *T. aestuarii* DSM 15283^T; 7, *S. marina* CECT 7688^T; 8, *S. haliotis* DSM 28453^T; 9, *T. halodurans* DSM 26915^T.

	1	2	3	4	5	6	7	8	9
Entner-Doudoroff pathway, classical	−	−	+	−	+	+	−	+	+
α- and β-galactosidades	+	−	−	+	+	−	+‡	+	+
Carbon monoxide oxidation (CODH)	+	+	+	−	−	+	+	+	+
CO_2 fixation, Calvin cycle									
RuBisCo	−	−	+	−	+	−	−	−	−
Phosphoribulokinase	−	−	+	−	−	−	−	−	−
Protocatechuate dioxygenase and 3,4 dihydroxyphenylacetate 2,3 dioxygenase	−	−	+	−	−	+	−	−	+
Denitrification	−	−	+*	−	+	+§	−	+*	−
Aryl sulfatase	−	−	−	−	−	+	+	+	−
DMSP lyase (DMSP to DMS + acrylate)	−	−	−	v	−	+	−	+	+
Sox genes (sulfur oxidation)	+	−	+	−†	−	+	−	−	+
HvrA (transcriptional activator for *puf/puh*)	+	+	+	−	−	+	+	+	+
Type IV secretion system	+	+	−	−	−	−	+	−	+
Restriction-Modification systems	I, III	−	I	I	−	I	I	I	−
Gene Transfer Agent, GTA (no Capsid gene)	+	−	+	+	+	+	+	+	+
RepABC plasmid replication module	+	+	−	+	+	+	+	−	+
Multidrug efflux pumps	+	+	+	−	+	+	+	+	−
Phospholipids	PG, PC	PG, PC	PG, PC	PG, PC	PG, PC	PG, PC, DPG	PG, PC, DPG	PG, PC, DPG	PG, PC, PE
Polyamine biosynthetic enzymes:									
Carboxy *nor*-spermidine dehydrogenase and decarboxylase	−	+	+	+	+	+	+	− −	+
G+C (mol%)	59.8	58.7	58.4	54.4	56.3	60.4	57.4	58.0	58.0

*Only N_2O reduction pathway.
§Only NO and N_2O reduction pathway.
†soxA, X, W and H, missing.
‡Only β-galactosidase. +, positive, –, negative, v, variable between strains.

neighbors in both cases, thus, they should be recognized as new *Thalassobius* species: we propose the name of *Thalassobius activus*, for strains CECT 5113^T and CECT 5114, with CECT 5113^T as type strain. Strains CECT 5118^T and CECT 5120 constitute another new species, for which we propose the name of *Thalassobius autumnalis*, with CECT 5118^T as type strain. Protologues for both species are included at the end of the section.

Phylogenomic Analysis of Genera *Thalassobius*, *Shimia,* and *Thalassococcus*

As previously pointed, another goal of this study was to study the relationships between the species of the genera *Thalassobius*, *Shimia,* and *Thalassococcus* by using genome-derived data. All publicly available type strain genomes of species in these three genera have been considered together with the genome from an unclassified *Shimia* species (Kanukollu et al., 2016) that has been proven to constitute a separate genomic species (**Tables 2**, **3**). The tree generated with BCG54 using *Thalassobacter stenotrophicus* CECT 5294^T (CYRX01) and *Nereida ignava* CECT 5292^T (CVQV01) as outgroup is shown in **Figure 3**. The tree confirms the phylogenetic relationships previously suggested by 16S rRNA gene sequence comparison and genomic relatedness, confirming the close position of *T. activus* to *T. maritimus* and of *T. autumnalis* to *T. mediterraneus*. The grouping of *T. mediterraneus* plus *T. autumnalis* with *T. gelatinovorus*, a regular finding with other approaches, is also confirmed. All three *Shimia* species appear inside the *Thalassobius* genus, as closest relatives of *T. aestuarii*, while *T. halodurans* separates from *Thalassobius* and *Shimia*, and merges with the pair used as an outgroup, discarding the topology that *Thalassococcus* spp. occupied in the 16S rRNA tree. The results obtained from the BCG54 tree do not sustain monophyly for *Thalassobius* as currently defined. Furthermore, monophyly for *Thalassobius—Shimia* grouping is not resolved, as the group formed by *T. mediterraneus-T. autumnalis-T. gelatinovorus* does not merge with the rest of species before the last node. A better resolution of this node might be attained if the genomes of all *Shimia, Thalassobius,* and *Thalassococcus* were available, but, for the moment, we only can conclude that there are three possible monophyletic groups to be considered: the core *Thalassobius* spp. or *Thalassobius* spp. *sensu stricto* (containing the type species *T. mediterraneus* plus *T. autumnalis* and *T. gelatinovorus*), the one of *T. maritimus* plus *T. activus* sp. nov. and the one of *Shimia*

spp. plus *T. aestuarii*. A strict taxonomic procedure would split *Thalassobius* spp. in three genera, with *T. aestuarii* (and perhaps others, as *T. abyssi, T. aquaeponti,* and *T. litoreus*) being reclassified as new combinations of *Shimia*, and *T. maritimus* plus *T. activus* being proposed as a new genus. But as already commented, this would be premature, since the addition of more type strain genomes would eventually reveal if the whole group constitutes a monophyletic lineage at higher levels, thus allowing a reclassification of all *Shimia* spp as *Thalassobius* new combinations, a solution that, in our opinion, would be also possible and more clarifying than the splitting option, which would imply to describe more and more indistinguishable genera within the *Roseobacter* group.

Genome Derived Features of *Thalassobius* spp., *Shimia* spp., and *Thalassococcus halodurans*: Common and Differential Features

A survey of the gene content of the *Thalassobius* spp., *Shimia* spp., and *T. halodurans* draft genomes was conducted in order to enlighten the taxonomic status of the three genera. We looked for traits especially used for taxonomic work at the genus level. Genomic traits that are differential among species and/or genera are presented in **Table 6**. A list of features shared by all three genera representatives is included as **Table 7**.

All three genera representatives share characteristics owned by the whole *Rhodobacteraceae* family, as the peptidoglycan composition (with *m*-DAP at 3rd position), predominant ubiquinone (Q10), major cellular fatty acid ($C_{18:1}$ *ω7c*/$C_{18:1}$ *ω6c*) and a G+C molar content over 50%. Maybe, the lack of *minCDE* genes should be added to this list of traits common to all *Rhodobacteraceae*: the absence of these genes in various *Rhodobacteraceae* was already noted by Lutkenhaus (2007) and we have been unable, so far, to find any member of this group that contained this part of the cytoskeletal machinery. Obviously, these shared common traits (symplesiomorphies) should not be considered as relevant to back a genera fusion proposal. But there are other that are shared by these particular genera (synapomorphies): we would highlight, for example, the GTA gene set composition, characterized by the common absence of capsid gene, which seems to be exclusive of these genera in contrast to the other >40 genomes of *Roseobacter* group members that could be compared in the Seed Viewer. The ability to accumulate polyhydroxyalkanoates, although not exclusive, would be another common character, not shared with all other *Rhodobacteraceae*. Other candidate traits are indicated in **Table 7**.

On the other hand, several genomic differences were found that allow species discrimination, as well as differentiation of the three phylogenetic clusters revealed by gene and multigenic trees: polar lipid synthesis, for example, pull apart the *T. aestuarii-Shimia* spp. group by the exclusive presence of DPG synthetic ability, while *Thalassococcus* could be differentiated by the unique presence of gene *psd,* encoding a phosphatidyl ethanolamine-synthetizing enzyme, phosphatidyl serine decarboxylase (EC 4.1.1.65), which is absent from all other species. The absence of genes encoding for PE biosynthesis in all *Thalassobius* species was surprising, as Park et al. (2012) reported the detection of PE in the type strains of four species (*T. maritimus, T. aestuarii, T. gelatinovorus,* and *T. mediterraneus*) by using TLC chromatography. But genomes of these type strains did not contain phosphatidylserine decarboxylase gene and PE synthesis is not predicted by RAST for any of the eight *Thalassobius* strains whose genomes have been explored. Genes for key enzymes of the alternative routes for PE synthesis, ethanolamine phosphotransferase (EC 2.7.8.1) or L-serine phosphatidylethanolamine phosphatidyl transferase (EC 2.7.8.29) were also undetected. Undoubtedly, this is a point that needs more attention, because of the importance that is given to polar lipid composition in bacterial taxonomy.

Absence of carbon monoxide dehydrogenase and *hvrA* gene are characteristic of *T. maritimus-T. activus*, while, the presence of aryl sulfatase is exclusive of the *Shimia* spp.-*T. aestuarii* group.

It is interesting to highlight that *T. aestuarii* shares with *T. gelatinovorus* the ability to degrade aromatic compounds, as confirmed through Aromadeg (Duarte et al., 2014) showing a complete that both type strains contain genes coding for extradiol dioxygenases of Viccinal Chelate superfamily and Lig B superfamily (for degrading monocyclic substrates in the first case and homoprorocatechuate and protocatechuate, in the second) and Rieske non-heme iron oxygenases of the phthalate dioxygenases family. However, only *T. aestuarii* DSM 15283^T genome harbors an extradiol dioxygenase of the Cupin superfamily, which anables to degrade gentisate and, on the other side, *T. gelatinovorus* CECT 4357^T possesses an additional phthalate oxygenase and a benzoate oxygenase. *T. halodurans* DSM 26915^T genome also encodes two extradiol dioxygenases of Viccinal Chelate superfamily and one Rieske non-heme iron oxygenases of the phthalate dioxygenases. The three *Shimia* genomes were also checked with Aromadeg (default parameters) but no protein had matches with the database.

In a recent phylogenomic study, Simon et al. (2017) analyse the evolutionary adaptation of *Rhodobacteraceae* to marine and non-marine habitats. Among other things, they conclude that the so-called *Roseobacter* group is not monophyletic,

TABLE 7 | A list of common genomic features shared by *Thalassobius-Shimia* group included in the study.

- A complete set of PHB synthesis and degradation machinery
- Type II secretion system, tad type
- *chv* two components response regulator
- Absence of *BChl a* production, together with presence of *regA* gene
- Absence of proteorhodopsin production
- Presence of a Genome Transfer Agent set (GTA) lacking Capsid gene
- Lack of *minCDE* system among cytoskeleton set of genes
- Putrescine biosynthetic ability (arginine decarboxylase, *speA,* plus agmatinase*, speB*)
- Phosphatidyl choline synthesis (*pcs*)
- Decaprenyl phosphate synthase (Q10)
- A complete ethylmalonyl CoA pathway for C2 assimilation
- Arsenate reductase
- Mercuric reductase

but their members derived from a common marine ancestor shared with other non-marine, non-halophilic *Rhodobacteraceae* (*Rhodobacter* and *Paracoccus*, mainly), the later representing an adaptation to non-marine habitats. A selection of genome-predicted, habitat-correlated enzymes is shown in their **Table 1**, including several lost in non-marine habitats plus others that were gained in marine habitats. The analysis of Simon et al. (2017) included no *Thalassobius*, *Shimia,* or *Thalassococcus* (the genome labeled as "*Thalassiobium* sp. R2A62" is completely unrelated to any true *Thalassobius*), so we have explored the distribution of the marine-related enzymes they reported in the genomes employed in this paper, as all of them pertain to marine inhabitants. Results are indicated in **Table 8**. As it can readily be observed, the five enzymes classified under the category of "Lost in non-marine habitats" are all present in the genomes of *Thalassobius* spp. (6 species), *Shimia* spp (3 species), and in the single *Thalassococcus* species included. The only significant exception was the genes for large, medium, and small carbon monoxide dehydrogenase chains, absent from the genomes of *T. maritimus* and *T. activus* sp. nov. (both strains), a trait already highlighted as characteristic of this small clade (**Figure 3**). On the other hand, enzymes apparently "gained in marine habitats" are not widely represented in *Thalassobius* or *Shimia* genomes. A few interesting exceptions are the betaine homocysteine S-methyl transferase (EC 2.1.1.5), involved in glycine-betaine catabolism, present in all these genomes, and the nitrile hydratase (EC 4.2.1.84), absent only from the genomes *T. activus* sp. nov. and *T. halodurans*, but present in the rest. Two additional occurrences are worth to note: γ-butyrobetaine dioxygenase (EC 1.14.11.1), participating in carnitine biosynthesis, is found exclusively in the genomes of *Thalassobius* spp. *sensu stricto* (*T. mediterraneus, T. autumnalis* sp. nov., and *T. gelatinovorus*) while aryl sulfatase (EC 3.1.6.1) is present only in *T. aestuarii* and *Shimia* spp. (as already mentioned); thus, these two enzymes could be used as discriminant traits to differentiate the respective groups.

So, this analysis proves there is room for either a proposal of reunification of (at least) *Shimia* spp. with *Thalassobius* spp. under the name of *Thalassobius*, but also for the splitting of *Thalassobius-Shimia* in three genera (one of them a new one for *T. maritimus-T. activus*). However, any decision should be postponed until the gap in genomes of type strains pertaining to the group is closed.

TABLE 8 | Habitat-related enzymes and their presence in the genomes of different *Thalassobius, Shimia* and *Thalassococcus* type and reference strains: 1, *T. autumnalis* sp. nov. CECT 5118^T and CECT 5120; 2, *T. mediterraneus* CECT 5383^T; 3, *T. gelatinovorus* CECT 4357^T; 4, *T. activus* sp. nov. CECT 5113^T and CECT 5114; 5, *T. maritimus* DSM 28223^T; 6, *T. aestuarii* DSM 15283^T; 7, *S. marina* CECT 7688^T; 8, *S. haliotis* DSM 28453^T; 9, *T. halodurans* DSM 26915^T.

Enzyme	EC number	1	2	3	4	5	6	7	8	9
Gained in marine habitats[a]										
Ectoine synthase	4.2.1.108	−	−	−	−	−	−	−	−	−
Betaine homocysteine S-methyl transferase	2.1.1.5	+	+	+	+	+	+	+	+	+
γ-butyrobetaine dioxygenase	1.14.11.1	+	+	+	−	−	−	−	−	−
TMA[b] corrinoid protein Co-methyl transferase	2.1.1.250	−	−	−	−	−	−	−	−	−
TMA[b] N-oxyde reductase	1.6.6.9	−	−	−	−	−	−	−	−	−
Nitrile hydratase	4.2.1.84	+	+	+	−	+	+	+	+	−
Aryl sulfatase	3.1.6.1	−	−	−	−	−	+	+	+	−
Precorrin 3B synthase	1.14.13.83	−	−	−	−	−	−	−	−	−
Lost in non-marine habitats[a]										
(S)-2-haloacid dehalogenase	3.8.1.2	+	+	+	+	+	+	+	+	+
Mercury (Hg) II reductase	1.16.1.1	+	+	+	+	+	+	+	+	+
Carbon monoxide dehydrogenase	1.2.99.2	+	+	+	−	−	+	+	+	+
Precorrin-8X methylmutase	5.4.1.2	+	+	+	+	+	+	+	+	+
Precorrin-4 C11 methyl transferase	2.1.1.133	+	+	+	+	+	+	−	+	+

[a] *According to* Simon et al. (2017).

[b] *TMA, trimethylamine. +, present; –, not detected in annotation.*

CONCLUSIONS

This study provides the first taxogenomic approach conducted on the genus *Thalassobius* and permits the classification of four isolates into two novel species, whose description is given below. An effort has been done to explore the possibilities of phenotypic inference opening a trend for similar studies. It also provides an analysis of the complexity for classification at the genus level within the *Roseobacter* group, here illustrated by the case of the genera *Thalassobius-Shimia-Thalassococcus*.

Description of *Thalassobius activus* sp. nov.

Thalassobius activus (ac.ti'vus. L. masc. adj. *activus*, active, referring to the metabolic activity of the type strain).

Cells are Gram-reaction-negative, non- motile coccobacilli. Aerobic and chemoorganotroph, they grow on Marine Agar as regular unpigmented circular colonies. Cells accumulate PHB. Mesophilic, able to grow from 15 to 26°C (optimum, 26°C), but not a 4 or 37°C. Slightly halophilic, requires sodium, calcium and magnesium salts for growth. Maximum salinity for growth is 6% (optimum 3.5–5%), no growth is obtained without salt, at 7% or above. Oxidase and catalase are positive. Fermentation of carbohydrates, nitrate reduction and denitrification are negative. Negative for urea, casein, gelatin, starch, Tween-80 and DNA hydrolysis. Indole production from tryptophan and arginine dihydrolase are negative, PNPG and alkaline phosphatase are positive.

Carbon sources used for growth included D-cellobiose, sucrose, D-melibiose, D-mannitol, *m*-inositol, acetate, pyruvate, propionate, butyrate, *t*-aconitate, citrate, 2-oxoglutarate, succinate, lactate, 3-hydroxybutyrate, glycine, L-leucine, L-serine, L-threonine, L-arginine, L-glutamate, L-alanine, L-tyrosine, L-ornithine, 4-aminobutyrate, and L-aspartate. No growth is obtained with D-ribose, L-arabinose, D-xylose, D-glucose, D-fructose, D-galactose, D-trehalose, D-mannose, L-rhamnose, maltose, lactose, salicin, amygdalin, D-glycerol, D-sorbitol, D-gluconate, D-glucuronate, D-galacturonate, D-glycerate, D-saccharate, L-citrulline, L-histidine, L-lysine, L-sarcosine, or putrescine.

Major cellular fatty acids are $C_{18:1}$ *ω7c*/*ω6c* and $C_{10:0}$ 3OH.

The predominant respiratory quinone, Q10 was inferred from annotated gene encoding decaprenyldiphosphate synthase (EC 2.5.1.91). Major polar lipids, PG and PC, were inferred from annotated genes encoding CTP-phosphatidate cytidyl transferase (EC 2.7.7.41), phosphatidyl glycerophosphate synthase (EC 2.7.8.5), phosphatidyl glycerophosphatase (EC 3.1.3.27), and phosphatidyl choline synthase (EC 2.7.8.24), in the genomes of strains CECT 5113^T and CECT 5114.

The G+C content of the DNA is 54.4–54.5 mol% (partial genome sequences).

The type strain, CECT 5113^T (=11SM13^T =LMG 29900^T) was isolated from coastal seawater, Mediterranean Sea, at Vinaroz coast, Spain.

Description of *Thalassobius autumnalis* sp. nov.

Thalassobius autumnalis (au.tum.na'lis. L. adj. *autumnus*, fall, autumn, after the season of isolation).

Cells are Gram-reaction negative, non-motile coccobacilli, aerobic, and chemoorganotroph. Grows on Marine Agar forming regular circular colonies. Some strains produce a brown diffusible pigment on prolonged incubation. Polyhydroxybutyrate (PHB) is accumulated in the cells. Mesophilic and slightly halophilic, growth is observed from 15 to 37°C (optimum, 26°C), but not at 4 or 40°C and up to 6% total salinity in media supplemented with sea salts (optimum 3.5–5%). Does not grow at 7% or more salinity. It displays complex ionic requirements, as it is unable to grow either without added salts or with addition of simple salts (NaCl, KCl, $MgCl_2$, $CaCl_2$). Oxidase and catalase positive. Nitrate reduction to nitrite or N_2 is negative. Does not ferment carbohydrates. Negative for urea, casein, gelatin, starch, Tween-80, and DNA hydrolysis. Indole production from tryptophan and arginine dihydrolase are negative, PNPG (β-galactosidase) and leucine arylamidase are positive.

The following sole carbon and energy sources are used for growth: D-ribose, D-xylose, D-glucose, D-mannose, D-galactose, D-cellobiose, sucrose, salicin, N-acetyl-D-glucosamine, D-glycerol, *m*-inositol, acetate, pyruvate, propionate, butyrate, *t*-aconitate, citrate, 2-oxoglutarate, succinate, malate, lactate, 3-hydroxybutyrate, L-leucine, L-serine, L-glutamate, L-alanine, L-tyrosine, L-ornithine, 4-aminobutyrate, L-sarcosine, and putrescine. Growth is negative with L-arabinose, D-fructose, D-trehalose, L-rhamnose, maltose, lactose, D-melibiose, D-amygdalin, D-mannitol, D-sorbitol, D-gluconate, D-glucuronate, D-galacturonate, D-glycerate, D-saccharate, glycine, L-threonine, L-arginine, L-citrulline, L-histidine, L-lysine, and L-aspartate.

Major cellular fatty acids are $C_{18:1}$ *ω7c*/*ω6c* and $C_{18:1}$ *ω7c* 11-methyl.

The predominant respiratory quinone, Q10 was inferred from annotated gene encoding decaprenyldiphosphate synthase (EC 2.5.1.91). Major polar lipids, PG and PC, were inferred from annotated genes encoding CTP-phosphatidate cytidyl transferase (EC 2.7.7.41), phosphatidyl glycerophosphate synthase (EC 2.7.8.5), phosphatidyl glycerophosphatase (EC 3.1.3.27), and phosphatidyl choline synthase (EC 2.7.8.24), in the genomes of strains CECT 5118^T and CECT 5120.

The G+C content of the DNA is 59.8 mol% (partial genome sequences).

The type strain, CECT 5118^T (=XSM11^T =LMG 29904^T) was isolated from coastal seawater, Mediterranean Sea, at Vinaroz coast, Spain.

AUTHOR CONTRIBUTIONS

DA and MP: Designed the study; LR-T: Obtained the genomes and did the mainstream processing; TL, LR-T, and MP: Further analyzed the annotations; MP: Drafted the manuscript and conducted the phenotypic testing; TL and DA: Did the phylogenomic analysis; all authors corrected and approved the manuscript.

FUNDING

The study has been funded through projects TAXPROMAR 2010 CGL2010-18123/BOS (Spanish Ministry of Economy and Competitivity) to MP and PROMETEO 2012-040 (Generalitat Valenciana).

ACKNOWLEDGMENTS

We thank M. A. Ruvira and L. Jiménez-Gadea for technical support.

REFERENCES

Arahal, D. R., Macián, M. C., Garay, E., and Pujalte, M. J. (2005). *Thalassobius mediterraneus* gen. nov., sp. nov., and reclassification of *Ruegeria gelatinovorans* as *Thalassobius gelatinovorus* comb. nov. *Int. J. Syst. Evol. Microbiol.* 55, 2371–2376. doi: 10.1099/ijs.0.63842-0

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

Baumann, P., and Baumann, L. (1981). "The marine gram-negative eubacteria: genera Photobacterium, Beneckea, Alteromonas, Pseudomonas and Alcaligenes," in *The Prokaryotes*, *Vol. 2*, eds M. P. Starr, H. Stolp, H. G. Trueper, A. Balows, and H. Schleger (Berlin; Heidelberg: Springer), 1302–1331.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44, D471–D480. doi: 10.1093/nar/gkv1164

Chen, M. H., Sheu, S. Y., Chen, C. A., Wang, J. T., and Chen, W. M. (2011). *Shimia isoporae* sp. nov., isolated from the reef-building coral *Isopora palifera*. *Int. J. Syst. Evol. Microbiol.* 61, 823–827. doi: 10.1099/ijs.0.022848-0

Choi, D. H., and Cho, B. C. (2006). *Shimia marina* sp. nov., a novel bacterium of the *Roseobacter clade* isolated from biofilm in a coastal fish farm. *Int. J. Syst. Evol. Microbiol.* 56, 1869–1873. doi: 10.1099/ijs.0.64235-0

Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Soranzo, N., Travaglione, A., et al. (2014). Orione, a web-based framework for NGS analysis in microbiology. *Bioinformatics* 30, 1928–1929. doi: 10.1093/bioinformatics/btu135

Duarte, M., Jauregui, R., Vilchez-Vargas, R., Junca, H., and Pieper, D. H. (2014). AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database* 2014:bau118. doi: 10.1093/database/bau118

Farmer, J. J. III, and Hickman-Brenner, F. W. (2006). "The genera vibrio and photobacterium," in *The Prokaryotes*, *3rd Edn.*, eds M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (New York, NY: Springer), 508–563.

Hameed, A., Shahina, M., Lin, S. Y., Lai, W. A., Hsu, Y. H., Liu, Y. C., et al. (2013). *Shimia biformata* sp. nov., isolated from surface seawater, and emended description of the genus Shimia Choi and Cho 2006. *Int. J. Syst. Evol. Microbiol.* 63, 4533–4539. doi: 10.1099/ijs.0.053553-0

Hyun, D. W., Kim, M. S., Shin, N. R., Kim, J. Y., Kim, P. S., Whon, T. W., et al. (2013). *Shimia haliotis* sp. nov., a bacterium isolated from the gut of an abalone, *Haliotis discus* hannai. *Int. J. Syst. Evol. Microbiol.* 63, 4248–4253. doi: 10.1099/ijs.0.053140-0

Kanukollu, S., Voget, S., Pohlner, M., Vandieken, V., Petersen, J., Kyrpides, N., et al. (2016). Genome sequence of Shimia str. SK013, a representative of the Roseobacter group isolated from marine sediment. *Stand. Genom. Sci.* 11:25. doi: 10.1186/s40793-016-0143-0

Lee, I., Kim, Y. O., Park, S.-C., and Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103. doi: 10.1099/ijsem.0.000760

Lee, O. O., Tsoi, M. M. Y., Li, X., Wong, P. K., and Qian, P. Y. (2007). *Thalassococcus halodurans* gen. nov., sp., nov., a novel halotolerant member of the Roseobacter clade isolated from the marine sponge *Halichondria panicea* at Friday Harbor, USA. *Int. J. Syst. Evol. Microbiol.* 57, 1919–1924. doi: 10.1099/ijs.0.64801-0

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004). ARB: a software environment for sequence data. *Nucl. Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

Luo, H., and Moran, M. A. (2014). Evolutionary ecology of the marine Roseobacter clade. *Microbiol. Mol. Biol. Rev.* 78, 573–587. doi: 10.1128/MMBR.00020-14

Lutkenhaus, J. (2007). Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. *Ann. Rev. Biochem.* 76, 539–562. doi: 10.1146/annurev.biochem.75.103004.142652

Macián, M. C., Arahal, D. R., Garay, E., Ludwig, W., Schleifer, K. H., and Pujalte, M. J. (2005). *Thalassobacter stenotrophicus* gen. nov., sp. nov., a novel marine α-proteobacterium isolated from Mediterranean Sea water. *Int. J. Syst. Evol. Microbiol.* 55, 105–110. doi: 10.1099/ijs.0.63275-0

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

MIDI (2008). *Sherlock Microbial Identification System Operating Manual, Version 6.1.* Newark, DE: MIDI Inc.

Newton, R. J., Griffin, L. E., Bowlwes, K. M., Meile, C., Gifford, S., Givens, C. E., et al. (2010). Genome characteristics of a generalist marine bacterial lineage. *ISME J.* 4, 784–798. doi: 10.1038/ismej.2009.150

Nogi, Y., Mori, K., Makita, H., and Hatada, Y. (2016). *Thalassobius abyssi* sp. nov., a marine bacterium isolated from cold-seep sediment. *Int. J. Syst. Evol. Microbiol.* 66, 574–579. doi: 10.1099/ijsem.0.000758

Nogi, Y., Mori, K., Uchida, H., and Hatada, Y. (2015). *Shimia sagamensis* sp. nov., a marine bacterium isolated from cold-seep sediment. *Int. J. Syst. Evol. Microbiol.* 65, 2786–2790. doi: 10.1099/ijs.0.000333

Ortigosa, M., Garay, E., and Pujalte, M. J. (1994). Numerical taxonomy of aerobic, gram-negative bacteria associated with oysters and surrounding seawater of the Mediterranean coast. *Syst. Appl. Microbiol.* 17, 589–600. doi: 10.1016/S0723-2020(11)80081-0

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

Park, I., Cha, I. T., and Seo, M. J. (2016). *Thalassobius litorarius* sp. nov., isolated from a tidal flat. *Int. J. Syst. Evol. Microbiol.* 66, 1666–1672. doi: 10.1099/ijsem.0.000931

Park, S., Jung, Y. T., Kim, S. I., and Yoon, J. H. (2013). *Thalassococcus lentus* sp. nov., an alphaproteobacterium isolated from seawater of a seaweed farm. *Antonie van Leeuwenhoek* 103, 465–473. doi: 10.1007/s10482-012-9826-8

Park, S., Jung, Y. T., Won, S. M., Park, J. M., and Yoon, J. H. (2014). *Thalassobius aquaeponti* sp. nov., an alphaproteobacterium isolated from seawater. *Antonie van Leeuwenhoek* 106, 535–542. doi: 10.1007/s10482-014-0225-1

Park, S., Lee, M. H., Lee, J. S., Oh, T. K., and Yoon, J. H. (2012). *Thalassobius maritimus* sp. nov., isolated from seawater. *Int. J. Syst. Evol. Microbiol.* 62, 8–12. doi: 10.1099/ijs.0.029199-0

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

Petersen, J., and Wagner-Döbler, I. (2017). Plasmid transfer in the ocean - a case study from the Roseobacter group. *Front. Microbiol.* 8:1350. doi: 10.3389/fmicb.2017.01350

Pujalte, M. J., Lucena, T., Ruvira, M. A., Arahal, D. R., and Macián, M. C. (2014). "The family Rhodobacteraceae," in *The Prokaryotes—Alphaproteobacteria and Betaproteobacteria*, *4th Edn.*, *Vol. 8*, eds E. Rosenberg, E. F. DeLong, E. Stackebrandt, S. Lory, and F. Thompson (Berlin: Springer Verlag), 439–512.

Richter, M., Rosselló-Móra, R., Glöckner, F. O., and Peplies, J. (2015). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Rodrigo-Torres, L., Pujalte, M. J., and Arahal, D. R. (2016a). Draft genome sequence of *Thalassobius mediterraneus* CECT 5383T, a poly-beta-hydroxybutyrate producer. *Genomics Data* 7, 237–239. doi: 10.1016/j.gdata.2016.01.013

Rodrigo-Torres, L., Pujalte, M. J., and Arahal, D. R. (2016b). Draft genome sequence of *Shimia marina* CECT 7688T. *Mar. Genomics* 28, 83–86. doi: 10.1016/j.margen.2016.01.006

Rodrigo-Torres, L., Pujalte, M. J., and Arahal, D. R. (2017). Draft genome sequence of *Thalassobius gelatinovorus* CECT 4357T, a roseobacter with the potential ability to degrade polycyclic aromatic hydrocarbons. *Gene Rep.* 9, 32–36. doi: 10.1016/j.genrep.2017.08.005

Rüger, H. J., and Höfle, M. (1992). Marine star-shaped-aggregate-forming bacteria: *Agrobacterium atlanticum* sp. nov.; *Agrobacterium meteori* sp. nov.; *Agrobacterium ferrugineum* sp. nov., nom. rev.; *Agrobacterium gelatinovorum* sp. nov., nom. rev.; and *Agrobacterium stellulatum* sp. nov., nom. rev. *Int. J. Syst. Bacteriol.* 42, 133–143. doi: 10.1099/00207713-42-1-133

Sasser, M. (1990). *Identification of Bacteria by Gas Chromatography of Cellular Fatty Acids, MIDI Technical Note 101*. Newark, DE: MIDI Inc.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Simon, M., Scheuner, C., Meier-Kolthoff, J. P., Brinkhoff, T., Wagner-Döbler, I., Ulbrich, M., et al. (2017). Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 11, 1483–1499. doi: 10.1038/ismej.2016.198

Smibert, R. M., and Krieg, N. R. (2007). "Phenotypic characterization," in *Manual of Methods for General Bacteriology*, eds C. A. Reddy, T. J. Beveridge, J. A. Breznak, G. Marzluf, T. M. Schmidt, and L. R. Snyder (Washington, DC: American Society for Microbiology), 607–654.

Spiekermann, P., Rehm, B. H. A., Kalschenuer, R., Baumeister, D., and Steinbüchel, A. (1999). A sensitive, viable-colony staining method usig Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds. *Arch. Microbiol.* 171, 73–80. doi: 10.1007/s002030050681

Tang, K., Huang, H., Jiao, N., and Wu, C. H. (2010). Phylogenomic analysis of marine Roseobacters. *PLoS ONE* 5:e11604. doi: 10.1371/journal.pone.0011604

Uchino, Y., Hirata, A., Yokota, A., and Sugiyama, J. (1998). Reclassification of marine agrobacterium species: proposals of *Stappia stellulata* gen. nov., comb. nov., *Stappia aggregata* sp. nov., nom. rev., *Ruegeria atlantica* gen. nov. comb. nov., *Ruegeria gelatinovora* comb. nov., *Ruegeria algicola* comb. nov., and *Ahrensia kieliense* gen. nov., sp. nov., nom. rev. *J. Gen. Appl. Microbiol.* 44, 201–210. doi: 10.2323/jgam.44.201

Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., Glöckner, F. O. et al. (2010). Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. *Syst. Appl. Microbiol.* 33, 291–299. doi: 10.1016/j.syapm.2010.08.001

Yi, H., and Chun, J. (2006). *Thalassobius aestuarii* sp. nov., isolated from tidal flat sediment. *J. Microbiol.* 44, 171–176.

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA and whole genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

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 Pujalte, Lucena, Rodrigo-Torres and Arahal. 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: 07 December 2017
doi: 10.3389/fmicb.2017.02422

Clarification of Taxonomic Status within the *Pseudomonas syringae* Species Group Based on a Phylogenomic Analysis

Margarita Gomila[1], *Antonio Busquets*[1], *Magdalena Mulet*[1], *Elena García-Valdés*[1,2] *and Jorge Lalucat*[1,2]*

[1] Microbiology, Department of Biology, Universitat de les Illes Balears, Palma de Mallorca, Spain, [2] Institut Mediterrani d'Estudis Avançats (Consejo Superior de Investigaciones Científicas—Universidad de las Islas Baleares), Palma de Mallorca, Spain

The *Pseudomonas syringae* phylogenetic group comprises 15 recognized bacterial species and more than 60 pathovars. The classification and identification of strains is relevant for practical reasons but also for understanding the epidemiology and ecology of this group of plant pathogenic bacteria. Genome-based taxonomic analyses have been introduced recently to clarify the taxonomy of the whole genus. A set of 139 draft and complete genome sequences of strains belonging to all species of the *P. syringae* group available in public databases were analyzed, together with the genomes of closely related species used as outgroups. Comparative genomics based on the genome sequences of the species type strains in the group allowed the delineation of phylogenomic species and demonstrated that a high proportion of strains included in the study are misclassified. Furthermore, representatives of at least 7 putative novel species were detected. It was also confirmed that *P. ficuserectae*, *P. meliae,* and *P. savastanoi* are later synonyms of *P. amygdali* and that "*P. coronafaciens*" should be revived as a nomenspecies.

Keywords: ***P. syringae,*** **phylogenetic group, phylogenomic species, core genome, pangenome, ANIb, GGDC, MLSA**

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Boris Alexander Vinatzer, Virginia Tech, United States
Alexander Ignatov, R&D Center "Phytoengineering" LLS, Russia

***Correspondence:**
Jorge Lalucat
jlalucat@uib.es

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

Received: *15 October 2017*
Accepted: *22 November 2017*
Published: *07 December 2017*

Citation:
Gomila M, Busquets A, Mulet M, García-Valdés E and Lalucat J (2017) Clarification of Taxonomic Status within the Pseudomonas syringae Species Group Based on a Phylogenomic Analysis. Front. Microbiol. 8:2422. doi: 10.3389/fmicb.2017.02422

INTRODUCTION

The genus *Pseudomonas* is divided into two phylogenetic lineages (*Pseudomonas aeruginosa* and *Pseudomonas fluorescens*) based on inferred evolutionary relationships by using multilocus sequence analysis (MLSA) of four housekeeping genes (Mulet et al., 2010). The *P. fluorescens* lineage contains six phylogenetic groups, one of them represented by *Pseudomonas syringae*, and includes most of the phytopathogens within the genus *Pseudomonas* (Bull et al., 2010).

P. syringae was described by Van Hall (1902) and several closely related species have since been described. In the Approved List of Bacterial Names (Skerman et al., 1980), three other species of phytopathogenic *Pseudomonas* were also included: *Pseudomonas cichorii* (Stapp, 1928), *Pseudomonas viridiflava* (Burkholder, 1939), *Pseudomonas caricapapayae* (Robbs, 1956), and *Pseudomonas amygdali* (Psallidas and Panagopoulos, 1975). "*Pseudomonas coronafaciens*" (Schaad and Cunfer, 1979) was not included in the Approved List of Bacterial Names and is not recognized as a valid species name. Until that moment, species characterizations and proposals have been performed using physiological, biochemical, serological, and pathological traits. Later, several other species closely related to *P. syringae* were proposed and validated: *Pseudomonas*

meliae (Ogimi, 1977), *Pseudomonas savastanoi* (Gardan et al., 1992), *Pseudomonas ficuserectae* (Goto, 1983), *Pseudomonas avellanae* (Janse et al., 1996), *Pseudomonas cannabina* (Gardan et al., 1999), *Pseudomonas tremae* (Gardan et al., 1999), *Pseudomonas congelans* (Behrendt et al., 2003), *Pseudomonas asturiensis* (González et al., 2013), *Pseudomonas cerasi* (Kałuzna et al., 2016), and *Pseudomonas caspiana* (Busquets et al., 2017). The *P. syringae* species complex is usually considered to include all these taxonomically closely related species.

Molecular techniques based on experimental DNA-DNA hybridizations (DDH) or on DNA sequence analysis are essential in determining actual taxonomy. DDH were used first in the *P. syringae* group of species by Pecnold and Grogan (1973) and when *P. savastanoi* was proposed (Gardan et al., 1992). Gardan and colleagues established eight genomic groups, called genomospecies, based on DDH analysis (Gardan et al., 1999) that allowed the reclassification of strains previously known as pathovars of *P. syringae* as the new species *P. cannabina* and *P. tremae*. A phylogenetic study based on the 16S rRNA gene sequences of species in the genus was applied first by Moore et al. (1996) to propose a phylogenetic scheme within the genus, but until the description of *P. avellanae*, sequence analyses were not included in new species proposals in the *P. syringae* species complex. Due to the limitations in sequence variation in the 16S rRNA gene, other genes have been used for species delineation, especially the *rpoD* gene (Yamamoto et al., 2000; Mulet et al., 2010; Parkinson et al., 2011) and the *cts* gene (Berge et al., 2014). These analyses have allowed the delineation of phylogenetic groups, or phylogroups, within the species complex. Multilocus sequence analyses (MLSA) based on the sequences of three or four housekeeping genes have also been very successful in clarifying the phylogeny of strains in the *Pseudomonas* genus (Mulet et al., 2010; Bull et al., 2011; Berge et al., 2014). More specifically, Almeida et al. (2010) have developed the Plant associated microbes database (PAMDB) that contains sequences for MLST and MLSA accessible in a useful website. As a result of the molecular techniques, many strains have been reclassified and a more stable phylogenetic classification has become possible. Determining the precise taxonomic affiliations of strains in the *P. syringae* species complex can be difficult when pathovars are considered (Baltrus, 2016; Vinatzer et al., 2017). Currently, the species in the *P. syringae* phylogenetic group are subdivided into over 60 pathovars defined by pathogenic characters, 15 genomospecies are defined by DDH, 13 phylogroups are defined by MLSA using 3 or 4 genes, and 15 validly described species are accepted in the List of Prokaryotic Names with Standing in Nomenclature (Parte, 2014; http://www.bacterio.net/). Vinatzer and Bull (2009) have published a comprehensive history of the taxonomy of plant pathogenic bacteria, the use of MLSA and the impact of genomic approaches on taxonomy of plant pathogenic bacteria.

Genome-based taxonomic analyses have been recently introduced, and several algorithms are currently used for strain comparisons, such as the average nucleotide identity based on BLAST or MUMmer algorithms (ANIb, ANIm) and genome-to-genome distance calculations (GGDC), and are substituted for experimental DDH (Konstantinidis and Tiedje, 2007). Comparative genomics provides another tool that allows core genome and pangenome analyses at different levels of classification in a phylogenomic approach, that is, phylogenetic inference by combining many genes (Jeffroy et al., 2006). Recently, it has been proposed the use of similarity-based codes, called life identification numbers (LINs) to name individual bacterial isolates in the *P. syringae* species complex (Vinatzer et al., 2017).

As noted by Morris et al. (2017), "delineation of pertinent phylogenetic contours of plant pathogenic bacteria and naming of strains independent of their presumed life style is one of the five challenges for understanding the ecology of plant pathogenic bacteria." With the goal of clarifying the taxonomic delineation of species in the *P. syringae* phylogenetic group, 139 genomes of the 15 recognized species assigned to this group that are available in public databases have been analyzed by a phylogenomic approach. At least one member of each phylogroup described in the *P. syringae* phylogenetic branch by Berge et al. (2014) was included in the analyses if it was available in the public databases. "*P. coronafaciens*" strains and the three closely related species in the *Pseudomonas lutea* group (*P. graminis*, *P. lutea,* and *P. abietaniphila*) were also included, as well as an unclassified *Pseudomonas* sp. strain S25 isolated in our laboratory. MLSA and several *in silico* algorithms for genome comparisons (e.g., ANIb, GGDC) allowed the clustering of strains in 6 clear genomic branches. Core genome and pangenome analyses have been performed in the present study for the whole *P. syringae* phylogenetic group, for 5 of the 6 individual genomic branches and for 7 proposed phylogenomic species to explore their usefulness to delineate inter- and intra-species relationships. We included in our study the genome sequences of the type strains in the *P. syringae* group and 19 of the 56 pathotype strains recently published by Thakur et al. (2016) with the main purpose to clarify the species delineation, without considering all the pathovars. Species affiliation of the pathotypes is a prerequisite for the posterior study of the phylogeny of the pathovars.

MATERIALS AND METHODS

Data Collection and Genome Sequences

Draft and complete genome sequences of 139 strains belonging to different species of the *P. syringae* group available in the NCBI database were analyzed, including the genomes of 3 *P. aeruginosa* strains and 2 *P. stutzeri* strains used as outgroup. The 139 selected strains included the genomes of 15 species type strains of the *P. syringae* phylogenetic group, *Pseudomonas* sp. S25 and representatives of "*P. coronafaciens*." Genomes of strains in the *P. lutea* phylogenetic group (*P. lutea*, *P. graminis,* and *P. abietaniphila*) were used as an outgroup in the analysis because they belong to the closest phylogenetic group to the *P. syringae* group (Gomila et al., 2015). Six type strain genomes were analyzed in duplicate: five were representatives of the same type strain but from two different culture collections, and the sequence of *P. cannabina* was deposited twice by 2 different authors. The set of 139 genome sequences of *Pseudomonas* was retrieved from the GenBank database on 30th April 2017. The list of the 139

complete or draft genomes analyzed and additional details are provided in Supplemental Table S1.

Three-Gene Multilocus Sequence Analysis (3-Gene MLSA)

An MLSA based on the analysis of the partial sequences of the 16S rRNA, *gyrB* and *rpoD* genes was performed. The sequences of the 16S rRNA, *gyrB*, and *rpoD* genes were extracted from each genome studied and compared with the corresponding sequences of all *Pseudomonas* species type strains (161) described through 2016. Sequences are available in the public National Center for Biotechnology Information (NCBI) database. A concatenated gene tree was constructed using the individual alignments in the following order: 16S rRNA (1,309 nt), *gyrB* (803 nt), and *rpoD* (791 nt) by methods previously described (Gomila et al., 2015). Genomes that did not contain the 16S rRNA, *gyrB*, and *rpoD* gene sequences were removed from the MLSA concatenated analysis.

Whole-Genome Comparisons

In silico tools were used for genomic species delineation. Average nucleotide identity based on BLAST algorithm (ANIb) was calculated between all pairs of genomes, using the JSpecies software tool available at the webpage http://www.imedea.uib.es/jspecies (Konstantinidis and Tiedje, 2007; Richter and Rosselló-Móra, 2009). The recommended species cut-off was 95%. The similarity matrix obtained with all pairwise genomic comparisons was used to generate a UPGMA dendrogram using the PAST software. GGDC was performed between genome pairs on specific sets of genomes using the GGDC 2.0 update available in the web service http://ggdc.dsmz.de (Meier-Kolthoff et al., 2013)

Phylogenomic Comparisons

Pangenome Analysis and Clustering

A comparative genomic analysis was performed using the GET_HOMOLOGUES software described by Contreras-Moreira and Vinuesa (2013). All genomes were annotated with PROKKA for comparison purposes (Seemann, 2014), and the protein amino acid sequences obtained were compared using the criterion of 50% similarity over 50% of coverage alignment. Core genome and pangenome analyses were performed with three different clustering algorithms, bi-directional best-hits (BDBH), COGtriangle (COG), and OrthoMCL (OMCL). The four clusters determined from the analyses were defined as previously described (Koonin and Wolf, 2008; Kaas et al., 2012): core, soft core, shell, and cloud. Core genome and pangenome analyses were performed for all the genomes analyzed and for subsets of them.

Phylogenomic Analysis (Core MLSA)

All proteins codified by genes of the core genome that were present in monocopy were aligned, and the resulting alignments were concatenated. Elimination of poorly aligned positions and divergent regions of protein sequences were performed with Gblocks (Castresana, 2000), and the phylogenetic tree was constructed with the PhyML program (Guindon et al., 2010). Analysis of the concatenated amino acid sequences of the core proteins (core MLSA) was performed for all genomes and for the different delineated subsets.

Average Amino Acid Identities among Homologous CDSs

A GET_HOMOLOGUES script was used to estimate the average amino acid identities of CDSs between individual members of specific pangenome clusters. Gower's distance matrix were determined based on the percent amino acid identities of protein coding genes in the different genome branches using a script from GET_HOMOLOGUES Those distance matrices obtained were further illustrated as heatmaps, showing similarities and differences between genomes.

RESULTS

Genome Characteristics

Genome characteristics of the strains studied are summarized in Supplemental Table S1. The genome sequences of the 15 species type strains so far described in the *P. syringae* phylogenetic group were included. At least one member of 11 of the 13 phylogroups described in the *P. syringae* phylogenetic branch by Berge et al. (2014) was included in the analyses. Genomes of phylogroups 8 and 12 were not available in public databases. As a control, the sequences of the genomes of the *P. cichorii*, *P. viridiflava*, *P. congelans,* and *P. meliae* type strains were studied in duplicate, i.e., two type strains from two different culture collections. Two genome sequences of *P. cannabina* ICMP 2823^T with two different accession numbers were also studied. The studied genomes included 121 genomes with a status of "contig" (the number of contigs ranged from 5 to 5,099; mean: 617 contigs) and 6 with a status of "complete genome." The chromosome sizes ranged from 4,713,747 to 7,317,256 bp (mean: 5,976,989 bp) and the GC content in mol % ranged from 56.95 to 59.38 (mean: 58.34). Plasmids were reported in the databases for only 3 of the 6 closed genomes: *P. savastanoi* pv. *savastanoi* strain 1448A (2 plasmids, 3% of the genome content), *P. cerasi* strain 58^T (6 plasmids, 7% of the genome), and *P. syringae* pv. *tomato* DC3000 (2 plasmids, 2% of the genome). The plasmids were not included in the comparative analyses.

The authenticity of the type strains studied was checked by analyzing their affiliation in the *Pseudomonas* 3-gene MLSA tree (Gomila et al., 2015). All type strains were affiliated with the previously determined gene sequences with the exception of 2 species type strains, *P. tremae* and *P. lutea*. The genome of *P. tremae* ICMP 9151^T clustered close to those of *"P. coronafaciens"* strains, and the sequences were different from those published for *P. tremae* LMG 22121^T. Therefore, the published sequences of the *cts*, *gyrB*, *rpoB*, *rpoD*, aconitase, and 16S rRNA genes of the type strains of three different culture collections (LMG 22121^T, CFBP 3229^T, NCPPB 3465^T) were compared with the corresponding sequences of the *P. tremae* ICMP 9151^T genome. The sequences were only 88–98% identical. We concluded that the status of the species type strain of *P. tremae* ICMP 9151^T must be revised, and therefore it was not further considered as a type strain in the present study. Two genome sequences are available for *P. lutea* type strains.

Surprisingly, *P. lutea* LMG 21974^T was an outlier. The 16S rDNA gene sequence of strain LMG 21974^T was 99% identical to that of *Pseudomonas poae* DSM 14936^T in the *P. fluorescens* phylogenetic group. Several housekeeping genes of strain LMG 21974^T were analyzed, and it was concluded that the deposited *P. lutea* LMG 21974^T genome did not belong to the *P. lutea* phylogenetic group. The rest of the duplicated genome sequences were concordant.

Phylogenomic Analysis with Outgroups

All 139 genomes were phylogenetically analyzed using different strategies: (i) 3-gene MLSA of the partial sequences of the 16S rRNA, *gyrB* and *rpoD* genes, (ii) a phylogenomic tree based on the concatenation of all single-copy conserved protein sequences that conforms the core genome of the 139 genomes analyzed (core MLSA), and (iii) by ANIb.

(i) The 3-gene MLSA phylogenetic analysis included the 15 species type strains in the *P. syringae* phylogenetic group that are validly described and were used in a previous publication (Gomila et al., 2015) combined with the 139 *Pseudomonas* complete or draft genomes available in databases. A phylogenetic tree (Supplemental Figure S1) was generated based on the concatenated sequences with a total length of 2,796 nucleotides. One hundred and nine of the 139 strains (78% of the genomes analyzed) were affiliated with the corresponding species type strain, and their species assignments were considered correct.

(ii) One hundred and forty-nine monocopy genes were defined in the core genome of the whole set of 139 genomes. The phylogenomic tree obtained after the concatenation of the amino acid sequences of the 149 monocopy genes (core MLSA) is shown in Supplemental Figure S2. The *P. aeruginosa* and *P. stutzeri* genomes were used as outgroups. From the 33,744 positions obtained after the concatenation of the individual alignments, 93% of them were finally analyzed (31,400 positions). Bootstrap values were indicated on the nodes. In this phylogenomic tree, six main clusters or phylogenomic branches could be detected, indicated in Roman numerals from I to VI (Supplemental Figure S2).

(iii) Average nucleotide identities based on BLAST (ANIb) were calculated for the 139 genomes, obtaining a square matrix with 19,321 pairwise comparison values (Supplemental Table S2). A dendrogram was generated for this matrix to assess phylogenetic coherence (**Figure 1**). The ANIb dendrogram showed high topological congruence compared with the 3-gene MLSA phylogenetic tree and with the core MLSA phylogenetic tree. All duplicated type strains clustered together with ANIb values higher than 99.87%. The reference genomes of *P. aeruginosa*, *P. stutzeri,* and the species in the *P. lutea* group clustered outside the *P. syringae* group. All ANIb percentage values calculated were plotted on a graph that demonstrated a clear gap between 89 and 93% (Supplemental Figure S3). Only 206 values were observed between 93 and 96%, <2%, and corresponded to pairwise comparisons among strains in groups I and II. Six genomic branches were again delineated in the *P. syringae* group based on the observed tree branching in the ANIb dendrogram with an ANIb cut-off of 93%, such that members of different genomic branches could not have an ANIb of >93%. The genomic branches corresponded to the same six main phylogenetic branches detected in the 3-gene MLSA and in the core-gene phylogenetic tree. Therefore, each branch was considered a homogeneous phylogenomic branch, without taxonomic implications, to later facilitate comparative genomic analyses. The only exception detected was genomic branch II, which was divided into two clusters by ANIb value. The boundary values of the phylogenomic branches (minimal value among strains of different phylogenomic branches) were higher than 4.8%. The usually accepted cut-off for species delineation based on the ANIb lies between 95 and 96% (Richter and Rosselló-Móra, 2009). In branch I (represented by *P. syringae*) and branch IV (represented by *P. amygdali*), no clear gap at the 95–96% ANIb cut-off could be delineated; therefore, we were not able to distinguish genomic species (genospecies or genomospecies) by using only the ANIb value. Groupings of strains with intrabranch values higher than 94.3% in the other four branches were separated by clear gaps, and each group was considered a phylogenomic species. Seventeen clusters with ANIb intrabranch values higher than 94.3% could be differentiated, and their boundary values were higher than 4.4% with the exception of 9 strains in genomic branch I, which included the *P. syringae* type strain, as indicated in **Table 1** and Supplemental Table S2. Each of the 17 clusters was considered a phylogenomic species. Genomic branches I, IV, V, and VI contain more than one validly described species.

The results obtained for the three methodologies applied were compared. Strain clustering was maintained in the 3-gene MLSA at a cut-off of 97% (**Table 1**), which is in agreement with previous studies (Gomila et al., 2015), but some slight differences can be observed in the branching order (Supplemental Figure S1). Two strains with a 3-gene MLSA value lower than 97% cannot be assigned to the same species. Phylogenetic similarities in the analysis of the three concatenated genes were compared with the ANIb similarities calculated in the whole genome analysis and the results plotted in Supplemental Figure S4. A good correspondence between the ANIb and 3-gene MLSA indices could be observed. The six main clusters or genomic branches observed in the ANIb results were also detected in the core-gene MLSA tree, although phylogenomic branch III was divided into two closely related branches in the ANIb analysis.

GGDC similarities were also calculated for all genomes included in each ANIb/core MLSA genomic branch in order to clarify phylogenetic assignments to species. The results are shown in Supplemental Table S3 and were highly concordant with the ANIb values, accepting a species cut-off value of 70% as recommended by Meier-Kolthoff et al. (2013). It is worth mentioning that the 2 genomes available for the *P. cannabina* type strain were only 70.9% similar in the GGDC analysis but were almost identical in the ANIb (99.9% similar) and MLSA analyses (100% identical). This discrepancy has to be attributed to the *in silico* methodologies or to the quality of the genome sequences.

The combined use of the 4 indices allowed the delineation of 19 phylogenomic species, and these are described below in the context of each phylogenomic branch. ANIb values among members of different phylogenomic species were lower than 96%. Genomic branch I (intrabranch values: 93.11–98.18% for ANIb

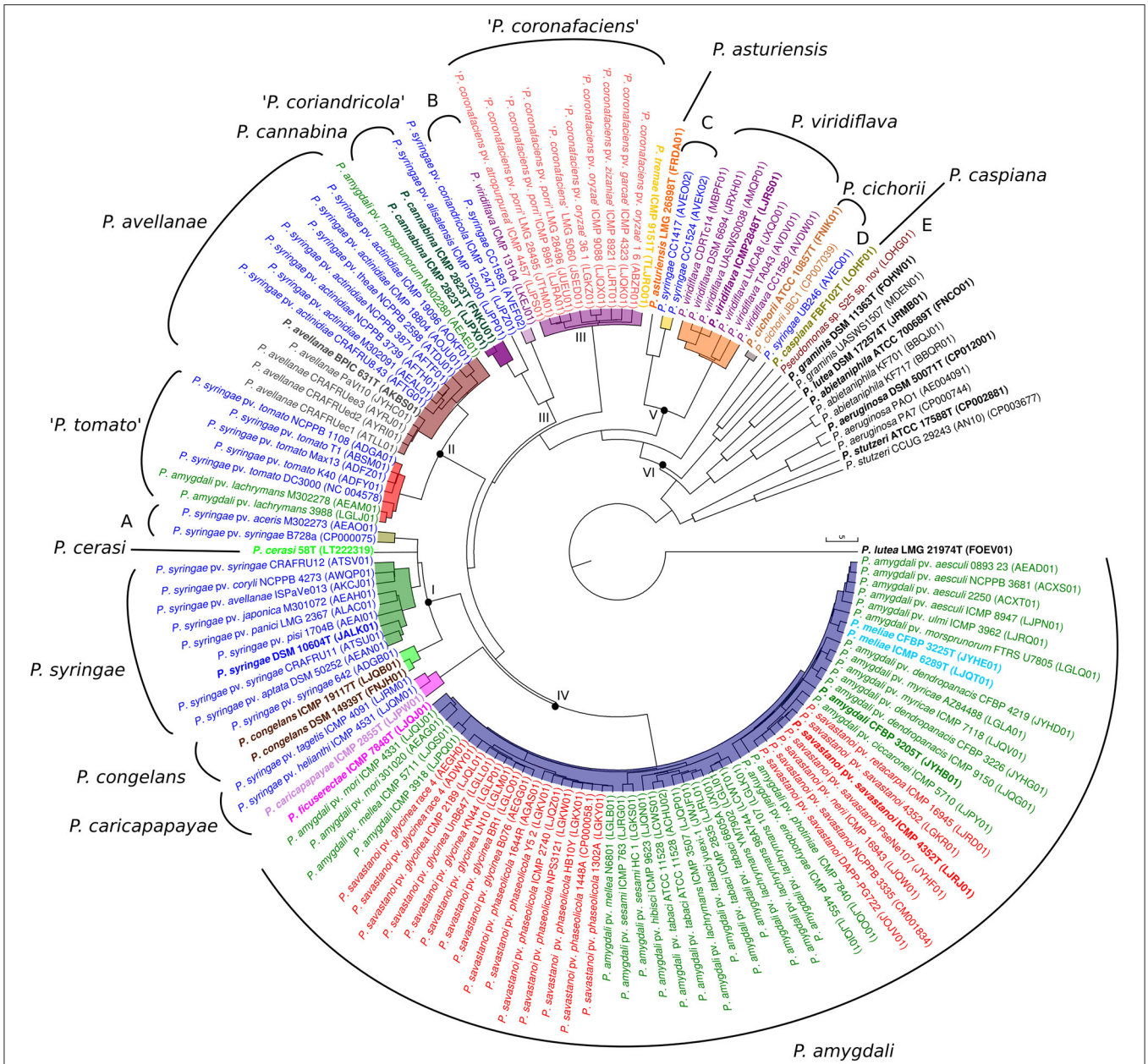


FIGURE 1 | UPGMA dendrogram based on ANIb values of all pairwise comparisons. Each species name, as submitted in the database, is labeled with different colors. Roman numerals at the corresponding nodes indicate phylogenomic branches defined. Phylogenomic species inside each phylogenetic branch are highlighted with different colors. Species type strains are labeled in bold. Accession numbers of the corresponding genomes are given in brackets. Proposed phylogenomic species are indicated in the external circle. Putative novel species are marked in quotation marks or by capital letters (A–E).

and 97.18–99.93% for 3-gene MLSA) included 15 strains divided into 4 groups, 3 of them belonging to recognized taxonomically described species: the *P. syringae* type strain and 8 closely related strains, *P. cerasi* (1 strain), and *P. congelans* (2 strains; the type strain is duplicated). The fourth group, designated as unnamed group A and represented by strain B728a, includes 2 strains. The boundaries at 95% ANIb were diffuse within this branch, although the 4 clusters could be clearly distinguished with GGDC values lower than the accepted cut-off of 70% (intrabranch values between the 4 groups ranged between 54.9 and 62.8%).

Genomic branch II, represented by the *P. avellanae* type strain and 19 other strains, presented intrabranch values of 94.30–99.99% for ANIb, 98.21–100% for 3-gene MLSA, and 59–99% for GGDC. Two homogeneous and clear sub-branches could be distinguished at a cut-off lower than 96% in ANIb (94.30–95.41%), which corresponded to 95.09% in 3-gene MLSA and values lower of 64% in GGDC and can be considered

TABLE 1 | ANIb, GGDC, 3 genes-MLSA, and core MLSA indices for the delineation of the proposed phylogenomic species in each genomic branch.

Genomic Branch	Phylogenomic species and representative strains	Nr of strains	ANIb		GGDC		3 genes–MLSA		Core MLSA	
			Minimal intra-cluster ANIb (%)	Closest inter-cluster ANIb (%)	Minimal intra-cluster GGDC (%)	Closest inter-cluster GGDC (%)	Minimal intra-cluster MLSA (%)	Closest inter-cluster MLSA (%)	Minimal intra-cluster Core (%)	Closest inter-cluster Core (%)
I	*P. congelans*	3	98.42	94.15	87.7	58.7	99.64	98.64	99.7	98.7
	P. syringae	9	94.79	95.15	68.4	62.8	98.75	98.39	98.8	98.9
	P. cerasi	1	–	95.07	–	61.2	–	97.47	–	99
	Species A (strain B728a)	2	98.27	95.21	87.6	62.8	99.25	98.96	99.6	98.9
II	*P. avellanae*	13	96.66	95.41	73.2	65	99.03	98.85	99.2	99.1
	"*P. tomato*" (DC3000)	7	98.41	95.41	89	65	99.35	98.85	99.7	99.1
III	*P.cannabina*	3	97.49	95.17	70.9	62.8	98.89	98.42	99.5	98.9
	"P. coriandricola" (ICMP 12471)	1	–	95.17	–	62.8	–	98.42	–	99
	species B (strain CC1583)	2	98.91	89.53	90.9	38.8	99.71	96.81	99.7	97.5
	"P. coronafaciens" (LMG 5060)	11	97.68	86.45	74	32.9	99.28	96.84	99.7	95.9
IV	*P. amygdali, P. meliae*, *P. savastanoi*, *P. ficuserectae*	57	96.79	89.2	76.4	39	97.45	97.54	98.8	98.1
	P. caricapapayae	3	97.3	89.38	78.8	39	98.97	97.49	99.5	98.1
V	*P. asturiensis*	1	100	94.63	–	59	–	98.57	–	98.8
	species C (strain CC1417)	2	99.05	94.63	92.7	59	99.64	98.57	99.9	98.8
	P.viridiflava	7	96.39	86.69	71.6	32.8	98.24	95.2	99.6	96.8
VI	*P. cichorii*	2	99.94	81.35	100	23.6	100	92.26	100	91.3
	species D (strain UB246)	1	–	88.52	–	36.8	–	97.31	–	97.6
	P. caspiana	1	–	88.52	–	32.6	–	96.31	–	97.5
	species E (strain S25)	1	–	86.93	–	23.5	–	97.31	–	97.5

The representative strain for the unnamed phylogenomic species is indicated in brackets.

phylogenomic species. Seven strains, represented by *P. syringae* pv. *tomato* DC3000, conformed a possible phylogenomic species. They grouped with similarities higher than 98.41%. The other phylogenomic species, represented by the *P. avellanae* type strain, together with 12 additional strains, grouped at 96.6% of ANIb, and in both cases, GGDC values were higher than 70%. Each phylogenomic species was circumscribed by uniform intrabranch values.

Intrabranch values of genomic branch III were 85.15–99.9% ANIb, 93.90–100% 3-gene MLSA, and 30–99.9% GGDC. This genomic branch included 17 strains distributed in 4 possible phylogenomic species: one included 2 *P. cannabina* type strain genome sequences and another strain identified as *P. syringae*; *P. syringae* pv. coriandricola ICMP 12471 was a singleton; 11 strains (10 of them classified as "*P. coronafaciens*," not taxonomically validly described, and the supposed type strain of *P. tremae*); and an unnamed phylogenomic species B (2 strains). The 4 sub-branches were also clearly distinguished at a threshold of 62% in GGDC.

Intrabranch values of genomic branch IV were 88.56–99.99% for ANIb, 96.13–100% for 3-gene MLSA, and 37.3–99.8% for GGDC. It included 5 validly described species type strains: the *P. meliae*, *P. amygdali, P. savastanoi,* and *P. ficuserectae* type strains, grouped together with 53 additional strains forming a unique phylogenomic species with clear boundaries from the rest of the strains; and *P. caricapapayae* (together with two additional strains). The 2 proposed phylogenomic species were clearly separated at the established cut-offs of 95% ANIb, 97% MLSA, and 70% GGDC.

Genomic branch V included two validly described species type strains with intrabranch values of 80.1–99.97% for ANIb, 91.68–100% for 3-gene MLSA, and 32.6–92.7% for GGDC. *P. asturiensis* was a singleton, and *P. viridiflava* was represented by 7 strains. Two strains formed a separate sub-branch (possible phylogenomic species C) close to *P. asturiensis*. The phylogenomic species were separated at cut-offs of 95% ANIb, 97% MLSA, and 70% GGDC.

Genomic branch VI was more distant and diverse. It included the type strains of *P. cichorii* and *P. caspiana*, together with other 2 strains: *Pseudomonas* sp. S25 and *P. syringae* UB246. The 4 strains were clearly separated at the accepted species cut-offs for ANIb, GGDC and 3-gene MLSA and have to be considered four different species.

As discussed later, these results suggested that a high proportion of genomes (53 of 127, 42%) were submitted in the databases with a species name affiliation different from that suggested by their affiliation in the ANIb, GGDC, 3-gene MLSA, and core MLSA dendrograms.

Species Assignations of Pathovars

Sixty-two pathovars of *P. syringae* are listed in the "Comprehensive list of names of plant pathogenic bacteria, 1980–2007" (Bull et al., 2010). Twenty-seven strains of *P. syringae* assigned to 15 different pathovars were included in the present study (3 pathovars with more than 1 representative: pv. *actinidiae* 6 strains, pv. *tomato* 5 strains, and pv. *syringae* 4 strains) to emphasize that the correct species affiliation is a prerequisite for the posterior study of the phylogeny of the pathovars. The 6 pv. *actinidiae* strains clustered together with *P. avellanae* strains. The 5 strains of pv. *tomato* clustered also together in one phylogenomic species included in genomic branch II. On the contrary, the 4 strains of pv. *syringae* affiliated to three different phylogenomic species (*P. congelans*, *P. syringae,* and the proposed new phylogenomic species A) in branch I. The rest of pathovars represented by single strains were distributed in branches I, II, III, and IV. Only 8 of the pathovars were affiliated with *P. syringae* phylogenomic species (e.g., *P. syringae* pv. *tagetis* ICMP 4091 and *P. syringae* pv. *helianthi* ICMP 4531 belonged to *P. caricapapayae*). It is worthy of note that five of the six *P. syringae* pathotype strains included in this study did not affiliate with *P. syringae* genomic species: *P. syringae pv. tagetis* ICMP 4091 belongs to *P. caricapapayae*; *P. syringae pv. actinidiae* NCPPB 3739 belongs to *P. avellanae*; *P. syringae pv. alisalensis* ICMP 15200 belongs to *P. cannabin*a; and finally *P. syringae pv. coriandricola* ICMP 12471 was a putative new phylogenomic species. This suggested the need to reclassify the misclassified strains.

Most of the 14 pathovars of the 33 *P. amygdali* strains clustered in the same phylogenomic species with the exception of *P. amygdali* pv. *morsprunorum*, that clustered with *P. avellanae* strains, and 2 *P. amygdali* pv. *lachrymans* that clustered with *P. syringae* pv. *tomato* strains. The 5 pathovars of the 23 *P. savastanoi* strains clustered together in the same phylogenomic species represented by *P. amygdali*. The 5 pathovars of the 9 strains of *P. coronafaciens* affiliated to the same phylogenomic species. Only 3 pathovars (*lachrymans*, *morsprunorum,* and *syringae*) were assigned to more than 1 phylogenomic species.

Core Genome and Pangenome Analysis of the *P. syringae* Group

To facilitate later comparative genomic analyses and due to the good correspondence between the 4 methods, the 6 genomic branches were maintained for further comparative genomic analysis (**Figure 1**, Supplemental Figure S2 and **Table 1**). Similarities between the members of different branches were always lower than 90% in ANIb, lower than 97% in 3-gene MLSA and lower than 70% in GGDC analyses.

Core genome and pangenome analyses were performed for the whole set of strains selected belonging to the *P. syringae* group (127 genomes) and for each of the five genomic branches (I–V) delineated by the previous methodologies. Members of group VI were not included in the analyses because they were more distantly related and only 5 strains representing 4 phylogenomic species were available. Each set of genomes was analyzed with the GET_HOMOLOGUES software. Different images were produced for each pangenome analysis: (1) a Venn diagram of core genomes generated by the three algorithms BDBH, COG, and OMCL and of pangenomes generated by COG and OMCL algorithm, (2) the core genome size was estimated with the Tettelin and Willenbrock fits and the pangenome size with the Tettelin fit, and (3) the partition of the OMCL pangenomic matrix into shell, cloud, soft core, and core compartments (**Figure 2**, Supplemental Figures S5–S9).

The core genome of the 127 strains in the *P. syringae* phylogenetic group contained 343 genes. Two hundred and nineteen of them were in monocopy and were concatenated and analyzed to establish their phylogenetic relationships as previously described. The clustering of strains in the 6 phylogenetic groups was identical to those obtained with the other indices (ANIb, GGDC, 3-gene MLSA, and core MLSA of 139 genomes), showing the same branching order and supported by high bootstrap values (100) (**Figure 3**). Genomic branch II was the only exception, being separated from groups I, III, and IV with a bootstrap value of only 12. Bull et al. (2011) also observed this result. The 5 main phylogenomic branches were also supported by a high number of shared genes, representing the core genome proteins as a minimum of 20% of the whole genome, as indicated in **Table 2**.

Each branch was analyzed separately to assess the potential use of the shared genes for species delineation. **Table 2** summarizes the number of core genome and pangenome genes calculated for all genomes and for each specific group analyzed. **Table 2** shows the genes present in the "soft core" (genes present in 95% of the genome analyzed), in the "shell" cluster (genes moderately common in the pangenome, present in >10% and <89% of the genomes and in the "cloud" cluster (genes present in very few of the genomes analyzed, 2 or less). The "core" and "soft core" clusters include highly conserved genes with phylogenetic information (Bezuidt et al., 2016). The shell and the cloud clusters represent subsets of the flexible genome, which reflect the adaptation of strains to particular environments and also the evolutionary history these organisms. The 127 genomes of the *P. syringae* phylogenetic group show a high percentage of genes in the flexible genome, indicating that these strains are able to share genes and are highly diverse.

Average amino acid identity matrices were calculated using protein-CDSs within the 5 phylogenomic branches (Supplemental Figures S5–S9). The heatmap shows the clustering of genomes into different groups based on average similarities and differences of their CDS amino acid identities. In this case, core and flexible gene pools are combined. The clustering of the genomes in each phylogenomic branch follows the same groupings observed with the other methodologies (3-gene MLSA, core MLSA, ANIb, and GGDC). That is, in phylogenomic branch I, the four phylogenomic species detected with the previous methodologies could also be differentiated by clear boundaries.

The amino acid sequences of the core genes for each phylogenomic branch were concatenated, and the phylogenetic tree was constructed with PhyML. The phylogenetic trees are depicted in Supplemental Figures S5–S9. The clustering

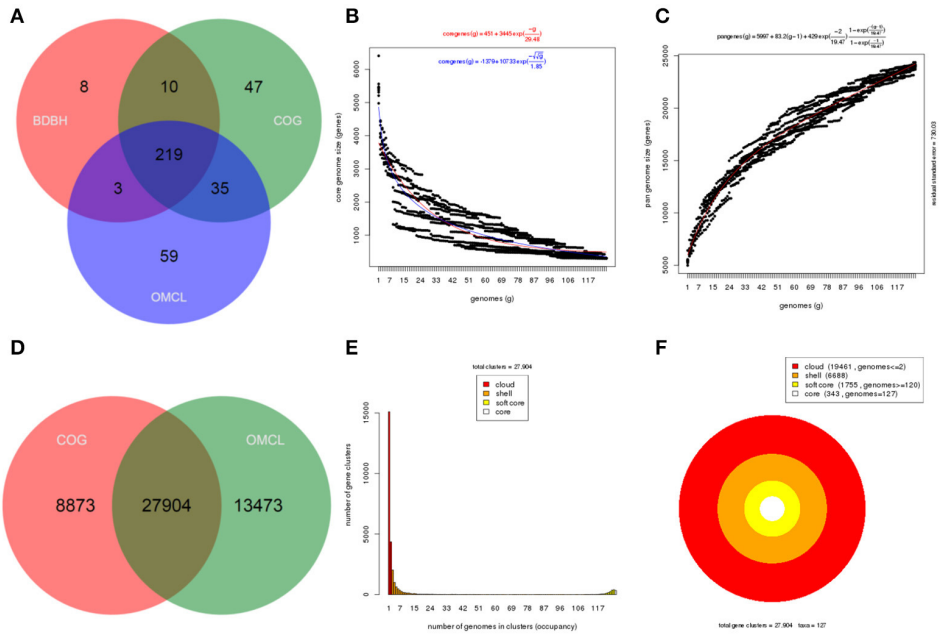


FIGURE 2 | Core and pangenome analysis of the 127 strains in the *P. syringae* phylogenetic group. **(A)** Venn diagram of core genomes generated by the BDBH, COG, and OMCL strategies. **(B)** Estimate of core genome size with the Tettelin (blue) and Willenbrock (red) fits. **(C)** Estimate of pangenome size with the Tettelin fit. **(D)** Venn analysis of pangenomes generated by COG and OMCL. **(E,F)** Partition of the OMCL pangenomic matrix into shell, cloud, soft-core, and core compartments. These plots can be easily created with GET_HOMOLOGUES auxiliary scripts, as explained in the manual.

observed was consistent with the results obtained previously and supported the proposed phylogenomic species.

The core genome and pangenome were also analyzed for 6 delineated phylogenomic species that contained at least 7 strains each. Four of them included the type strains of the validly described species *P. syringae*, *P. avellanae*, *P. amygdali,* and *P. viridiflava*. Additionally, genomes of the proposed phylogenomic species "*P. coronafaciens*" and "*P. tomato*," were also studied. The results are shown in **Table 3** and Supplemental Figures S10–S15. The pool of conserved genes decreased with increasing genome number and represents at least 17% of the individual genomes. The percentage of the core genes in the 6 proposed phylogenomic species ranged between 17.67 and 72.60%. The low percentage of conserved genes might be directly associated with the difficulty to phylogenetically assign some strains to the correct phylogenomic species.

This low percentage of conserved genes correlates with a higher number of cloud genes that contribute to the flexible genome.

DISCUSSION

Although the taxonomy of *Pseudomonas*, and more specifically of the *P. syringae* phylogenetic group, has been extensively analyzed, significant uncertainties remain regarding the genus boundaries and species composition of this heterogeneous taxon. Many species have been named without adequate descriptions, or their identifications have not been updated with more modern techniques. Phylogenomic insights published by Gomila et al. (2015) and more recently by Tran et al. (2017) have substantially improved the knowledge of the whole genus. Vinatzer et al. (2017) used genome similarities to study the taxonomy of plant-pathogenic bacteria and constructed core genome phylogenies for plant-pathogenic bacteria. The composition of distinct species and whether *P. syringae* is a cohesive unit has been debated for a long time (Janse et al., 1996; Bull et al., 2010). Several recent publications have tried to clarify this situation (Baltrus, 2016; Baltrus et al., 2017), also at the pathovar classification level (Thakur et al., 2016).

Strains were originally identified phenotypically as members of the *P. syringae* complex if they were fluorescent pseudomonads, positive for levan sucrase activity, negative for oxidase activity, unable to rot potato, able to produce arginine dihydrolase and able to cause a hypersensitive response on tobacco (the LOPAT group 1 strains; Lelliott et al., 1966; Sands et al., 1970). In 1975, numerous formerly distinct LOPAT group 1 plant pathogenic species were combined into the species *P. syringae* (Lapage et al., 1975), and the confusion increased due to subspecific pathovar names have been given to distinct pathogenic characters and host of isolation (Young, 2008). At that time a large number of nomenspecies of these bacteria were defined and became widely regarded as host-adapted pathogenic varieties (pathovars). Consequently, the Approved Lists of Bacterial Names did not list most of these nomenspecies, which thus lost standing in nomenclature. Main reason were the absence of deposited strains in culture collections, lack of adequate phenotypic descriptions and phenotypic traits that distinguished the proposed species names. The International Society of Plant Pathologists published a checklist of the earlier nomenspecies and pathovars (Dye et al., 1980) and advised that

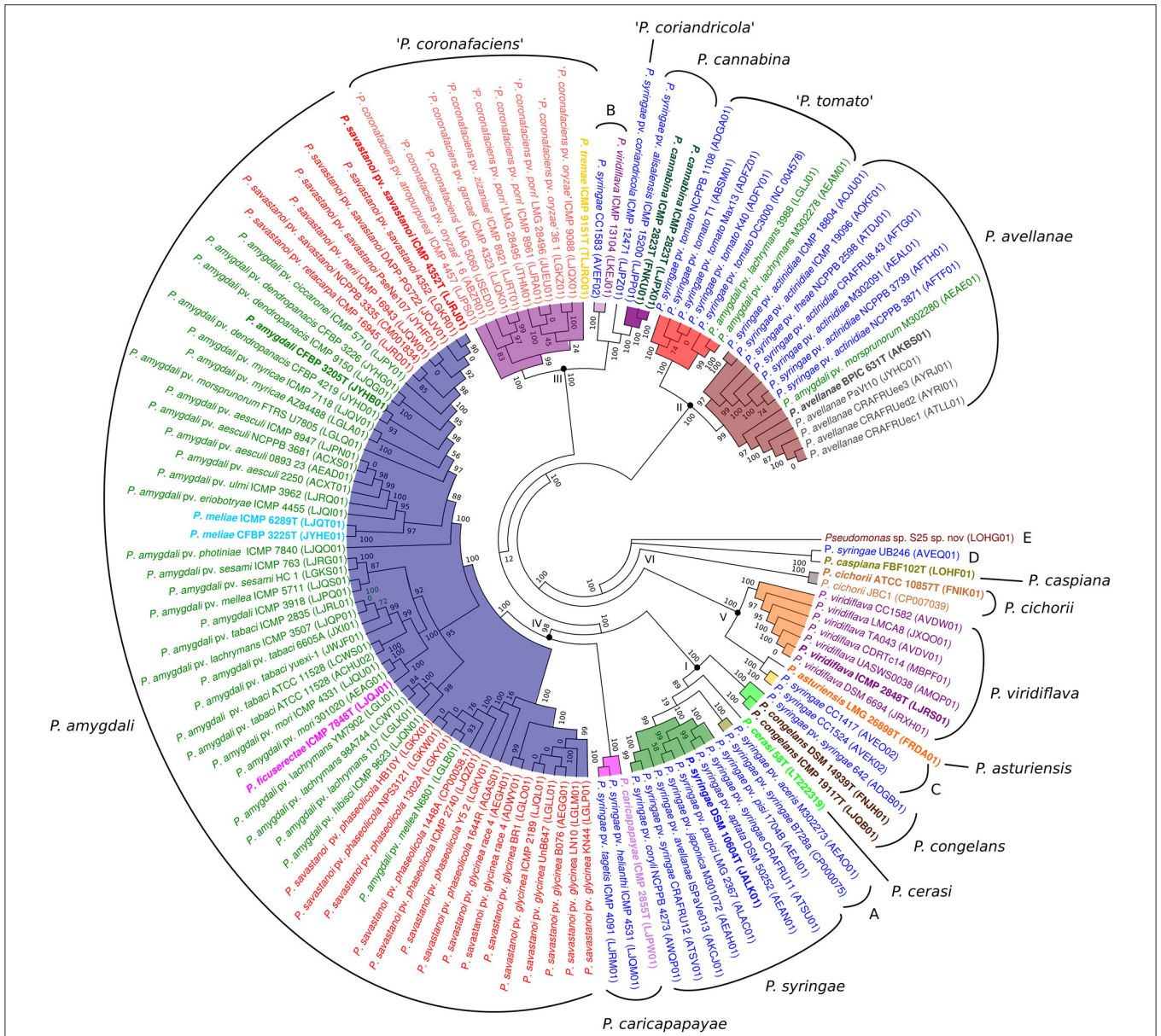


FIGURE 3 | Phylogenetic tree of the concatenated amino acid sequences of 219 monocopy proteins of the core genome defined in the 127 genomes analyzed. Forty-three thousand one hundred thirty-three amino acid positions were used to construct the tree. Each species name, as submitted in the database, is labeled with different colors. Roman numerals at the corresponding nodes indicate phylogenomic branches defined. Phylogenomic species inside each phylogenetic branch are highlighted with different colors. Species type strains are labeled in bold. Accession numbers of the corresponding genomes are given in brackets. Proposed phylogenomic species are indicated in the external circle. Putative novel species are marked in quotation marks or by capital letters (A–E). All bootstrap values are indicated in the nodes.

such names should be revived only for the original bacteria (Lapage et al., 1992).

Classification based only on phenotype has led to increased taxonomic confusion, as more *P. syringae* strains have been isolated from different environments, including non-diseased tissues and environmental sources, such as rivers, lakes, snowfields, and clouds (Morris et al., 2008). Phenotypic diversity of strains in relevant species has also been demonstrated (Demba Diallo et al., 2012; Bartoli et al., 2014) and many strains cannot be easily classified. Currently, the *P. syringae* species complex is subdivided into over 60 pathovars defined by pathogenic characters, nine genomospecies defined by DDH and 13 phylogenetic groups (phylogroups) defined by multilocus sequence analysis (Sarkar and Guttman, 2004; Hwang et al., 2005; Almeida et al., 2010; Bull et al., 2011; Berge et al., 2014).

The aim of this work is to try to circumscribe the *P. syringae* species complex and classify its strains into species according to the taxonomic rules and thresholds actually accepted in

TABLE 2 | Core and pangenome analyses of the 127 strains included in the *P. syringae* phylogenetic group, as well as for 5 of the individual groups defined (I–V).

	***P. syringae* phylogenetic group**	**Group I**	**Group II**	**Group III**	**Group IV**	**Group V**
Nr of strains	127	15	20	17	57	10
Coregenome proteins	343	1,938	1,229	2,367	1,694	3,900
Pangenome proteins	27,904	13,150	13,493	11,873	16,810	7,704
Cloud	19,461	8,139	7,559	6,366	9,826	2,329
Shell	6,688	1,888	3,715	2,189	4,014	1,254
Soft core	1,755	3,123	2,219	3,318	2,970	4,121
% Conserved genes	6.29	23.75	16.45	27.95	17.67	53.49
% Flexible genome	93.71	76.25	83.55	72.05	82.33	46.51

Numbers of genes in the shell, cloud, soft-core, and core compartments are indicated. Percentages of conserved genes and flexible genome are also given.

TABLE 3 | Core and pangenome analyses of the phylogenomic species proposed with more than 6 strains in the *P. syringae* phylogenetic group.

	Group I	**Group II**		**Group III**	**Group IV**	**Group V**
	P. syringae	***"P. tomato"***	***P. avellanae***	***"P. coronafaciens"***	***P. amygdali***	***P. viridiflava***
Nr of strains	9	7	13	11	57	7
Coregenome proteins	2,185	4,197	1,329	2,888	1,694	4,438
Pangenome proteins	11,966	7,022	12,361	9,674	16,810	6,452
Cloud	7,210	1,605	7,007	4,512	9,826	1,244
Shell	1,394	875	2,952	1,145	4,014	524
Soft core	3,392	4,542	2,402	4,017	2,970	4,684
% Conserved genes	28.35	64.68	19.43	41.52	17.67	72.60
% Flexible genome	71.90	35.32	80.57	58.48	82.33	27.40

Numbers of genes in the shell, cloud, soft-core, and core compartments are indicated. Percentages of conserved genes and flexible genome are also given.

taxonomy. Considering all methods tested together, we were able to circumscribe 6 phylogenomic branches within the *P. syringae* phylogenetic group.

The first branch, branch I, included 15 strains divided into 4 groups corresponding to four different phylogenomic species: *P. syringae*, *P. congelans*, *P. cerasi,* and novel species A. The first one includes the *P. syringae* type strain, together with only 9 strains previously identified as *P. syringae*. The rest of the *P. syringae* strains included in the present study (23) must be reassigned to other species (see Supplemental Table S1).

A second branch, branch II, contains 20 strains: 5 *P. avellanae* strains, including the species type strain, together with 1 strain classified as *P. amygdali* pv. *morsprunorum* (M302280), 5 strains of *P. syringae* pv. *tomato*, 2 of *P. amygdali* pv. *lachrymans*, 7 of *P. syringae* pv. *actinidiae*, and 1 of *P. syringae* pv. *theae*. We have included in this study five strains not considered in the previous publication by Scortichini et al. (2013) on *P. avellanae* genomes. The 3-gene MLSA values for all strains of this group ranged from 99.46 to 100% and divided these 20 strains into two subclusters. These 2 subclusters are maintained in the analysis of 219 concatenated core genes. All strains in the branch have ANIb similarity values ranging from 94.3 to 100%, but 2 sub-branches can be delineated: one at 97.1–100%, which corresponds to the group of the *P. avellanae* type strain, and another at 98.4–100%, which corresponds to strains of *P. syringae* pv. *tomato* and *P. amygdali* pv. *lachrymans*. These two sub-branches can also be delineated with the GGDC values; they are in accordance with Gardan's genomospecies 3 and 8, respectively, and with the phylogroups established by Parkinson and Berge. If we apply the currently accepted species threshold, all the strains of the group might be assigned to the species *P. avellanae*, but attending to all indices tested and the boundaries of the two subclusters, the possibility to differentiate two species or 2 subspecies must be considered. We have distinguished 2 distinct phylogenomic species: one is *P. avellanae*, and we propose the provisional operative name of "*Pseudomonas tomato*" for the branch that includes strain DC3000, pending a deeper taxonomic analysis. To formally propose a new species clear phenotypic characteristics that differentiate the new species with its closely related species have to be found. In our experience, the whole-cell protein profiles obtained by MALDI-TOF mass spectrometry can be a method of choice to phenotypically discriminate new species in the genus *Pseudomonas* (Mulet et al., 2012).

Branch III includes 4 phylogenomic species: 3 strains in one cluster must be assigned to *P. cannabina*; 1 strain deposited as *P. viridiflava* and another as *P. syringae* must be considered representatives of new species B; *P. syringae* pv. *coriandricola* ICMP 13104 represents a phylogenomic species provisionally named "*Pseudomonas coriandricola*"; 10 strains initially assigned to "*P. coronafaciens*" cluster together in all analyses and constitute the fourth phylogenomic species. This

group includes the genome of *P. tremae* ICMP 9151^T, not considered type strain in the present study. Therefore, the genome of *P. tremae* LMG 22121^T was sequenced and ANIb and GGDC results demonstrated that it has to be included in genomic branch IV (results not shown), that strengthen the possible misclassification of strain ICMP 9151^T. The "*P. coronafaciens*" strains belong to phylogroup 4 of Parkinson and Berge and to genomospecies 4 of Gardan. The species "*P. coronafaciens*" has been proposed by Schaad and Cunfer (1979) based on phenotypic characteristics, and the indices studied clearly support the revival of "*P. coronafaciens*" as a nomenspecies.

Two phylogenomic species can be delineated in branch IV, which correspond to phylogroup 6 of Parkinson and Berge and genomospecies 7 of Gardan. Two strains and the *P. caricapapayae* type strain must be assigned to this species. The other phylogenomic species is more abundant and very homogeneous and contains 4 accepted nomenspecies. As already noted by Gardan et al. (1999), strains in this phylogenomic species have been assigned to genomospecies 3 and must be considered *P. amygdali* strains. The other 3 are later synonyms (*P. ficuserectae*, *P. meliae,* and *P. savastanoi*). As mentioned before *P. tremae* should be also considered in this group as a later synonym once will be assessed the correct genome.

Three well-defined phylogenomic species were distinguished in branch V. One was formed by *P. asturiensis* strains, another for a novel species C with 2 strains, and the other by *P. viridiflava* strains. Strains in this branch shared at least 53% of the genes in the pangenome. The strains *P. syringae* CC1417 and *P. syringae* CC1524 are considered non-phytopathogens, and *P. asturiensis* LMG 26898^T is phytopathogenic. The three strains are closely related in all the indices tested and were isolated from distant geographical areas (Montana, USA; France and Spain, respectively) and ecological habitats (rocks in waterfall in pristine woods, stream water and soybeans, respectively; Morris et al., 2008; González et al., 2013; Baltrus et al., 2014). The ANIb and GGDC indices are near the borderline of the species acceptance cut-off and share at least 87% of the pangenome in the Gower analysis. Consequently, strains CC1417 and CC1524 are not members of the *P. syringae* phylogenomic species and might be considered strains of *P. asturiensis*. However, the differences in plant pathogenicity and other practical reasons suggest that the taxonomic status of both strains merit further analyses before a definitive classification and must be considered for the moment representatives of putative new species C.

In branch VI, four phylogenomic species have been defined: the *P. caspiana* type strain is a representative of the other 4 strains of the species (Beiki et al., 2016; Busquets et al., 2017); *P. cichorii* with two strains, *Pseudomonas* sp. S25 and *P. syringae* UB246 are singletons, and more closely related strains are needed for a definitive taxonomic assignment of both these strains, which are assigned to putative new species D and E.

Overall, we were able to distinguish 19 phylogenomic species in the *P. syringae* phylogenetic group distributed within 6 phylogenomic branches. Two strains are assigned to 2 different phylogenomic species when the following criteria are accomplished: (i) ANIb value is lower than 94.5%, (ii) GGDC values lower than 68%, and (iii) 3-gene MLSA similarity lower than 98%. ANIb values between 94.5 and 96% might be analyzed carefully with respect to other characteristics such as GGDC, the core genome and pangenome analyses. In general, very good agreement has been found between these phylogenomic species, the phylogroups of Parkinson and Berge, and the genomospecies of Gardan. In fact, Bull et al. (2011) showed also how MLSA quite accurately reflects the genomospecies described by Gardan et al. (1999) by experimental DNA-DNA hybridizations.

A strain was assigned to a given phylogenomic species when it fell into one of the 19 phylogenomic species delineated as described above. In 58% of strains, there was an agreement between strain name and genomospecies. Furthermore, 23 out of 32 (72%) strains deposited as *P. syringae* were not assigned to the *P. syringae* phylogenomic species but were scattered among 10 different phylogenomic species. This fact points out the importance of correctly assigning a genome to the right species. Thanks to NGS technologies, a remarkable increase in the number of sequenced genomes, both draft and complete, are available, but the correct assignment of the sequenced strains to the corresponding species with the accepted taxonomic tools is important before comparative analyses with other genomes can be performed.

Genomic data are very useful in the actual taxonomy to delineate phylogenomic species that merits the species status. However, it is possible that many species will be separated in several species, even when the abundance of species names can be confusing. In this sense, the use by the experts in phytopathology will consolidate or not the use of these new species names. In many practical issues it can be maintained the less precise concept of *P. syringae* species complex for all of them, although it is essential a proper naming of bacterial species in order to establish a truly systematic taxonomy and avoid confusions in the scientific communities.

CONCLUSIONS

Comparative genomics is a very useful tool for the establishment of a stable taxonomy, and we demonstrate its usefulness for the plant pathogenic bacteria studied in the present manuscript. Although further taxonomic studies are needed to support formal proposals, based on the present study of strains in the *P. syringae* phylogenetic group, we suggest that *P. ficuserectae*, *P. meliae,* and *P. savastanoi* are later synonyms of *P. amygdali* and, therefore, the group includes 11 recognized nomenspecies: *P. amygdali, P. asturiensis*, *P. avellanae*, *P. cannabina*, *P. caricapapayae*, *P. caspiana*, *P. cerasi, P. cichorii*, *P. congelans*, *P. syringae,* and *P. viridiflava*. Additionally, "*P. coronafaciens*" should be revived as a nomenspecies, and 27 strains representing 7 putative new species must be considered.

AUTHOR CONTRIBUTORS

MG, EG-V, and JL: conceived and designed the research project, and analyzed the data; AB and MM performed

the experiments; MG: performed the bioinformatic analyses of the data; MG, AB, MM, EG-V, and JL: interpreted the results and contributed to the writing of the manuscript.

FUNDING

Financial support was obtained from the Spanish MINECO through project CGL2015-70925-P (with FEDER cofunding). MG was supported by a postdoctoral contract from the Conselleria d'Innovació, Recerca i Turisme del Govern de les Illes Balears and the European Social Fund.

ACKNOWLEDGMENTS

We wish to thank to Dr. Bruno Contreras-Moreira and Dr. Pablo Vinuesa for the useful help with the Get_Homologues scripts. We would like to thank the reviewers for their detailed comments and suggestions for the manuscript. Their constructive criticism and valuable comments have improved the manuscript content.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02422/full#supplementary-material

REFERENCES

Almeida, N. F., Yan, S., Cai, R., Clarke, C. R., Morris, C. E., Schaad, N. W., et al. (2010). PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. *Phytopathology* 100, 208–215. doi: 10.1094/PHYTO-100-3-0208

Baltrus, D. A. (2016). Divorcing strain classification from species names. *Trends Microbiol.* 24, 431–439. doi: 10.1016/j.tim.2016.02.004

Baltrus, D. A., McCann, H. C., and Guttman, D. S. (2017). Evolution, genomics and epidemiology of *Pseudomonas syringae*. Challenges in bacterial molecular pathology. *Mol. Plant. Pathol.* 18, 152–168. doi: 10.1111/mpp.12506

Baltrus, D. A., Yourstone, S., Lind, A., Guilbaud, C., Sands, D. C., Jones, C. D., et al. (2014). Draft genome sequences of a phylogenetically diverse suite of *Pseudomonas syringae* strains from multiple source populations. *Genome Announc.* 23:e01195-13. doi: 10.1128/genomeA.01195-13

Bartoli, C., Berge, O., Monteil, C. L., Guilbaud, C., Balestra, G. M., Varvaro, L., et al. (2014). The *Pseudomonas viridiflava* phylogroups in the *P. syringae* species complex are characterized by genetic variability and phenotypic plasticity of pathogenicity-related traits. *Environ. Microbiol.* 16, 2301–2315. doi: 10.1111/1462-2920.12433

Behrendt, U., Ulrich, A., and Schumann, P. (2003). Fluorescent pseudomonads associated with the phyllosphere of grasses; *Pseudomonas trivialis* sp. nov., *Pseudomonas poae* sp. nov. and *Pseudomonas congelans* sp. nov. *Int. J. Syst. Evol. Microbiol.* 53, 1461–1469. doi: 10.1099/ijs.0.02567-0

Beiki, F., Busquets, A., Gomila, M., Rahimian, H., Lalucat, J., and García-Valdés, E. (2016). New *Pseudomonas* spp. are pathogenic to citrus. *PLoS ONE* 26:e0148796. doi: 10.1371/journal.pone.0148796

Berge, O., Monteil, C. L., Bartoli, C., Chandeysson, C., Guilbaud, C., Sands, D. C., et al. (2014). A user's guide to a data base of the diversity of *Pseudomonas syringae* and its application to classifying strains in this phylogenetic complex. *PLoS ONE* 3:e105547. doi: 10.1371/journal.pone.0105547

Bezuidt, O. K., Pierneef, R., Gomri, A. M., Adesioye, F., Makhalanyane, T. P., Kharroub, K., et al. (2016). The *Geobacillus* pan-Genome: implications for the evolution of the genus. *Front. Microbiol.* 7:723. doi: 10.3389/fmicb.2016.00723

Bull, C. T., Clarke, C. R., Cai, R., Vinatzer, B. A., Jardini, T. M., and Koike, S. T. (2011). Multilocus sequence typing of *Pseudomonas syringae* sensu lato confirms previously described genomospecies and permits rapid identification of *P. syringae* pv. coriandricola and *P. syringae* pv. *apii* causing bacterial leaf spot on parsley. *Phytopathology* 101, 847–858. doi: 10.1094/PHYTO-11-10-0318

Bull, C. T., De Boer, S. H., Denny, T. P., Firrao, G., Fischer-Le Saux, M., Saddler, G. S., et al. (2010). Comprehensive list of names of plant pathogenic bacteria, 1980-2007. *J. Plant Pathol.* 92, 551–592. doi: 10.4454/jpp.v92i3.302

Burkholder, W. H. (1939). The bacterial diseases of the bean. *Mem. Cornell University Agric. Exp. Stat.* 127, 1–88.

Busquets, A., Gomila, M., Beiki, F., Mulet, M., Rahimian, H., García-Valdés, E., et al. (2017). *Pseudomonas caspiana* sp. nov., a citrus pathogen in the *Pseudomonas syringae* phylogenetic group. *Syst. Appl. Microbiol.* 40, 266–273. doi: 10.1016/j.syapm.2017.04.002

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Demba Diallo, M., Monteil, C. L., Vinatzer, B. A., Clarke, C. R., Glaux, C., Guilbaud, C., et al. (2012). *Pseudomonas syringae* naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic. *ISME J.* 6, 1325–1335. doi: 10.1038/ismej.2011.202

Dye, D. W., Bradbury, J. F., Goto, M., Hayward, A. C., Lelliott, R. A., and Schroth, M. N. (1980). International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Rev. Plant Pathol.* 59, 153–168.

Gardan, L., Bollet, C., Abu Ghorrah, M., Grimont, F., and Grimont, P. A. D. (1992). DNA relatedness among the pathovar strains of *Pseudomonas syringae* subsp. savastanoi Janse (1982) and proposal of *Pseudomonas savastanoi* sp. nov. *Int. J. Syst. Bacteriol.* 42, 606–612. doi: 10.1099/00207713-42-4-606

Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F., and Grimont, P. A. D. (1999). DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (*ex* Sutic and Dowson 1959). *Int. J. Syst. Evol. Microbiol.* 49, 469–478. doi: 10.1099/00207713-49-2-469

Gomila, M., Peña, A., Mulet, M., Lalucat, J., and García-Valdés, E. (2015). Phylogenomics and systematics in *Pseudomonas*. *Front. Microbiol.* 6:214. doi: 10.3389/fmicb.2015.00214

González, A. J., Cleenwerck, I., De Vos, P., and Fernández-Sanz, A. M. (2013). *Pseudomonas asturiensis* sp. nov., isolated from soybean and weeds. *Syst. Appl. Microbiol.* 36, 320–324. doi: 10.1016/j.syapm.2013.04.004

Goto, M. (1983). *Pseudomonas ficuserectae* sp. nov., the causal agent of bacterial leaf spot of *Ficus erecta* Thunb. *Int. J. Syst. Evol. Microbiol.* 33, 546–550. doi: 10.1099/00207713-33-3-546

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hwang, M. S., Morgan, R. L., Sarkar, S. F., Wang, P. W., and Guttman, D. S. (2005). Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 71, 5182–5191. doi: 10.1128/AEM.71.9.5182-5191.2005

Janse, J. D., Rossi, P., Angelucci, L., Scortichini, M., Derks, J. H. J., Akkermans, A. D. L., et al. (1996). Reclassification of *Pseudomonas syringae* pv. *avellanae* as *Pseudomonas avellanae* (spec. nov.), the bacterium causing canker of hazelnut (*Corylus avellana* L.). *Syst. Appl. Microbiol.* 19, 589–595. doi: 10.1016/S0723-2020(96)80030-0

Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231. doi: 10.1016/j.tig.2006.02.003

Kaas, R. S., Friis, C., Ussery, D. W., and Aarestrup, F. M. (2012). Estimating variation within the genes and inferring the phylogeny of

186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics*. 13:577. doi: 10.1186/1471-2164-13-577

Kałuzna, M., Willems, A., Pothier, J. F., Ruinelli, M., Sobiczewski, P., and Puławska, J. (2016). *Pseudomonas cerasi* sp. nov. (non Griffin, 1911) isolated from diseased tissue of cherry. *Syst. Appl. Microbiol.* 39, 370–377. doi: 10.1016/j.syapm.2016.05.005

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol.* 10, 504–509. doi: 10.1016/j.mib.2007.08.006

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and Archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719. doi: 10.1093/nar/gkn668

Lapage, S. P., Sneath, P. H. A., Lessel, E. F., Skerman, V. B. D., Seeliger, H. P. R., and Clark, W. A. (1975). *International Code of Nomenclature of Bacteria, 1976 Revision*. Washington, DC: American Society for Microbiology.

Lapage, S. P., Sneath, P. H. A., Lessel, E. F., Skerman, V. B. D., Seeliger, H. P. R., and Clark, W. A. (eds.). (1992). "Chapter 3, Rules of nomenclature with recommendations," in *International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision* (Washington, DC: American Society for Microbiology Press).

Lelliott, R. A., Billing, E., and Hayward, A. C. (1966). A determinative scheme for the fluorescent plant pathogenic pseudomonads. *J. Appl. Bacteriol*. 29, 470–489. doi: 10.1111/j.1365-2672.1966.tb03499.x

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Moore, E. R. B., Mau, M., Arnscheidt, A., Böttger, E. C., Hutson, R. A., Collins, M. D., et al. (1996). The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationships. *Syst. Appl. Microbiol*. 19, 478–492. doi: 10.1016/S0723-2020(96)80021-X

Morris, C. E., Barny, M. A., Berge, O., Kinkel, L. L., and Lacroix, C. (2017). Frontiers for research on the ecology of plant-pathogenic bacteria: fundamentals for sustainability. *Mol. Plant Pathol.* 18, 308–319. doi: 10.1111/mpp.12508

Morris, C. E., Sands, D. C., Vinatzer, B. A., Glaux, C., Guilbaud, C., Buffière, A., et al. (2008). The life history of the plant pathogen *Pseudomonas syringae* is linked to the water cycle. *ISME J.* 2, 321–334. doi: 10.1038/ismej.2007.113

Mulet, M., Lalucat, J., and García-Valdés, E. (2010). DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol*. 12, 1513–1530. doi: 10.1111/j.1462-2920.2010.02181.x

Mulet, M., Gomila, M., Scotta, C., Sánchez, D., Lalucat, J., and García-Valdés, E. (2012). Concordance between whole-cell matrix-assisted laser-desorption/ionization time of flight mass spectrometry and multi locus sequence analysis approaches in the species discrimination in the genus *Pseudomonas*. *System. Appl. Microbiol.* 35, 455–464. doi: 10.1016/j.syapm.2012.08.007

Ogimi, C. (1977). Studies on bacterial gall of chinaberry (*Melia Azedarach* Lin.), caused by *Pseudomonas meliae* n. sp. *Bull. Coll. Agric. Univ. Ryukyus* 24, 497–556.

Parkinson, N., Bryant, R., Bew, J., and Elphinstone, J. (2011). Rapid phylogenetic identification of members of the *Pseudomonas syringae* species complex using the *rpoD* locus. *Plant Pathol.* 60, 338–344. doi: 10.1111/j.1365-3059.2010.02366.x

Parte, A. C. (2014). LPSN-list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res*. 42, D613–D616. doi: 10.1093/nar/gkt1111

Pecnold, P. C., and Grogan, R. G. (1973). Deoxyribonucleic acid homology groups among phytopathogenic *Pseudomonus* species. *Int. J. Syst. Evol. Microbiol*. 23, 111–121.

Psallidas, P. G., and Panagopoulos, C. G. (1975). A new bacteriosis of almond caused by *Pseudomonas amygdali* sp. nov. *Ann. Instit. Phytopathol. Benaki* 11, 94–108.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A*. 106, 19126–19131. doi: 10.1073/pnas.0906412106

Robbs, C. F. (1956). Uma nova doenca bacteriana do mamoeiro (*Carica papaya* L.). *Revist. Soc. Brasil. Agron*. 12, 73–76.

Sands, D. C., Schroth, M. N., and Hildebrand, D. C. (1970). Taxonomy of phytopathogenic pseudomonads. *J. Bacteriol*. 101, 9–23.

Sarkar, S. F., and Guttman, D. S. (2004). Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl. Environ. Microbiol.* 70, 1999–2012. doi: 10.1128/AEM.70.4.1999-2012.2004

Schaad, N. W., and Cunfer, B. M. (1979). Synonymy of *Pseudomonas coronafaciens*, *Pseudomonas coronafaciens* pathovar zeae, *Pseudomonas coronafaciens* subsp. atropurpurea, and *Pseudomonas striafaciens*. *Int. J. Syst. Bacteriol*. 29, 213–221. doi: 10.1099/00207713-29-3-213

Scortichini, M., Marcelletti, S., Ferrante, P., and Firrao, G. (2013). A genomic redefinition of *Pseudomonas avellanae* species. *PLoS ONE* 25:e75794. doi: 10.1371/journal.pone.0075794

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Skerman, V. B. D., McGowan, V., and Sneath, P. H. A. (1980). Approved list of bacterial names. *Int. J. Syst. Bacteriol.* 30, 225–420. doi: 10.1099/00207713-30-1-225

Stapp, C. (1928). "Schizomycetes (Spaltpilze oder Bakterien)", in *Handbuch der Pflanzenkrankheiten, 5th Edn*, *Vol. 2*, ed P. Sorauer (Berlin: Parey), 1–295.

Thakur, S., Weir, B. S., and Guttman, D. S. (2016). Phytopathogen genome announcement: draft genome sequences of 62 *Pseudomonas syringae* type and pathotype strains. *Mol. Plant Microbe Interact.* 29, 243–246. doi: 10.1094/MPMI-01-16-0013-TA

Tran, P. N., Savka, M. A., and Gan, H. M. (2017). In-silico taxonomic classification of 373 genomes reveals species misidentification and new genospecies within the genus *Pseudomonas*. *Front. Microbiol.* 8:1296. doi: 10.3389/fmicb.2017.01296

Van Hall, C. J. J. (1902). *Bijdragen tot de Kennis der Bakterieele Plantenziekten*. Amsterdam: Coöp. Drukk.

Vinatzer, B. A., and Bull, C. T. (2009). "The impact of genomic approaches on our understanding of diversity and taxonomy of plant pathogenic bacteria," in *Plant Pathogenic Bacteria: Genomics and Molecular Biology*, ed R. W. Jackson (Caister: Academic Press), 37–61.

Vinatzer, B. A., Weisberg, A. J., Monteil, C. L., Elmarakeby, H. A., Sheppard, S. K., and Health, L. S. (2017). A proposal for a genome similarity-based taxonomy for plant-pathogenic bacteria that is sufficiently precise to reflect phylogeny, host range, and outbreak affiliation applied to *Pseudomonas syringae* sensu lato as a proof of concept. *Phytopathology* 107, 18–28. doi: 10.1094/PHYTO-07-16-0252-R

Yamamoto, S., Kasai, H., Arnold, D. L., Jackson, R. W., Vivian, A., and Harayama, S. (2000). Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* 146, 2385–2394. doi: 10.1099/00221287-146-10-2385

Young, J. M. (2008). An overview of bacterial nomenclature with special reference to plant pathogens. *Syst. Appl. Microbiol.* 31, 405–424. doi: 10.1016/j.syapm.2008.09.005

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 Gomila, Busquets, Mulet, García-Valdés and Lalucat. 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: 25 June 2018
doi: 10.3389/fmicb.2018.01360

Defining the Species *Micromonospora saelicesensis* and *Micromonospora noduli* Under the Framework of Genomics

Raúl Riesco[1], *Lorena Carro*[1], *Brenda Román-Ponce*[1], *Carlos Prieto*[2], *Jochen Blom*[3], *Hans-Peter Klenk*[4], *Philippe Normand*[5] *and Martha E. Trujillo*[1]*

[1] Departament of Microbiology and Genetics, Edificio Departamental, University of Salamanca, Salamanca, Spain, [2] Servicio de Bioinformática, NUCLEUS, Edificio I+D+i, University of Salamanca, Salamanca, Spain, [3] Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany, [4] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [5] Centre National de la Recherche Scientifique-UMR5557 Ecologie Microbienne, Université de Lyon, Université Lyon1, Villeurbanne, France

OPEN ACCESS

Edited by:
Jesus L. Romalde,
Universidade de Santiago de Compostela, Spain

Reviewed by:
Marc Gerard Chevrette,
University of Wisconsin-Madison, United States
Javier Pascual,
Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Germany

***Correspondence:**
Martha E. Trujillo
mett@usal.es

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

Received: 26 April 2018
Accepted: 05 June 2018
Published: 25 June 2018

Citation:
Riesco R, Carro L, Román-Ponce B, Prieto C, Blom J, Klenk H-P, Normand P and Trujillo ME (2018) Defining the Species Micromonospora saelicesensis and Micromonospora noduli Under the Framework of Genomics. Front. Microbiol. 9:1360. doi: 10.3389/fmicb.2018.01360

The type isolates of species *Micromonospora saelicesensis* and *Micromonospora noduli* are Gram-stain positive actinobacteria that were originally isolated from nitrogen fixing nodules of the legumes *Lupinus angustifolius* and *Pisum sativum*, respectively. These two species are very closely related and questions arise as to whether they should be merged into a single species. To better delineate the relationship of *M. saelicesensis* and *M. noduli*, 10 strains isolated from plant tissue (nodules and leaves) and identified by their 16S rRNA gene sequences as either *M. saelicensesis* or *M. noduli*, based on a cut-off value of ≥99.5% were selected for whole-genome sequencing and compared with the type strains of *M. saelicesensis* Lupac 09^T and *M. noduli* GUI43^T using overall genome relatedness indices (OGRI) which included ANI, OrthoANI and digital DNA-DNA hybridization. Whole- and core-genome phylogenomic analyses were also carried out. These results were compared with the topologies of the 16S rRNA and *gyrB* gene phylogenies. Good correlation was found between all trees except for the 16S rRNA gene. Overall results also supported the current classification of *M. saelicesensis* and *M. noduli* as separate species. Especially useful was the core-genome phylogenetic analyses based on 92 genes and the dDDH results which were highly correlated. The importance of using more than one strain for a better definition of a species was also shown. A series of *in vitro* phenotypic assays performed at different times were compared with *in silico* predictions based on genomic data. *In vitro* phenotypic tests showed discrepancies among the independent studies, confirming the lack of reproducibility even when tests were performed in the same laboratory. On the other hand, the use of *in silico* predictions proved useful for defining a stable phenotype profile among the strains analyzed. These results provide a working framework for defining *Micromonospora* species at the genomic and phenotypic level.

Keywords: *Micromonospora*, genome sequencing, phylogenomic analysis, nitrogen-fixing nodule, taxonomy, species delimitation

INTRODUCTION

To define a species, current prokaryotic taxonomy integrates multiple aspects of a microorganism that include phenotypic and genotypic data (Chun and Rainey, 2014). This approach, known as polyphasic taxonomy (Colwell, 1970; Vandamme et al., 1996) has contributed for several decades to improve classification and identification schemes, however, its limitations and pitfalls, particularly in relation to reproducibility of some methods and/or the difficulty of data storage have been timely addressed (Sutcliffe et al., 2012; Vandamme and Peeters, 2014; Thompson et al., 2015).

The introduction and improvement of cost-effective whole-genome sequencing methods provide a new working framework. Unlike DNA-DNA hybridization, that was heralded as the "golden standard" for defining genomic species in 1987 (Wayne et al., 1987), genomic data can be stored and made available to the scientific community for subsequent comparisons (Chun and Rainey, 2014). Furthermore, genomic data can also be used to predict phenotypic traits which can then be tested in the laboratory, reducing the need to perform labor-intensive and non-repoducible tests (Sutcliffe et al., 2012; Amaral et al., 2014).

Another problem is the definition of species based on single-strain representatives. This approach does not allow the recognition of intra-species diversity and limits the proposal for a sound and testable definition of a prokaryotic species. While the number of genomes representing bacterial species, in most cases, include only the type strain, it is also necessary to study several members within the same species to better understand intraspecies variation. Unfortunately, for most species, only one strain (the type strain) has been described and single-strain descriptions hinder the possibility for such studies.

The genus *Micromonospora* represented by Gram-stain positive, filamentous and sporulating actinobacteria, belongs to the family *Micromonosporaceae* of the order *Micromonosporales* in the phylum *Actinobacteria* (Genilloud, 2015a,b,c). The type species of the genus is *Micromonospora chalcea* and currently includes 81 species with validly published names http://www.bacterio.net/micromonospora.html (Parte, 2014). Most of these species have been described in the past 10 years with representative strains isolated from diverse habitats such as soil (Li and Hong, 2016; Lee and Whang, 2017), aquatic habitats (Trujillo et al., 2005; de Menezes et al., 2012), plant tissues (Carro et al., 2012, 2013; Kittiwongwattana et al., 2015; Trujillo et al., 2015; Kaewkla et al., 2017) and other environments (Hirsch et al., 2004; Nimaichand et al., 2013; Lin et al., 2015). Recently, a revised classification of the genus *Micromonospora* based on genome sequence data has been proposed (Carro et al., 2018).

In 2007, three strains recovered from internal nodular tissue of the plant *Lupinus angustifolius* were formally described as *Micromonospora saelicesensis* (Trujillo et al., 2007). Recent studies have shown that this species is widely distributed in legumes (e.g., *Trifolium*, *Lupinus*, *Pisum*, etc.), especially in nodules (Trujillo et al., 2015). *Micromonospora noduli* described as a single representative, strain GUI43^T, isolated from the nodular tissue of *Pisum sativum* was found to be closely related to *M. saelicesensis* (Carro et al., 2016). While the DNA-DNA hybridization value of 63.4% (62.3 reciprocal) is below the accepted threshold of 70% (Wayne et al., 1987) these two species share many features, and the question arises whether they should be merged into a single species. Therefore, this study was designed to determine the level of taxonomic relationship between ten strains initially identified as *M. saelicesensis* or *M. noduli* using a 16S rRNA gene sequence comparison with a similarity threshold of ≥99.5%. Draft whole-genome sequences were obtained for all strains, including the type strain *M. noduli* GUI43^T and data was analyzed using a combination of overall genome related indices (OGRI) and phylogenomic analyses. Furthermore, to obtain information about intra-species variation, especially at the phenotypic level, these studies were complemented with physiological and biochemical data. The integration of all these studies support the current status of *M. saelicesensis* and *M. noduli* as different species and the approach presented in this work provides a good method for the definition of species in the genus *Micromonospora*.

MATERIALS AND METHODS

Isolation of Strains

The list of *Micromonospora* strains used in this study is given in **Table 1**. All strains, except for LAH08, were isolated from nitrogen fixing nodules of five different legumes between 2003 and 2015 as described previously (Trujillo et al., 2010). Isolation of strain LAH08 from the leaves of *L. angustifolius* was done after surface sterilization of the plant material by immersing in 70% ethanol (v/v) for 1 min, transferred to 3.5% w/v sodium hypochlorite solution for 2 min and rinsed five times with sterile distilled water. Sample was crushed with a sterile homogenizing pestle in a microtube and the resulting slurry plated onto yeast extract-humic acid agar (de la Vega, 2010).

16S rRNA Gene Sequencing and Analysis

DNA extraction (REDExtract-N-Amp Plant PCR kit [Sigma]), 16S rRNA gene amplification and sequencing were carried out as previously explained (Trujillo et al., 2010). Assembled sequences were compared against the Ezbiocloud Database (Yoon et al., 2017) and other public platforms (Genbank, EMBL, etc.), and aligned with ClustalX2 (Thompson et al., 1997). *Catellatospora citrea* IMSNU 22008^T, a member of the family *Micromonosporaceae* was used as outgroup.

Phylogenetic analyses were performed using MEGA (v 7.0.14) (Kumar et al., 2016); distances were calculated with the Kimura 2-parameter and tree topologies were based on the Maximum Likelihood algorithm (Felsenstein, 1981). Total analysis included 1318 positions and a bootstrap (Felsenstein, 1985) sampling of 1,000.

Whole-Genome Sequencing, Assembly, and Annotation

DNA was isolated from 1 g of bacterial cultures grown in ISP 2 broth (Shirling and Gottlieb, 1966) at 28°C for 5–7 days. Cell lysis was done in 5 ml EC buffer containing 60 μl of lysozyme (300 mg/ml, Sigma-Aldrich, USA) and 50 μl of mutanolysin (1,000 U/ml), and incubated at 37°C for 90 min. Five ml of 2% SDS

TABLE 1 | Source of strains used in this study.

Strain	Host plant	Isolation	Plant collection site	Geographical coordinates	16S rRNA gene sequence identity		References
Lupac 09^T	*Lupinus angustifolius*	Nodule	Saelices	40° 40' 06" N; 6° 38' 02" W	*M. saelicesensis* Lupac09^T	100%	Trujillo et al., 2007
					M. noduli GUI43^T	99.2%	
GAR05	*Cicer arietinum*	Nodule	Cabrerizos	40° 58' 43" N; 5° 36' 46" W	*M. saelicesensis* Lupac09^T	99.9%	This study
					M. noduli GUI43^T	99.8%	
GAR06	*C. arietinum*	Nodule	Cabrerizos	40° 58' 43" N; 5° 36' 46" W	*M. saelicesensis* Lupac09^T	99.9%	This study
					M. noduli GUI43^T	99.9%	
Lupac 06	*L. angustifolius*	Nodule	Saelices	40° 40' 06" N; 6° 38' 02" W	*M. saelicesensis* Lupac09^T	99.9%	Trujillo et al., 2007
					M. noduli GUI43^T	99.7%	
PSN01	*Pisum sativum*	Nodule	Salamanca	40°58′07″ N; 5°39′49″ W	*M. saelicesensis* Lupac09^T	99.9%	This study
					M. noduli GUI43^T	99.6%	
PSN13	*P. sativum*	Nodule	Salamanca	40°58′07″ N; 5°39′49″ W	*M. saelicesensis* Lupac09^T	99.9%	This study
					M. noduli GUI43^T	99.8%	
GUI43^T	*P. sativum*	Nodule	Cañizal	41° 10' 04" N; 5° 22' 08" W	*M. noduli* GUI43^T	100%	Carro et al., 2016
					M. *saelicesensis* Lupac09^T	99.2%	
LAH08	*L. angustifolius*	Leaf	Cabrerizos	40° 58' 43" N; 5° 36' 46" W	*M. noduli* GUI43^T	99.9%	This study
					M. saelicesensis Lupac09^T	99.9%	
Lupac 07	*L. angustifolius*	Nodule	Saelices	40° 40' 06" N; 6° 38' 02" W	*M. noduli* GUI43^T	99.8%	Trujillo et al., 2007
					M. saelicesensis Lupac09^T	99.7%	
MED15	*Medicago* sp.	Nodule	Salamanca	40°58′07″ N; 5°39′49″ W	*M. noduli* GUI43^T	100%	This study
					M. saelicesensis Lupac09^T	99.2%	
ONO23	*Ononis* sp.	Nodule	Cabrerizos	40° 58' 43" N; 5° 36' 46" W	*M. noduli* GUI43^T	100%	This study
					M. saelicesensis Lupac09^T	99.8%	
ONO86	*Ononis* sp.	Nodule	Cabrerizos	40° 58' 43" N; 5° 36' 46" W	*M. noduli* GUI43^T	99.9%	This study
					M. saelicesensis Lupac09^T	99.8%	

(w/v) and 200 μl of proteinase k (10 mg/ml) were added with gentle mixing for protein precipitation and incubated at 55°C for 3 h. Samples were extracted with phenol:chloroform:isoamyl alcohol (25:24:1 v/v), treated with 35 μl of RNAse A (10 mg/ml) and precipitated with 70% ethanol. Draft genome sequences were determined by MiSeq (300 bp paired end) (Chunlab, Inc.). Libraries were prepared using TruSeq DNA LT Sample Prep kit (Illumina, San Diego, CA, USA) for the Illumina system (Chunlab, Inc.). Illumina sequencing data were assembled with SPAdes 3.10.1 (Algorithmic Biology Lab, St. Petersburg Academic University of the Russian Academy of Sciences). Protein-coding sequences (CDSs) were predicted by Prodigal 2.6.2 (Hyatt et al., 2010). Genes coding for tRNA were searched using tRNAscan-SE 1.3.1 (Schattner et al., 2005). The rRNA and other non-coding RNAs were searched by a covariance model search with Rfam 12.0 database (Nawrocki et al., 2016). All genomes were functionally annotated using the new eggNOG-mapper (Huerta-Cepas et al., 2017) with HMMER mapping mode against actNOG and bacterial HMM databases using all Orthologs. To confirm annotation, the predicted CDSs were compared with Swissprot (Bateman et al., 2015), KEGG (Kanehisa et al., 2014), and SEED (Overbeek et al., 2005) databases using UBLAST program (Edgar, 2010). Principal component analysis was carried out with the COG data using ggfortify v 0.4.3 R package (Tang et al., 2016). Clustering was inferred with K-means clustering algorithm using cluster R package v 2.0.7-1 (Maechler et al., 2018).

CRISPR elements were retrieved using the online application CRISPR-finder, available in http://crispr.i2bc.paris-saclay.fr (Grissa et al., 2007) using default parameters. EDGAR 2.0 platform (Blom et al., 2016) was used to calculate the core genome, dispensable genome and singleton genes.

gyrB Gene Phylogeny

gyrB nucleotide gene sequences extracted from whole genome sequence data or downloaded from the public databases were used to construct a Maximum-Likelihood phylogenetic tree based on Kimura 2-parameter, using 1001 nucleotide positions and a bootstrap value of 1000. *Catellatospora citrea* DSM 44097^T, a member of the family *Micromonosporaceae* was used as outgroup.

OGRI Analyses

Average Nucleotide Identity (ANI) (Goris et al., 2007) and OrthoANI (Lee et al., 2016) comparisons were made with the Orthologous Average Nucleotide Identity Tool (OAT) v0.93 https://www.ezbiocloud.net/tools/orthoani. Digital DNA-DNA hybridizations (dDDH) and G+C content differences were obtained with Genome to Genome Distance Calculator (GGDC) v2.0 available at https://ggdc.dsmz.de/ggdc.php# using the recommended settings (Meier-Kolthoff et al., 2013b). A dDDH heatmap was constructed using ComplexHeatmap R package v 1.17.1 (Gu et al., 2016).

Whole-Genome Phylogenomic Analyses

Genome Blast Distance Phylogeny (GBDP) was used to calculate the intergenomic distances based on whole proteomes (Meier-Kolthoff et al., 2013b). Calculation of a distance matrix was done using the on-line GGDC server, with BLAST+ and recommended formula 2 (optimized for draft genome sequences) (Meier-Kolthoff et al., 2013b). Phylogenetic trees were constructed with FastMe tool (Lefort et al., 2015). Genome sequence accession numbers are provided in **Table S1**.

UBCG Phylogenomic Analysis

Ninety-two bacterial core genes based on the Up to date Bacterial Core Gene (UBCG) tool, https://www.ezbiocloud.net/tools/ubcg were used for phylogenomic tree reconstruction using default parameters (Na et al., 2018). The selection of the representative genes was based on 1429 complete genome sequences, covering 28 phyla and providing a set of genes present in the majority of the genomes or highly conserved single copy genes (Na et al., 2018).

Physiology

A set of physiological and biochemical tests reported to differentiate between *M. saelicesensis* and *M. noduli* were carried out; these included carbon source utilization, determination of enzymatic activity, NaCl and pH tolerance, temperature range growth and degradation of starch, Tween 20, Tween 80, tyrosine, and urea (Carro et al., 2016). All tests were done in triplicate.

Several carbon sources (19) were tested *in vitro* at different times in the laboratory (2016 and 2017) and compared with the results of the original description of *M. saelicesensis* to check for reproducibility (Trujillo et al., 2007). Draft-genome data was screened for genes coding for proteins for carbon metabolism of the carbon sources assayed.

Biolog Characterization

To generate phenotypic fingerprints of 71 carbon source utilization and 23 chemical sensitivity assays, the strains were tested at 28°C using GEN III Microplates in an Omnilog device (BIOLOG Inc., Hayward, CA, USA). The reference strains *Micromonospora saelicesensis* Lupac 09^T and *Micromonospora noduli* GUI43^T were included for parallel comparison. One week old cells were suspended in an inoculating fluid (IF C) provided by the manufacturer and inoculated in the GEN III Microplates at a cell density of 80% transmittance. Phenotype microarray mode was used to measure respiration rates yielding a total running time of 7 days using two independent replicates for each strain. Data were recovered and analyzed using the opm package for R, v.1.0.6 (Vaas et al., 2012, 2013). Clustering analyses of the phenotypic microarrays were constructed using the pvclust package for R v.1.2.2 (Suzuki and Shimodaira, 2015). Distinct behaviors between the two repetitions in the reactions were regarded as ambiguous.

RESULTS

16S rRNA Gene Sequence Analysis

The 16S rRNA gene sequences were used to determine the nearest phylogenetic neighbors based on overall sequence similarity in relation to currently described *Micromonospora* species. In all cases, the closest species were *M. saelicesensis* and *M. noduli* with similarity values of 99.2–100% (**Table 1**).

A phylogenetic tree constructed with the new sequences and those of 81 *Micromonospora* species described to date, distributed the 10 strains into two clusters: Group I contained the type strain *M. saelicesensis* Lupac 09^T and the isolates PSN13, GAR06, PSN01, Lupac 06, GAR05, and Lupac 07. Group II was formed with ONO 86, ONO23, LAH08, and MED15 and *M. noduli* GUI43^T (**Figure 1**). The topology clearly showed the close relationship between the two groups as visualized by the branch lengths which were almost inexistent. Group II (*M. noduli*) also showed a close relationship with the type strains of *Micromonospora profundi* isolated from a deep marine sediment (Veyisoglu et al., 2016) and *Micromonospora ureilytica* isolated from *Pisum sativum* (Carro et al., 2016), also recovered in this group. Reported DDH values between *M. saelicesensis* and *M. ureilytica* and *M. profundi* (Veyisoglu et al., 2016) were 28.4 and 56.9%, respectively. A DDH value of 50.9% was found between *M. noduli* GUI 43^T and *M. ureilytica* GUI23^T (Carro et al., 2016).

gyrB Gene Phylogeny

The phylogenetic tree constructed with the *gyrB* gene sequences showed a similar topology to the 16S rRNA gene tree with respect to the study strains (**Figure 2**). These were recovered in their respective groups defined in the 16S rRNA gene tree. The exception was strain Lupac 07, recovered in the *M. noduli* cluster (Group II) and this rearrangement was supported by a bootstrap value of 99%. This strain was originally classified as *M. saelicesensis* (Trujillo et al., 2007). The positions of *M. profundi* DS 3010^T and *M. ureilytica* GUI23^T also changed and moved out of the *M. noduli* cluster. As previously noted (Garcia et al., 2010; Carro et al., 2012), the *gyrB* gene phylogeny yielded a better resolution as observed by slightly larger distance branches, however, the topology of the remaining type strains was very different from that obtained using the 16S rRNA gene. This phylogeny was very similar to a tree constructed using a concatenated set of five housekeeping genes as proposed previously (Carro et al., 2012) (data not shown).

Comparative Genomic Characteristics

Eleven high quality draft genomes (depth >100X), including that of the type strain *M. noduli* GUI43^T were obtained. General genome characteristics of the sequenced strains are provided in **Table 2**. Genome sizes varied from 6.8 to 7.4 Mb, the largest genome being that of strain PSN13. The G+C mol% among all strains was very homogeneous. The values of the *M. saelicesensis* group ranged from 71.1 to 71.2% while the strains in the *M. noduli* cluster, including Lupac 07, varied from 70.9 to 71.1%. As observed, the G+C mol% values between the two species was less than 1%.

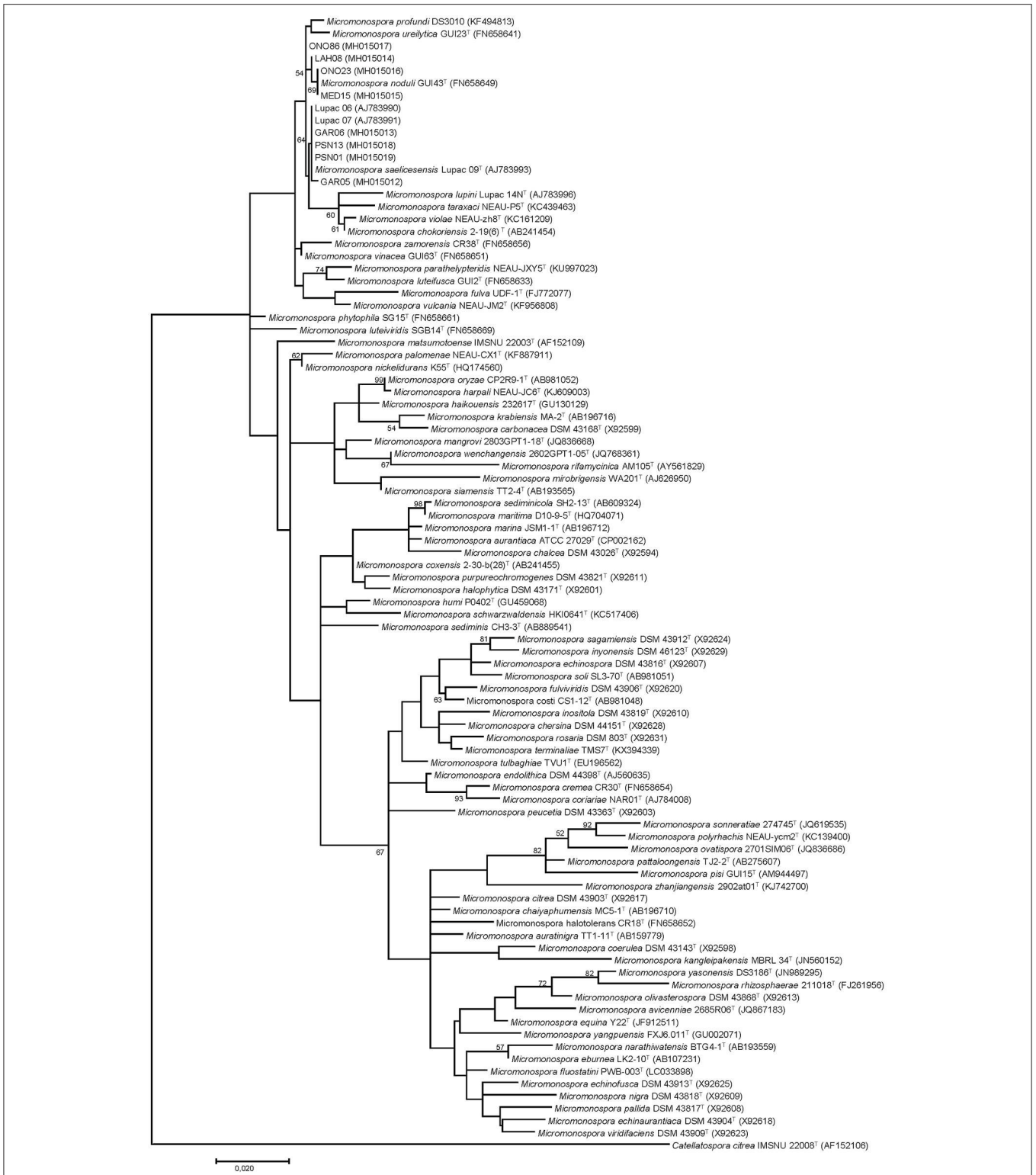


FIGURE 1 | Maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences showing the relationships between 81 *Micromonospora* type and the study strains. Distances were calculated with the Kimura 2-parameter. The tree is based on 1,318 nt. Bootstrap percentages ≥50% (1,000 samplings) are shown at nodes. Bar, 0.02 substitutions per nucleotide.

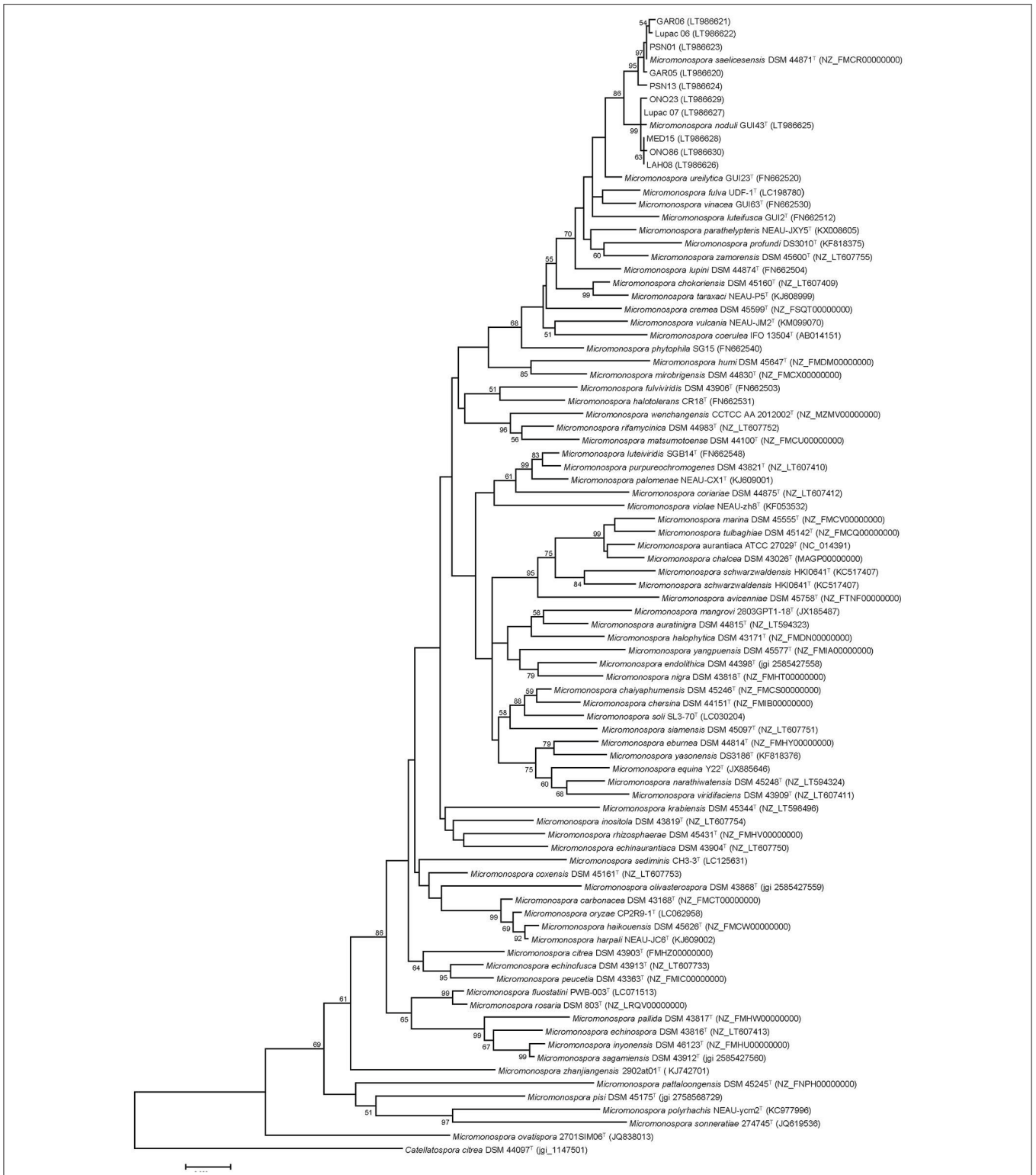


FIGURE 2 | Maximum-likelihood phylogenetic tree based on *gyr*B gene sequences. A total of 76 *Micromonospora* type strains and 10 non-type strains have been used for the analysis. Distances were calculated with the Kimura 2-parameter, using 999 nucleotide positions and a bootstrap of 1,000. Bar, 0.05 substitutions per nucleotide.

TABLE 2 | General genomic characteristics of M. *saelicesensis* and *M. noduli* strains.

Strain	Genome size (Mb)	G+C ratio (mol%)	Number of Contigs	CDS	rRNA operon	tRNA	Genes in COGs (%)*	Genes in COGs (%)**	Confirmed CRISPR count
M. saelicesensis DSM 44871^T	7.1	71.1	11	6568	9	67	87.13	86.13	4
GAR 05	7.0	71.2	75	6526	6	68	87.07	86.07	0
GAR 06	6.9	71.2	60	6410	8	67	89.58	88.63	4
Lupac 06	7.0	71.2	59	6495	4	65	87.62	86.61	1
PSN 01	6.8	71.1	456	6458	4	67	85.71	84.75	4
PSN 13	7.4	71.1	41	6823	3	64	86.91	85.90	1
M. noduli GUI43^T	7.2	70.9	224	6539	3	57	87.81	86.88	0
LAH 08	7.2	71.1	62	6627	4	56	88.13	87.14	3
Lupac 07	7.1	71.1	56	6500	4	51	86.95	86.06	2
MED 15	7.2	71.1	43	6554	4	56	87.97	87.00	1
ONO 23	7.2	71.0	135	6565	5	56	87.39	86.43	2
ONO 86	7.1	70.9	407	6591	4	55	85.48	84.60	2

mapped against **bact HMM database included in eggNOG 4.5.1.*
***mapped against **actNOG** HMM database included in eggNOG 4.5.1.*

The number of coding DNA sequences (CDS) was also very similar between both species representatives, strain GAR06 showed the lowest number with 6410 and strain PSN13 had the highest count with 6823, a difference of 413 CDS. A slightly larger difference was observed between the number of CDS in the group of *M. saelicesensis* (410) with respect to the *M. noduli* group (127). The number of rRNA operons was also higher in *M. saelicesensis*, with the type strain Lupac 09^T accounting for the highest number (9 operons), followed by GAR06 (8 operons) and GAR05 (6 operons). The number of rRNA operons in the *M. noduli* was lower (3–5 operons) with the type strain, GUI43^T having only 3 and strain ONO23 accounting for 5. In the case of tRNAs, it was observed that the number of these molecules was higher in all strains identified as *M. saelicesensis* (64–68) than in the *M. noduli* strains (51–57). A high number of tRNAs (77) was also reported for *Micromonospora lupini* Lupac 08^T also isolated from nitrogen fixing nodules (Trujillo et al., 2014), while tRNAs reported for available *Micromonospora* genomes ranged from 48 to 87 (Carro et al., 2018).

The core genome of the six strains identified as *M. saelicesensis* (Group I) was calculated to be 5313 genes (81.3%) considering an average genome of 6531 genes. The number of singletons ranged from 94 for GAR06 to 706 for PSN13. In the case of the *M. noduli* strains (group II), the core genome included 5759 genes (88.05%) for an average genome of 6540 genes. In this group, strain Lupac 07 had the lowest number of singletons, 84, while strain ONO86 showed the largest variation with 369 genes. The calculation of a core genome based on all strains dropped to 74.72% and contained 4884 genes (**Table S2**). The calculated pangenomes were 8405, 7857, and 9867 genes for *M. saelicesensis* (Group I), *M. noduli* (Group II) and the combination of both species, respectively (**Figure S1**). As expected, an increase in the number of genes in the global pangenome was observed when all strains were combined, suggesting an important degree of variation between the genomes. The progression of the pan- and core genome can be seen in **Figure S2**.

Over 85% of the CDS for each species group were classified into Clusters of Orthologous Groups (COGs). COG profiles were very similar in all strains and were assigned into 22 categories being K (transcription, 8.6–8.9%), G (carbohydrate metabolism, 6.3–6.7%) and E (amino acid and transport metabolism, 4.7–5.0%), the most abundant. This COG distribution was very similar to the COG profile of *M. lupini* Lupac 08 (Carro et al., 2018) a close phylogenetic neighbor of *M. saelicesensis* and *M. noduli*. Principal component analysis of the COG distribution is represented in **Figure 3** where both species groups are clearly separated, but with strain PNS13 recovered as an outlier. The categories K and G accounted for 65.38 and 15.27% of the variance, respectively.

OGRI Indices

Overall genomic relatedness indices (Chun and Rainey, 2014) were used to determine the relatedness between each pair of genomes used in the present study. In *M. saelicesensis* (Group I) ANI and OrthoANI values ranged from 97.82 to 99.13% and 97.96 to 99.19%, respectively, between the type and study strains. The *M. noduli* group (Group II) had ANI and OrthoANI values from 99.05 to 99.09% and 99.12 to 99.14% respectively (**Table 3**). In both cases, these values were above the recommended cut-off value of ∼96% for species recognition (Richter and Rosselló-Móra, 2009).

The ANI and OrthoANI values between the type strains *M. saelicesensis* and *M. noduli* were 96.64 and 96.82 respectively. Overall, pairwise comparison between the two groups showed the highest ANI and OrthoANI values corresponded to strains GAR05 and GUI43^T (96.68%, ANI) and PSN01 and ONO23 (96.90%, OrthoANI) (**Table S3**). Both results are slightly above the border line of 95–96% for the delineation of species. However, these results are comparable to OrthoANI values obtained for the genome pairs of *M. carbonacea* and *M. haikouensis* (95.16%), and *M. inyonensis* and *M. sagamiensis* (96.5%).

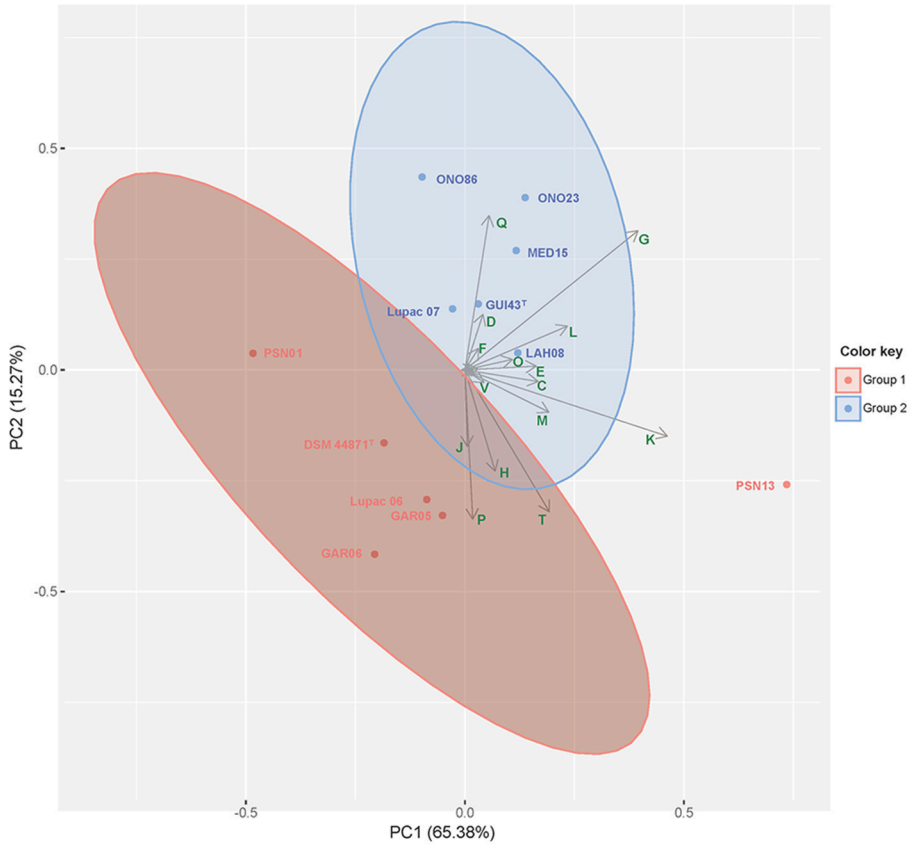


FIGURE 3 | Principal Component Analysis (PCA) of COGs as distributed in the genomes of the strains representing *M. saelicesensis* (Group I) and *M. noduli* (Group II). Clustering was inferred using the K-means clustering algorithm. Arrows represent the contribution of functional COG categories. COGs represented (C) Energy production and conversion, (D) Cell cycle control, cell division and chromosome partitioning, (E) Amino acid transport and metabolism, (F) Nucleotide transport and metabolism, (G) Carbohydrate transport and metabolism, (H) Coenzyme transport and metabolism, (J) Translation, ribosomal structure and biogenesis, (K) Transcription, (L) Replication, recombination and repair, (M) Cell wall/membrane/envelope biogenesis, (O) Post-translational modification, protein turnover and chaperones, (P) Inorganic ion transport and metabolism, (Q) Secondary metabolites biosynthesis, transport and catabolism, (T) Signal transduction mechanisms, (V) Defense mechanisms. The first two principal components accounting for 80.65% of the total variance are presented in the plot.

Species delineation based on dDDH values ranged from 81.0 to 93.7% for the six strains in *M. saelicesensis* (Group I) and 92.3–93.8% for the 5 components of *M. noduli* (Group II); all values clearly above the 70% recommended threshold (**Table S3**). dDDH values between the two species groups ranged from 71.0 to 71.8% (**Table 3**). Similar to ANI and OrthoANI results, dDDH values between the two groups were slightly above the border limit threshold value of 70% (68.1–74.5%). Overall pairwise comparisons of the study genomes and 48 additional *Micromonospora* type strains show the close relationship of the strains but clearly delineate each group within this 70–71% dDDH radius (**Figure 4**). A similar situation is observed between the species *Micromonospora sagamiensis* and *Micromonospora inyonensis* which share a dDDH value close to 70% (69.8%, dDDH; 61.3%, experimental DDH) (Kroppenstedt et al., 2005). Meier-Kolthoff and colleagues (Meier-Kolthoff et al., 2014), recently proposed the delineation of subspecies using genomic data. Specifically, these authors recommend a threshold of 79–80% to define subspecies in prokaryotic taxonomy. In the present study, the values obtained for the *Micromonospora* strains are much lower than this range and these strains are better classified as different species rather than subspecies.

Whole-Genome Phylogenomic Analysis

Phylogenomic tree reconstruction based on whole-genome distances calculated with the GBDP tool is presented in **Figure 5**. This tree included the 10 study genomes, all *Micromonospora* strains (type and non-type) published previously (Carro et al., 2018) and the genome sequences of the type strains *M. noduli* GUI43^T (this work), *M. avicinniae* DSM 45748^T, *M. pisi* DSM 45175^T, *M. pattaloongensis* DSM 45245^T, *M. rosaria* DSM 803^T and *M. wenchangensis* CCTCC AA 2012002^T. The composition of the *M. saelicesensis* and *M. noduli* groups defined in the *gyrB* gene tree were identical, including the position of Lupac 07 as a member of *M. noduli* (Group II). The overall topology of this tree and the one published by Carro et al. (2018) was very similar, however, the inclusion of 17 additional genomes, as expected, influenced the distribution of the type strains, especially the

TABLE 3 | OGRI indices: ANI, OrthoANI, and dDDH percentage values calculated between the type and study strains.

Strain	ANI ***M. saelicesensis* DSM 44871^T/*M. noduli* GUI43^T**	OrthoANI ***M. saelicesensis* DSM 44871^T/*M. noduli* GUI43^T**	dDDH ***M. saelicesensis* DSM 44871^T/*M. noduli* GUI43^T**
M. saelicesensis DSM 44871^T	100/96.64	100/96.82	100/71.20
GAR 05	99.11/96.68	99.16/96.82	92.50/71.50
GAR 06	99.09/96.61	99.18/96.79	92.90/71.30
Lupac 06	99.13/96.63	99.19/96.82	92.70/71.20
PSN 01	99.12/96.61	99.19/96.82	92.80/71.60
PSN 13	97.82/96.55	97.96/96.79	81.30/71.30
	***M. noduli* GUI43^T/*M. saelicesensis DSM* 44871^T**	***M. noduli* GUI43^T/*M. saelicesensis DSM* 44871^T**	***M. noduli* GUI43^T/*M. saelicesensis DSM* 44871^T**
M. noduli GUI43^T	100/96.64	100/96.82	100/71.20
LAH 08	99.09/96.23	99.15/96.83	92.50/71.30
Lupac 07	99.09/96.59	99.14/96.80	92.70/71.30
MED 15	99.09/96.59	99.13/96.78	92.60/71.10
ONO 23	99.05/96.62	99.12/96.87	92.30/71.30
ONO 86	99.05/96.59	99.14/96.23	92.80/71.50

inclusion of the six additional type strains. Nevertheless, three out of Carro's five defined groups (I, IV, and V) were almost completely recovered in the present phylogenomic analysis, the major rearrangements were observed in Carro's groups II and III. In the present phylogenomic analysis, the strains in group II (*M. purpureochromogenes*, *M. coxensis* and *M. halophytica*) fused with *M. rifamycinica* and *M. matsumotoense* (group III) and were joined by *M. wenchangensis* (new to the analysis). The instability of group III was already highlighted (Carro et al., 2018). This rearrangement reduced group III to *M. olivasterospora*, *M. carbonacea,* and *M. haikouensis*.

UBCG Phylogenomic Analysis

The same dataset as above was used to construct a phylogenetic tree based on a core genome set of 92 genes using the UBCG tool (Na et al., 2018). Most of the selected genes (67/92) fall in the translation COG category (J), coding for ribosomal proteins (25/92, 50S and 18/92, 30S), aminoacid-tRNA ligases (10/92) and elongation an initiator factors (4/92) (**Table S4**). Again, the ten strains were distributed in two groups of identical composition as that of the *gyrB* gene and whole-genome phylogenomic analyses with significant branch support as indicated by the bootstrap values and gene support indices (GSI) (**Figure 6**). GSI values indicate the reliability of the branches on the phylogenomic tree based on the total number of genes used to construct the tree (92 genes) (Na et al., 2018). The topology of this tree with respect to the composition of the two groups was the same as the whole-genome and *gyrB* gene trees, including the position of strain Lupac 07, recovered in the *M. noduli* group. The topology of the UBCG tree highly correlated to the topology of the whole-genome phylogenomic tree of this study. Especially interesting was the fact that the new redefined groups II and III were recovered in their entirety together with groups I, IV and V. In this analysis, a new group that contained strains from Carro's groups I (*M. mirobrigensis* and *M. siamensis*), III (*M. yangpuensis*) and IV (*M. krabiensis*), in addition to the newly included type strains *M. avicinniae* and *M. rosaria* was formed (**Figure 6**). Another important difference between the whole- and UBCG trees of this study was the position of *Salinispora pacifica* and *Salinispora arenicola* which in the latter tree was found associated to group IV. In this case, the up-to-date bacterial core gene analysis was not resolutive.

Phenotypic Profiles

Thirty-one phenotypic tests reported previously to be useful for the differentiation of the species *M. saelicesensis* and *M. noduli* (Carro et al., 2016) were carried out with all test strains. The number of characteristics that phenotypically differentiated between the two species was significantly reduced to one test when the number of strains compared increased (**Table S5**). Specifically, the use of rhamnose as a carbon source substrate was positive for all strains in the *M. noduli* group while the results were all negative for *M. saelicesensis* strains except for isolate PSN13 which was positive. The results of the remaining tests varied at the strain level and did not relate to their species identification.

Intra-species variability within each species group ranged from 0 to 33.3%. Range of pH growth and lipase production were the most variable tests in *M. saelicesensis*; utilization of serine as carbon source, degradation of tyrosine and pH growth range were the most variable tests for the *M. noduli* group.

Phenotypic profiles using the Biolog system were also determined for all strains. In this case, none of the 71 carbon sources or the 23 biochemical tests served to differentiate between the two species, given the variability observed among the duplicate tests (**Table S6**). Strain Lupac 06 was the most variable with 35.1% discrepancies recorded. Overall intraspecies variability for *M. saelicesensis* and *M noduli* was 25.5 and 26.6% respectively.

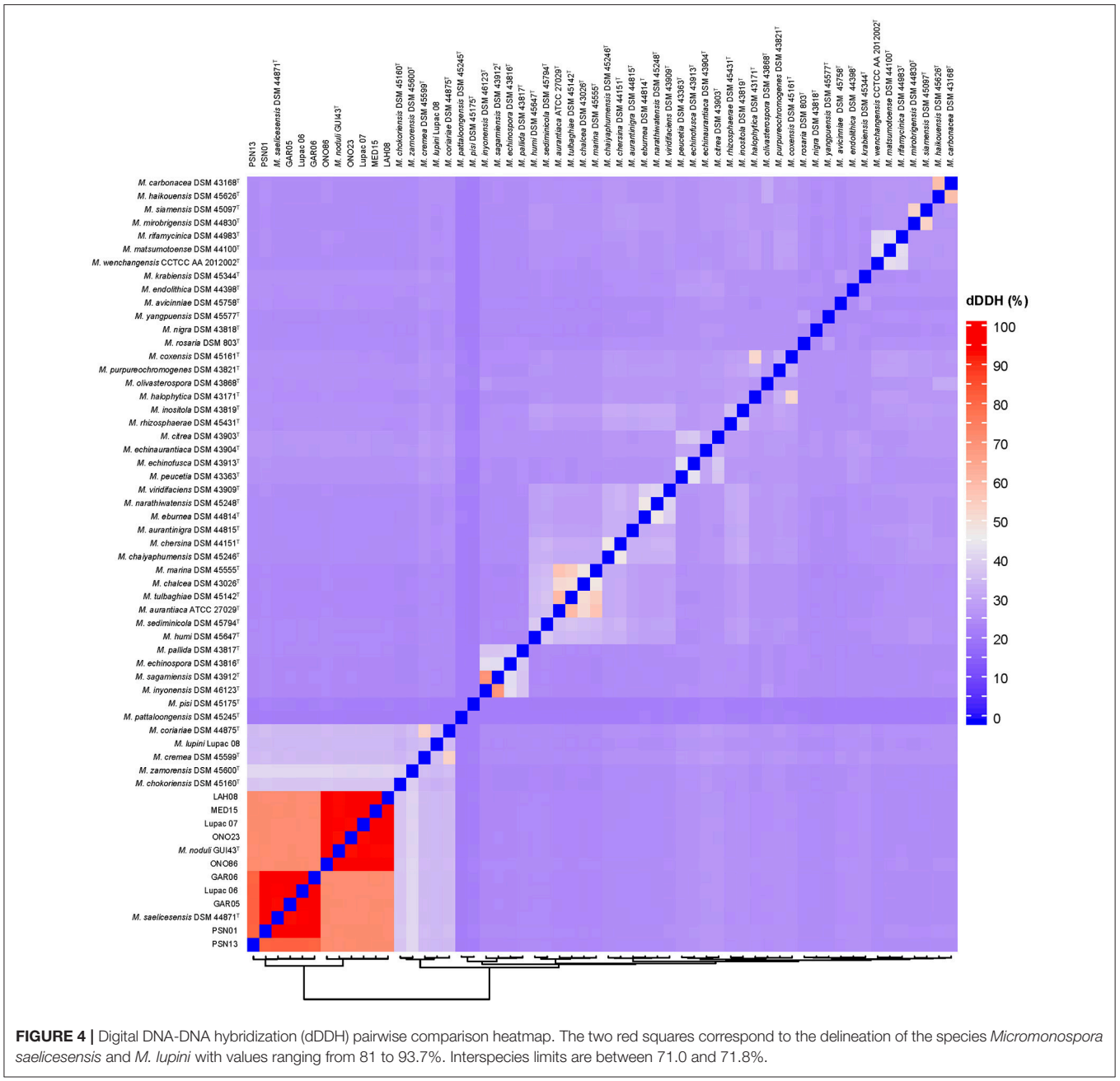


FIGURE 4 | Digital DNA-DNA hybridization (dDDH) pairwise comparison heatmap. The two red squares correspond to the delineation of the species *Micromonospora saelicesensis* and *M. lupini* with values ranging from 81 to 93.7%. Interspecies limits are between 71.0 and 71.8%.

Nineteen carbon sources were also assayed at different times (2007, 2016, and 2017) to check for reproducibility. Nine of the eleven strains tested expressed discrepant results over the different testing times. Three strains (Lupac 09^T, Lupac 06, and Lupac 07) showed the highest variation with 26% of the tests yielding conflicting results while MED15, LAH08, and PSN13 had the lowest variation (5.2%). The use of D-serine as carbon source was the least reproducible test with seven strains yielding conflicting results (**Table S7**).

Draft genomes of the test strains were screened for genes involved in the carbon metabolism of the corresponding 19 substrates assayed *in vitro*. The predicted phenotypes correlated 100% with the results obtained in the laboratory for 11 tests. However, in the case of L-alanine, L-arginine, L-histidine, L-lysine, myo-inositol, L-rhamnose, D-serine, and D-trehalose, discrepant results were found between wet lab and *in silico* predictions (**Figure S3**). In most cases, the genes were localized in the genome but the experimental results varied (+/–) suggesting that even when the tests were carried out in the same laboratory and using the same method, they were not 100% reproducible.

In the case of L-rhamnose, *in vitro* tests for strain GUI43^T were positive but the genes related to the metabolism of this compound were not located. This is probably explained by the

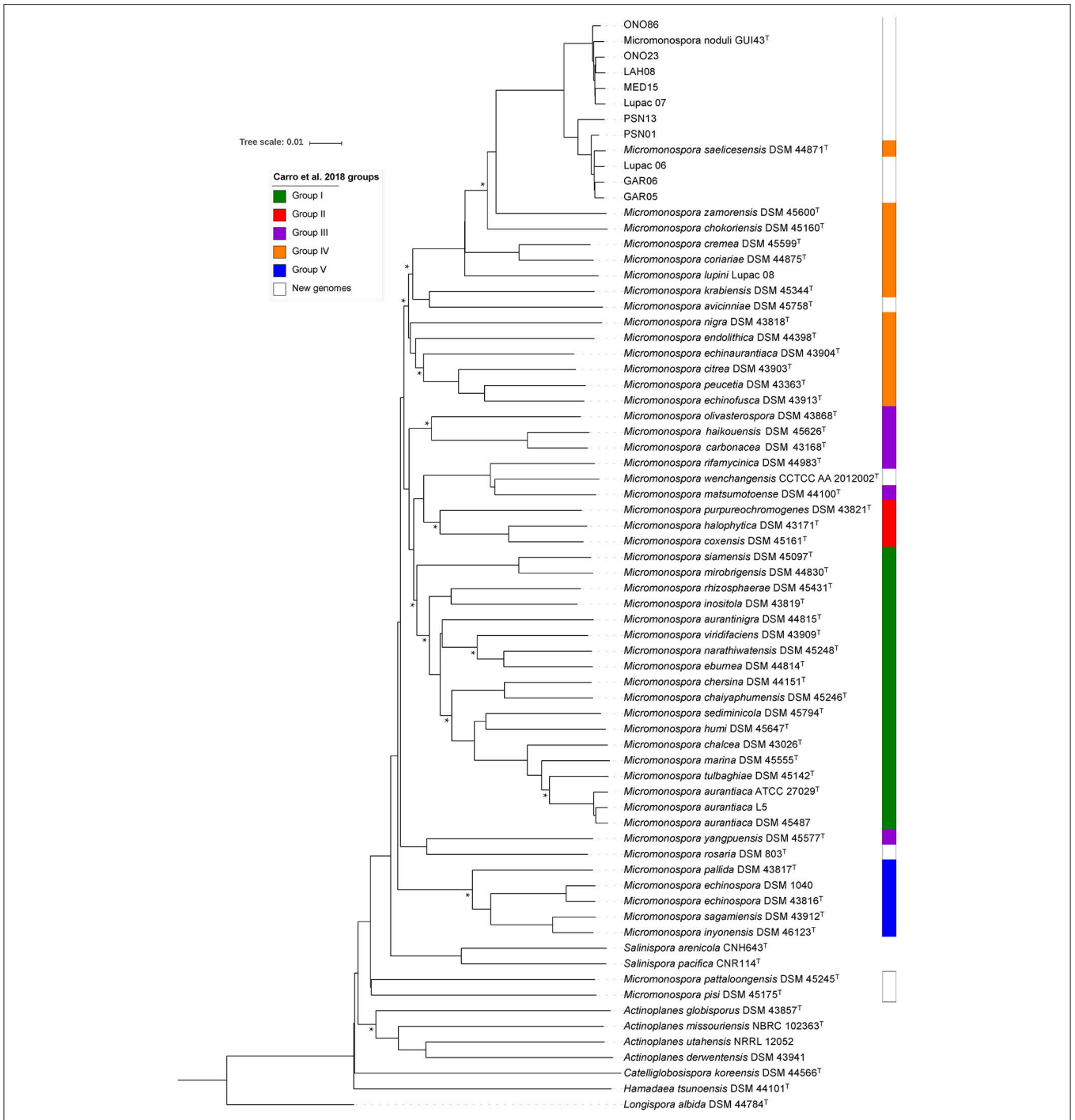


FIGURE 5 | Whole genome-sequence based phylogenomic tree constructed with the GBDP tool (see main text for details). Colors on the right side represent groups described in Carro et al. (2018). Asterisks represent conserved nodes between this tree and the core genome phylogenetic tree.

fact that draft-genomes were used and interpretation of genomic data should be done with precaution.

DISCUSSION

The genus *Micromonospora* is highly relevant in biotechnological applications in areas such as medicine, agriculture and biofuels (Hirsch and Valdés, 2010; Trujillo et al., 2015; Carro et al., 2018). At present, this taxon holds 81 species with validly published names (LPSN), most of them described based on a polyphasic approach (Colwell, 1970; Vandamme et al., 1996). Within this framework, DNA-DNA hybridization (DDH) has been considered the key test to decide if a new strain represents a new species, despite its well spelled limitations (Gevers et al.,

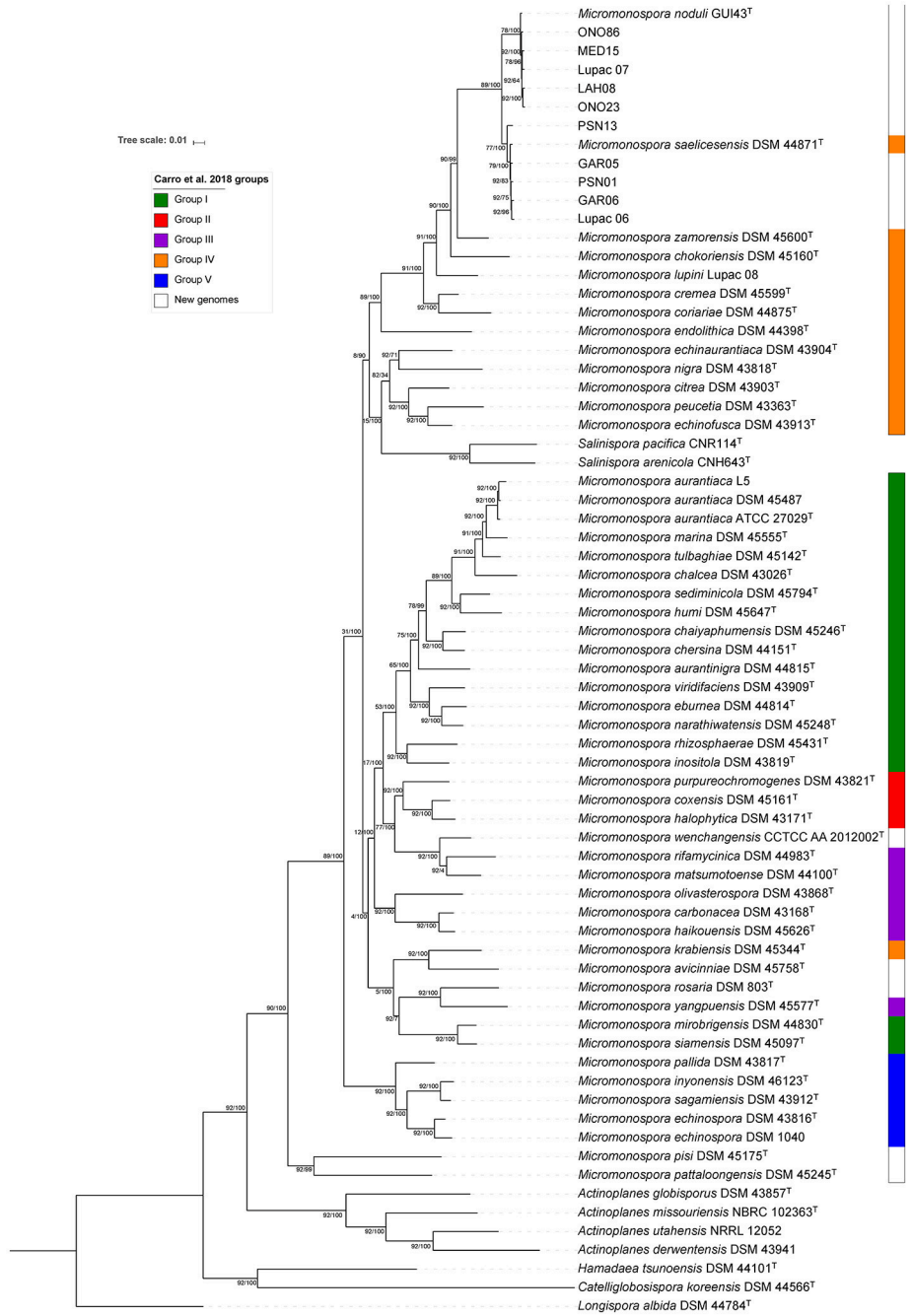


FIGURE 6 | Up-to-date bacterial core gene phylogeneomic tree reconstructed with 92 bacterial core genes. Tree has been formatted using itol platform. Colors on the right represent groups described in Carro et al. (2018). GSI support (left) and bootstrap values(right) are given at nodes.

2005; Meier-Kolthoff et al., 2013a). Given the drawbacks of DDH, it is not always straight forward to delineate the species limits, especially when DDH values close to the threshold. Therefore, the development of whole genome sequencing seems more appropriate to deduce relatedness by comparing genome sequences rather than performing DDH experiments (Vandamme and Peeters, 2014). Genomic data was recently used as the backbone to revisit the classification of the genus *Micromonospora* using a set of 45 draft genomes providing a useful dataset for comparison (Carro et al., 2018).

While 16S rRNA is limited in resolving phylogenetic relationships at the species level (Katayama et al., 2007; Hahnke et al., 2016; Carro et al., 2018; Na et al., 2018), it has provided a good starting point for taxonomic studies. In this work,

16S rRNA gene sequencing was used to identify the closest neighbors of ten *Micromonospora* strains isolated from various legumes (**Table 1**). The sequence similarity values indicated that *M. saelicesensis* or *M. noduli* were the two most closely related species although in some cases, similarity values were identical between the test and both type strains (e.g., GAR06 and LAH08). The 16S rRNA gene tree topology yielded two very tight groups which could be interpreted as a single one when the branch lengths from these clusters were compared against the lengths of the remaining 79 *Micromonospora* type strains included in the analysis.

The use of *gyrB* gene sequences to resolve phylogenetic relationships in the genus *Micromonospora* has been recommended by several authors (Kasai et al., 2000; Garcia et al., 2010; Carro et al., 2012) given its higher resolution when compared to 16S rRNA gene phylogeny. In this study, the *gyrB* gene tree topology showed a similar arrangement to the 16S rRNA gene tree with respect to the test strains, however several differences were observed. The branch lengths were slightly longer, but still very small when compared to the rest of the *Micromonospora* species included in the tree. The most relevant change was the position of strain Lupac 07, which, together with strains Lupac 06 and Lupac 09^T were originally classified as *M. saelicesensis* (Trujillo et al., 2007). The latter strains remained in the *M. saelicesensis* cluster but Lupac 07 moved to the *M. noduli* group. As expected, topologies of both trees in relation to the type strains were very different confirming that phylogenies based on single genes are very limited and unstable, making identification of nearest phylogenetic neighbors difficult.

The tree topologies based on the phylogenomic analyses of the UBCG (92 genes) and the whole draft genomes were similar. In both trees, strain Lupac 07 was recovered in the *M. noduli* group, strongly suggesting that this strain should be reclassified as a member of this species. The remaining 9 strains were recovered in the same species groups throughout all analyses.

In this study, both phylogenomic analyses contained a total of 70 genomes, including six additional *Micromonospora* type strains (see above). Overall, good agreement was found between the two phylogenies of this work and recently published data. In all cases, groups I, IV, and V previously defined (Carro et al., 2018) were recovered in their entirety with *M. avinniceae* (this analysis) joining group IV. The main difference between the three phylogenies was the composition of Carro's groups II and III which were clearly influenced by the addition of *M. rosaria* DSM 803^T and *M. wenchangensis* CCTCC AA 2012002^T, producing a new group recovered in both phylogenies of the present work. Nevertheless, the groups I, IV, and V remained very stable considering that 11 new genomes (*M. noduli* GUI43^T and 10 test strains) were added and these were assigned to group IV where *M. saelicesensis* Lupac 09 was originally assigned. These rearrangements reinforce the argument that classification and identification systems are data dependent and constant rearrangement should be expected as more data are added and alternative methods are applied (Carro et al., 2018).

The new analysis tool UBCG proved useful for the construction of phylogenomic analysis, showing good correlation with trees using whole-draft genome data even though it did not resolve well the position of the *Salinispora* representatives, however, this may be due to the small number of representatives in the data set. An advantage of this pipeline is the use of bootstrap and GSI values to support the phylogenetic branches. It is also expected that as more genome sequences are added to the database, the more resolutive it should become.

Genome relatedness indices (ANI, Ortho-ANI, and dDDH) were calculated to complement the phylogenomic analyses for species demarcation. Overall, the three methods showed good agreement and the two species groups defined in the *gyrB*, core-genome and whole-genome phylogenetic analyses supported the recognition of the 10 strains in two species.

Furthermore, these studies served to highlight the close relationship between the species *M. saelicesensis* and *M. noduli*. ANI values proposed for species delineation have been set to 95–96% as this range has been found to be correlated with the experimental DDH threshold of 70% (Goris et al., 2007; Richter and Rosselló-Móra, 2009). An alternative means to measure relatedness between two genomes is the calculation of dDDH using the GBDP method which appears to show a better correlation than ANI to the data derived from DDH experiments (Auch et al., 2010; Meier-Kolthoff et al., 2013b; Peeters et al., 2016).

In this work, the OGRI values were slightly above the recommended threshold for species delineation, if strictly applied, the study strains should be recognized as members of the same species. However, the consideration of other results in this work support the recognition of the strains as two separate species, *M. saelicesensis* and *M. noduli*. As previously expressed, thresholds are necessary for guidance but these should be applied in a flexible manner and considering other biological properties (Li et al., 2015). The present work is a good example for the interpretation and application of these values.

The use of phenotypic traits to identify and differentiate species in prokaryotic systematics is of limited value as previously discussed (Sutcliffe et al., 2013; Amaral et al., 2014; Vandamme and Peeters, 2014). In this work, several strains identified as one species, expressed different phenotypes, highlighting the problem of using diagnostic tables based on single strains to list differential characteristics between species. Information about intra-species variation is crucial for the development of stable diagnostic characteristics and the convenience of using more than a single isolate have been previously discussed (Sutcliffe et al., 2012; Oren and Garrity, 2014).

Our results confirm that the use of phenotypic tests, even when performed under the same conditions are not reliable for species differentiation due to the high variability observed within several members of the same species (Kumar et al., 2015). Instead, phenotypic studies should be regarded as complementary information to understand the biology of a microorganism and they should be restricted to strain characterization. Understandably, the inclusion of additional strains for the description of a taxon is often regarded as a burden because a lot of extra work is needed, especially when looking for differential phenotypic tests with questionable taxonomic value (Sutcliffe et al., 2012; Vandamme and Peeters, 2014).

Genomic information can be used to determine the intrinsic variability between a set of strains based on the core and pangenome profiles (Coenye et al., 2005; Sutcliffe et al., 2012; Oren and Garrity, 2014). In this work the calculation of these parameters has pointed out an important degree of variation between the species *M. saelicesensis* and *M. noduli* supporting their recognition as separate taxa. The complete elucidation of the gene functions within each group may provide an initial set of stable differential characteristics for each species, some of which may be phenotypically expressed.

CONCLUDING REMARKS

As additional data is generated, genome-based classifications should become more stable and provide a new working frame for the systematics of prokaryotes. The present study illustrates the advantage of using a diverse array of methods for the correct identification of new strains and the importance of using more than one isolate for a better characterization and definition of a species. OGRI values and especially dDDH values seem very appropriate for the delineation of prokaryotic species, but threshold numbers should be applied with a sufficient level of flexibility and considering other features inherent to a microorganism such as ecology, physiology, etc. (Li et al., 2015). There is no doubt that phenotypic information is useful for the good characterization of strains, but these studies should aim to provide information on the biology of a microorganism and not necessarily and not only to fill out a table with results of questionable value.

AUTHOR CONTRIBUTIONS

RR, LC and BR-P carried out experiments and bioinformatic analyses. CP and JB performed bioinformatic analyses. H-PK, PN, and MT designed the study and wrote the manuscript. All authors read the manuscript.

FUNDING

MT received financial support from the Ministerio de Economía y Competitivad (MICINN) under project CGL2014-52735-P. RR received a PhD scholarship from the University of Salamanca. BR-P and LC received a postdoctoral fellowship from CONACYT, México and University of Salamanca, respectively.

ACKNOWLEDGMENTS

Markus Göker is acknowledged for providing the iTOL annotation script.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01360/full#supplementary-material

Figure S1 | (A) Venn diagram showing the number of orthologous genes clusters that integrate the core and disposable genomes, and singletons of all strains in Group I (*Micromonospora saelicesensis*). **(B)** Venn diagram showing the number of orthologous gene clusters that integrate the core genome and disposable genomes, and singletons of all strains in Group II (*Micromonospora noduli*).

Figure S2 | Pan- and Core genome development plot of *Micromonospora noduli* and *Micromonospora saelicesensis* strains. The orange and blue lines show the progression in the pan- and core genomes as more genomes are added.

Figure S3 | Predicted phenotypes vs. experimental phenotypic data based on 19 carbon source substrates. *In silico* prediction negative, phenotype not expressed (purple); *in silico* prediction negative (genes not found), phenotype expressed (red); *in silico* prediction positive, phenotype not expressed (light green) and *in silico* prediction positive, phenotype expressed (green).

Table S1 | Genome sequence accession numbers of strains used in this work.

Table S2 | Number of orthologous genes that conform the pan genome, core genome and singletons of *Micromonospora saelicesensis* (Group 1) and *Micromonospora noduli* (Group II). In parenthesis, values expressed as percentages based on an average genome of 6531 genes for *M. saelicesensis* and 6540 genes for *M. noduli.*

Table S3 | Pair-wise OGRI values for ANI, OrthoANI and dDDH between *M. saelicesensis* (Group I) and *M. noduli* (Group II) strains.

Table S4 | List of genes used in core-genome phylogenomic analysis based on UBCG (Na et al., 2018).

Table S5 | Differential phenotypic characteristics between *M. saelicesensis* and *M. noduli* as reported by Carro and colleagues (Carro et al., 2016). +, Positive; –, Negative; w, Weak.

Table S6 | BIOLOG phenotypic profiles of *M. saelicesensis* and *M. noduli* strains. +, positive; –, negative; c, conflicting.

Table S7 | Carbon source substrates tested at different times using the same laboratory conditions. +, positive; –, negative; w, weak.

REFERENCES

Amaral, G. R. S., Dias, G. M., Wellington-Oguri, M., Chimetto, L., Campeão, M. E., Thompson, F. L., et al. (2014). Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. *Int. J. Syst. Evol. Microbiol.* 64, 357–365. doi: 10.1099/ijs.0.057927-0

Auch, A. F., von Jan, M., Klenk, H. P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., et al. (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44, W22–W28. doi: 10.1093/nar/gkw255

Carro, L., Nouioui, I., Sangal, V., Meier-Kolthoff, J. P., Trujillo, M. E., Montero-Calasanz, M. D. C., et al. (2018). Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential. *Sci. Rep.* 8:525. doi: 10.1038/s41598-017-17392-0

Carro, L., Pujic, P., Trujillo, M. E., and Normand, P. (2013). *Micromonospora* is a normal occupant of actinorhizal nodules. *J. Biosci.* 38, 685–693. doi: 10.1007/s12038-013-9359-y

Carro, L., Riesco, R., Spröer, C., and Trujillo, M. E. (2016). *Micromonospora ureilytica* sp. nov., *Micromonospora noduli* sp. nov. and *Micromonospora vinacea* sp. nov., isolated from Pisum sativum nodules. *Int. J. Syst. Evol. Microbiol.* 66, 3509–3514. doi: 10.1099/ijsem.0.001231

Carro, L., Spröer, C., Alonso, P., and Trujillo, M. E. (2012). Diversity of *Micromonospora* strains isolated from nitrogen fixing nodules and rhizosphere of *Pisum sativum* analyzed by multilocus sequence analysis. *Syst. Appl. Microbiol.* 35, 73–80. doi: 10.1016/j.syapm.2011.11.003

Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijs.0.054171-0

Coenye, T., Gevers, D., Van De Peer, Y., Vandamme, P., and Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol. Rev.* 29, 147–167. doi: 10.1016/j.fmrre.2004.11.004

Colwell, R. R. (1970). Polyphasic taxonomy of the genus vibrio: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. *J. Bacteriol.* 104, 410–433.

de la Vega, P. A. (2010). *Distribución, Caracterización E Importancia Ecologica de Micromonospora en Nódulos Fijadores de Nitrogeno de Lupinus*. Ph.D., thesis. Salamanca, Spain, University of Salamanca.

de Menezes, A. B., McDonald, J. E., Allison, H. E., and McCarthy, A. J. (2012). Importance of *Micromonospora* spp. as colonizers of cellulose in freshwater lakes as demonstrated by quantitative reverse transcriptase PCR of 16S rRNA. *Appl. Environ. Microbiol.* 78, 3495–3499. doi: 10.1128/AEM.07314-11

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359

Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.* 125, 1–15. doi: 10.1086/284325

Garcia, L. C., Martínez-Molina, E., and Trujillo, M. E. (2010). *Micromonospora pisi* sp. nov., isolated from root nodules of Pisum sativum. *Int. J. Syst. Evol. Microbiol.* 60, 331–337. doi: 10.1099/ijs.0.012708-0

Genilloud, O. (2015a). "Micromonospora," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, B. Hedlund, and S. Dedysh (Wiley). doi: 10.1002/9781118960608.gbm00148

Genilloud, O. (2015b). "Micromonosporaceae," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, B. Hedlund, and S. Dedysh (Wiley). doi: 10.1002/9781118960608.fbm00041

Genilloud, O. (2015c). "Micromonosporales ord. nov," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, B. Hedlund, and S. Dedysh (Wiley). doi: 10.1002/9781118960608.obm00015

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739. doi: 10.1038/nrmicro1236

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi: 10.1093/bioinformatics/btw313

Hahnke, R. L., Meier-Kolthoff, J. P., García-López, M., Mukherjee, S., Huntemann, M., Ivanova, N., et al. (2016). Genome-based taxonomic classification of *Bacteroidetes*. *Front. Microbiol.* 7:2003. doi: 10.3389/fmicb.2016.02003

Hirsch, A. M., and Valdés, M. (2010). *Micromonospora*: an important microbe for biomedicine and potentially for biocontrol and biofuels. *Soil Biol. Biochem.* 42, 536–542. doi: 10.1016/j.soilbio.2009.11.023

Hirsch, P., Mevs, U., Kroppenstedt, R. M., Schumann, P., and Stackebrandt, E. (2004). Cryptoendolithic actinomycetes from antarctic sandstone rock samples: *Micromonospora endolithica* sp. nov. and two isolates related to *Micromonospora coerulea* Jensen 1932. *Syst. Appl. Microbiol.* 27, 166–174. doi: 10.1078/072320204322881781

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

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

Kaewkla, O., Thamchaipinet, A., Milton, C., and Franco, M. (2017). *Micromonospora terminaliae* sp. nov., an endophytic actinobacterium isolated from the surface-sterilized stem of the medicinal plant Terminalia mucronata. *Int. J. Syst. Evol. Microbiol.* 67, 225–230. doi: 10.1099/ijsem.0.002402

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, 199–205. doi: 10.1093/nar/gkt1076

Kasai, H., Tamura, T., and Harayama, S. (2000). Intrageneric relationships among *Micromonospora* species deduced from *gyrB*-based phylogeny and DNA relatedness. *Int. J. Syst. Evol. Microbiol.* 50, 127–134. doi: 10.1099/00207713-50-1-127

Katayama, T., Tanaka, M., Moriizumi, J., Nakamura, T., Brouchkov, A., Douglas, T. A., et al. (2007). Phylogenetic analysis of bacteria preserved in a permafrost ice wedge for 25,000 years. *Appl. Environ. Microbiol.* 73, 2360–2363. doi: 10.1128/AEM.01715-06

Kittiwongwattana, C., Thanaboripat, D., Laosinwattana, C., Koohakan, P., Parinthawong, N., and Thawai, C. (2015). *Micromonospora oryzae* sp. nov., isolated from roots of upland rice. *Int. J. Syst. Evol. Microbiol.* 65, 3818–3823. doi: 10.1099/ijsem.0.000500

Kroppenstedt, R. M., Mayilraj, S., Wink, J. M., Kallow, W., Schumann, P., Secondini, C., et al. (2005). Eight new species of the genus *Micromonospora*, *Micromonospora citrea* sp. nov., *Micromonospora echinaurantiaca* sp. nov., *Micromonospora echinofusca* sp. nov. *Micromonospora fulviviridis* sp. nov., *Micromonospora inyonensis* sp. nov., *Micromonospora peucetia* sp. nov. *Syst. Appl. Microbiol.* 28, 328–339. doi: 10.1016/j.syapm.2004.12.011

Kumar, N., Lad, G., Giuntini, E., Kaye, M. E., Udomwong, P., Shamsani, N. J., et al. (2015). Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biol.* 5, 140133–140133. doi: 10.1098/rsob.140133

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

Lee, H., and Whang, K. (2017). *Micromonospora fulva* sp. nov., isolated from forest soil. *Int. J. Syst. Evol. Microbiol.* 67, 1746–1751. doi: 10.1099/ijsem.0.001858

Lee, I., Kim, Y. O., Park, S. C., and Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103. doi: 10.1099/ijsem.0.000760

Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800. doi: 10.1093/molbev/msv150

Li, L., and Hong, K. (2016). *Micromonospora ovatispora* sp. Nov. Isolated from mangrove soil. *Int. J. Syst. Evol. Microbiol.* 66, 889–893. doi: 10.1099/ijsem.0.000808

Li, X., Huang, Y., and Whitman, W. B. (2015). The relationship of the whole genome sequence identity to DNA hybridization varies between genera of prokaryotes. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* 107, 241–249. doi: 10.1007/s10482-014-0322-1

Lin, Y. B., Fan, M. C., Guo, Y. Q., Di, X. H., Dong, D. H., Zhang, X., et al. (2015). *Micromonospora nickelidurans* sp. nov., isolated from soil from a nickel-mining site. *Int. J. Syst. Evol. Microbiol.* 65, 4615–4620. doi: 10.1099/ijsem.0.000623

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2018). *Cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.7–1. Available online at: https://cran.r-project.org/web/packages/cluster/

Meier-Kolthoff, J. P., Göker, M., Spröer, C., and Klenk, H.-P. (2013a). When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* 195, 413–418. doi: 10.1007/s00203-013-0888-4

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013b). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., et al. (2014). Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand. Genomic Sci.* 9, 1–19. doi: 10.1186/1944-3277-9-2

Na, S.-I., Kim, Y. O., Yoon, S.-H., Ha, S., Baek, I., and Chun, J. (2018). UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J. Microbiol.* 56, 281–285. doi: 10.1007/s12275-018-8014-6

Nawrocki, E. P., Eddy, S. R., Nawrocki, E. P., and Eddy, S. R. (2016). Computational identification of functional RNA homologs in metagenomic data. *RNA Biol.* 6286, 1170–1179. doi: 10.4161/rna.25038

Nimaichand, S., Zhang, Y. G., Cheng, J., Li, L., Zhang, D. F., Zhou, E. M., et al. (2013). *Micromonospora kangleipakensis* sp. nov., isolated from a sample of limestone quarry. *Int. J. Syst. Evol. Microbiol.* 63, 4546–4551. doi: 10.1099/ijs.0.052746-0

Oren, A., and Garrity, G. M. (2014). Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie van Leeuwenhoek*. 106, 43–56. doi: 10.1007/s10482-013-0084-1

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

Parte, A. C. (2014). LPSN - List of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.* 42, 613–616. doi: 10.1093/nar/gkt1111

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S., and Vandamme, P. (2016). Phylogenomic study of *Burkholderia glathei*-like organisms, proposal of 13 novel Burkholderia species and emended descriptions of *Burkholderia sordidicola*, *Burkholderia zhejiangensis*, and *Burkholderia grimmiae*. *Front. Microbiol.* 7:877. doi: 10.3389/fmicb.2016.00877

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, 686–689. doi: 10.1093/nar/gki366

Shirling, E., and Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *Int. J. Syst. Bacteriol.* 16, 313–340. doi: 10.1099/00207713-16-3-313

Sutcliffe, I. C., Trujillo, M. E., and Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Int. J. Gen. Mol. Microbiol.* 101, 13–20. doi: 10.1007/s10482-011-9664-0

Sutcliffe, I. C., Trujillo, M. E., Whitman, W. B., and Goodfellow, M. (2013). A call to action for the International Committee on Systematics of Prokaryotes. *Trends Microbiol.* 21, 51–52. doi: 10.1016/j.tim.2012.11.004

Suzuki, R., and Shimodaira, H. (2015). *pvclust: Hierarchical Clustering With P-values via Multiscale Bootstrap Resampling*. Available online at: http://cran.r-project.org/web/packages/pvclust/index.html.

Tang, Y., Horikoshi, M., and Li, W. (2016). ggfortify: unified interface to visualize statistical results of popular R packages. *R. J.* 8, 474–485. Available online at: https://cran.r-project.org/web/packages/ggfortify/

Thompson, C., Amaral, G., Campeñao, M., Edwards, R., Polz, M., Bas, E., et al. (2015). Microbial taxonomy in the post-genomic era: rebuilding from scratch? *Arch. Microbiol.* 197, 359–370. doi: 10.1007/s00203-014-1071-2

Thompson, J., Gibson, T., Plewniak, F., Jeanmonguin, F., and Higgins, D. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Res.* 24, 4876–4882. doi: 10.1093/nar/25.24.4876

Trujillo, M. E., Alonso-Vega, P., Rodríguez, R., Carro, L., Cerda, E., Alonso, P., et al. (2010). The genus *Micromonospora* is widespread in legume root nodules: the example of *Lupinus angustifolius*. *ISME J.* 4, 1265–1281. doi: 10.1038/ismej.2010.55

Trujillo, M. E., Bacigalupe, R., Pujic, P., Igarashi, Y., Benito, P., Riesco, R., et al. (2014). Genome features of the endophytic actinobacterium *Micromonospora lupini* Strain Lupac 08: on the process of adaptation to an endophytic life style? *PLoS ONE* 9:e108522. doi: 10.1371/journal.pone.0108522

Trujillo, M. E., Fernández-Molinero, C., Velázquez, E., Kroppenstedt, R. M., Schumann, P., Mateos, P. F., et al. (2005). *Micromonospora mirobrigensis* sp. nov. *Int. J. Syst. Evol. Microbiol.* 55(Pt 2), 877–880. doi: 10.1099/ijs.0.63361-0

Trujillo, M. E., Kroppenstedt, R. M., Fernández-Molinero, C., Schumann, P., and Martínez-Molina, E. (2007). *Micromonospora lupini* sp. nov. and *Micromonospora saelicesensis* sp. nov., isolated from root nodules of Lupinus angustifolius. *Int. J. Syst. Evol. Microbiol.* 57, 2799–2804. doi: 10.1099/ijs.0.65192-0

Trujillo, M. E., Riesco, R., Benito, P., and Carro, L. (2015). Endophytic actinobacteria and the interaction of *Micromonospora* and nitrogen fixing plants. *Front. Microbiol.* 6:1341. doi: 10.3389/fmicb.2015.01341

Vaas, L. A. I., Sikorski, J., Hofner, B., Fiebig, A., Buddruhs, N., Klenk, H. P., et al. (2013). Opm: an R package for analysing OmniLog® phenotype microarray data. *Bioinformatics* 29, 1823–1824. doi: 10.1093/bioinformatics/btt291

Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M., and Klenk, H. P. (2012). Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS ONE* 7:e34846. doi: 10.1371/journal.pone.0034846

Vandamme, P., and Peeters, C. (2014). Time to revisit polyphasic taxonomy. *Int. J. Gen. Mol. Microbiol.* 106, 57–65. doi: 10.1007/s10482-014-0148-x

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60, 407–438.

Veyisoglu, A., Carro, L., Cetin, D., Guven, K., Spröer, C., Pötter, G., et al. (2016). *Micromonospora profundi* sp. nov., isolated from deep marine sediment. I*nt. J. Syst. Evol. Microbiol.* 66, 4735–4743. doi: 10.1099/ijsem.0.001419

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Evol. Microbiol.* 37, 463–464. doi: 10.1099/00207713-37-4-463

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

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 Riesco, Carro, Román-Ponce, Prieto, Blom, Klenk, Normand and Trujillo. 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: 08 May 2018
doi: 10.3389/fmicb.2018.00913

Discovery of Phloeophagus Beetles as a Source of *Pseudomonas* Strains That Produce Potentially New Bioactive Substances and Description of *Pseudomonas bohemica* sp. nov.

*Zaki Saati-Santamaría[1,2], Rubén López-Mondéjar[3], Alejandro Jiménez-Gómez[1,2], Alexandra Díez-Méndez[1,2], Tomáš Větrovský[3], José M. Igual[4,5], Encarna Velázquez[1,2,5], Miroslav Kolarik[3], Raúl Rivas[1,2,5] and Paula García-Fraile[1,2,3]**

[1] Microbiology and Genetics Department, University of Salamanca, Salamanca, Spain, [2] Spanish-Portuguese Institute for Agricultural Research (CIALE), Salamanca, Spain, [3] Institute of Microbiology of the Czech Academy of Sciences, Vestec, Czechia, [4] Institute of Natural Resources and Agrobiology of Salamanca, IRNASA-CSIC, Salamanca, Spain, [5] Associated R&D Unit, USAL-CSIC (IRNASA), Salamanca, Spain

OPEN ACCESS

Edited by:
Jesus L. Romalde,
Universidade de Santiago de Compostela, Spain

Reviewed by:
Learn-Han Lee,
Monash University Malaysia, Malaysia
Mari Carmen Macián,
Universitat de València, Spain

***Correspondence:**
Paula García-Fraile
garcia@biomed.cas.cz;
paulagf81@usal.es

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

Received: 15 December 2017
Accepted: 20 April 2018
Published: 08 May 2018

Citation:
Saati-Santamaría Z, López-Mondéjar R, Jiménez-Gómez A, Díez-Méndez A, Větrovský T, Igual JM, Velázquez E, Kolarik M, Rivas R and García-Fraile P (2018) Discovery of Phloeophagus Beetles as a Source of Pseudomonas Strains That Produce Potentially New Bioactive Substances and Description of Pseudomonas bohemica sp. nov. Front. Microbiol. 9:913. doi: 10.3389/fmicb.2018.00913

Antimicrobial resistance is a worldwide problem that threatens the effectiveness of treatments for microbial infection. Consequently, it is essential to study unexplored niches that can serve for the isolation of new microbial strains able to produce antimicrobial compounds to develop new drugs. Bark beetles live in phloem of host trees and establish symbioses with microorganisms that provide them with nutrients. In addition, some of their associated bacteria play a role in the beetle protection by producing substances that inhibit antagonists. In this study the capacity of several bacterial strains, isolated from the bark beetles *Ips acuminatus*, *Pityophthorus pityographus Cryphalus piceae*, and *Pityogenes bidentatus*, to produce antimicrobial compounds was analyzed. Several isolates exhibited the capacity to inhibit Gram-positive and Gram-negative bacteria, as well as fungi. The genome sequence analysis of three *Pseudomonas* isolates predicted the presence of several gene clusters implicated in the production of already described antimicrobials and moreover, the low similarity of some of these clusters with those previously described, suggests that they encode new undescribed substances, which may be useful for developing new antimicrobial agents. Moreover, these bacteria appear to have genetic machinery for producing antitumoral and antiviral substances. Finally, the strain IA19^T showed to represent a new species of the genus *Pseudomonas*. The 16S rRNA gene sequence analysis showed that its most closely related species include *Pseudomonas lutea, Pseudomonas graminis, Pseudomonas abietaniphila* and *Pseudomonas alkylphenolica*, *with* 98.6, 98.5 98.4, and 98.4% identity, respectively. MLSA of the housekeeping genes *gyr*B, *rpo*B, and *rpo*D confirmed that strain IA19^T clearly separates from its closest related species. Average nucleotide identity between strains IA19^T and *P. abietaniphila* ATCC 700689^T, *P. graminis* DSM 11363^T, *P. alkylphenolica* KL28^T and *P. lutea* DSM 17257^T were 85.3, 80.2, 79.0, and 72.1%, respectively. Growth occurs at 4-37°C and pH 6.5-8. Optimal growth occurs at 28°C,

pH 7–8 and up to 2.5% NaCl. Respiratory ubiquinones are Q9 (97%) and Q8 (3%). C16:0 and in summed feature 3 are the main fatty acids. Based on genotypic, phenotypic and chemotaxonomic characteristics, the description of *Pseudomonas bohemica* sp. nov. has been proposed. The type strain is IA19^T (=CECT 9403^T =LMG 30182^T).

Keywords: antimicrobials, anticarcinogenic, antiviral, genome mining, bark beetles, antibiotic resistance, NRPS-PKS, secondary metabolites

INTRODUCTION

During thousands of years, natural molecules obtained from plants, animals or microorganisms have been the main source of drugs to treat human illnesses. However, the field of chemistry has greatly influenced the availability of drugs by using combinatorial chemistry to produce new molecules (Cragg and Newman, 2009; Mishra and Tiwari, 2011; Dias et al., 2012; Harvey et al., 2015). Currently, natural sources are receiving more attention, as the production of new drugs chemically is becoming more restricted, and it is becoming more difficult to solve the problems that arise in association with Public Health (Ravelo and Braun, 2009; Harvey et al., 2015).

Antimicrobial resistance is a global problem that threatens the effectiveness of the treatment of microbial infections. During the last few decades, the number of multidrug resistant microbes is exponentially increasing, and the WHO warns that if this tendency persists, the number of human deaths derived from microbial infections will be higher by 2050 than those resulting from cancer. This means that antimicrobial resistance is one of the major menaces to public health at the present time (Levy and Marshall, 2004; Hoffman et al., 2015; Premanandh et al., 2016).

The complex characteristics of natural molecules could make them more suitable than chemically synthesized compounds for fighting disease, since they are more similar to the endogenous metabolites of an organism. Also, the complexity of natural molecules sometimes impedes their synthesize using chemical methods (Balamurugan et al., 2005; Ravelo and Braun, 2009; Drewry and Macarron, 2010; Harvey et al., 2015). Therefore, the study of unexplored ecological niches that can serve as a novel source for the isolation of new microbial strains able to produce antimicrobial compounds, which could serve as a basis for the development of new drugs effective in the treatment of microbial infections, is of utmost importance (Kennedy et al., 2007; Cragg and Newman, 2009; Piel, 2011; Harvey et al., 2015).

Bark beetles (*Curculionidae*, *Scolytinae*) belong to a group of insects that live in and feed on the phloem, the inner layer of the bark, in their host trees (Six, 2012). As many other bark beetles, they establish a symbiotic relationship with microorganisms that provide nutrients to the insect (García-Fraile, 2018). In addition, it has also been reported that some of the bacteria associated with these beetles play a role in the protection of the beetle holobiont – the insect and its microbial symbionts– by producing substances that inhibit the development of pathogens and other antagonists (García-Fraile, 2018).

Pseudomonas is a genus of bacteria found in a wide range of habitats (for a revision see Peix et al., 2018) and also includes bacteria that are commonly associated with bark beetles (Adams et al., 2009; Boone et al., 2013; Hu et al., 2014; Menéndez et al., 2015; Xu et al., 2015, 2016). Many strains belonging to this genus are used as biocontrol agents for inhibiting some plant pathogens (for a revision see Olorunleke et al., 2015), which further supports their capability to produce useful antimicrobials. Moreover, many drugs coming from *Pseudomonas* strains with interesting clinical applications have been described. Thus, the search for *Pseudomonas* strains associated with bark beetles is interesting from the point of view of discovering new antimicrobials and other compounds that may be of potential use to the pharmaceutical industry.

As regards, the aim of this work was to study the potential of bacterial bark beetles associates to inhibit Gram positive and negative bacteria and fungi. In addition, a genome sequence analysis was carried out to test the capacity of some of the most promising strains belonging to the genus *Pseudomonas* to produce antimicrobial compounds, which could lead to the development of new antibiotics. Finally, based on phenotypic, genotypic and chemotaxonomic tests, we describe one of the new isolates, strain IA19^T, as a new species of the genus *Pseudomonas*.

MATERIALS AND METHODS

Strains Used in This Study

Bacterial strains used in this study were isolated from adult bark beetles of several species -*Ips acuminatus*, *Cryphalus piceae*, *Pityophthorus pityographus*, and *Pityogenes bidentatus* (all Coleoptera, *Scolytinae*)-. The isolation of the bacterial associates of *C. piceae* and *P. pityographus* has been previously described in Fabryová et al. (2018). The newly obtained bacteria isolated in this study were obtained from *I. acuminatus* and *P. bidentatus* beetles, extracted from branches from *Pinus sylvestris* exhibiting the typical boring holes of bark beetles which were collected in May 2016 in Stará Boleslav, Czech Republic (coordinates: 50°12′59.5″N, 14°41′58.4″E). The branches were taken to the laboratory and then, under aseptic conditions, the bark was removed, and the bark beetles were sorted into 5 groups of 3 individuals each. The beetles were crushed using sterile toothpicks in 500 μl of sterile water. Serial dilutions were made using the suspensions obtained, and 100 μl of each dilution were plated onto Nutrient Agar (NA) and Tryptose Soy Agar (TSA).

The plates were incubated at 28°C for 2 weeks, and the emerging bacterial colonies with different morphologies were regularly passed to new plates in order to obtain pure cultures. The isolated strains were stored in a sterile 20% glycerol solution at −80°C and sub-cultured regularly in the corresponding isolation medium.

Bacterial Identification and Genotypic Analysis

To identify the 21 bacterial isolates obtained in this study, total DNA was extracted and the 16S rRNA gene was amplified and sequenced as previously reported (Rivas et al., 2007). Almost complete (∼1,400 bp) 16S rRNA sequences were compared with those available in GenBank using the BLASTn program (Altschul et al., 1990) and EzTaxon tool (Kim et al., 2012). The remaining 40 strains used in the antimicrobial screening were identified as detailed by Fabryová et al. (2018).

The amplification and sequencing of the housekeeping genes *gyr*B, *rpo*B, and *rpo*D was performed as described by Menéndez et al. (2015).

The phylogenetic analysis of the 16S rRNA gene sequence and the concatenated sequences of the *gyr*B, *rpo*B, and *rpo*D housekeeping genes of strain IA19^T and all the sequences of the closely related species of the genus *Pseudomonas* was done using the MEGA7 software (Kumar et al., 2016), based on the Clustal_W alignment (Thompson et al., 1994; Larkin et al., 2007). The distances were calculated using the Kimura's two-parameter model (Kimura, 1980) and the phylogenetic trees were generated using maximum-likelihood (ML; Rogers et al., 1998) and neighbor-joining (NJ; Saitou and Nei, 1987) analyses.

The average nucleotide identity (ANI) values between the genome sequence of strain IA19^T and the genome sequences of the type strains of the closest related species were estimated by using ANI Calculator in the EZBioCloud (http://www.ezbiocloud.net).

The mol % G+C content of DNA was determined from the complete genome sequence.

Chemotaxonomic Analysis

Biomass for the analysis of fatty acid methyl esters (FAME) and respiratory quinones of strain IA19^T and its closest related type was harvested after the strains were cultivated for 2 days on TSA medium at 28°C. For FAME analyses, the cells were collected from the plates and placed into sterile plastic tubes and freeze dried. The extraction of fatty acids was carried out as described by Sasser (1990) and analyzed using the Microbial Identification System (MIDI) Sherlock 6.1 together with the library RTSBA6. Quinones extraction and identification were performed at the Identification of Microorganisms Service at the DSMZ, were they are extracted using methanol:hexane (Tindall, 1990a,b), followed by phase separation into hexane, separated into their different classes by thin layer chromatography on silica gel (Macherey-Nagel Art. No. 805 023), using hexane:tert-butylmethylether (9:1 v/v) as solvent. UV absorbing bands corresponding to the different quinone classes are removed from the plate and analyzed by HPLC.

Bacterial Characterization

Gram-staining of strain IA19^T was carried out following the protocol described by Doetsch (1981), and motility was checked by phase-contrast microscopy after growing the cells at 22°C for 48 h in NA. The type of flagellation was determined by electron microscopy as previously described by García-Fraile et al. (2015).

Tryptose Soy Broth (TSB) medium supplemented with 0–10% (w/v) NaCl was used to assay salt tolerance in IA19^T. The same medium, with an adjusted final pH in the range of 4–10, was used for studying the growth capability of the strain at different pHs; in both cases the cultures were incubated at 28°C for up to 1 week. Also, cells were cultured on TSA plates at 4, 8, 22, 28, 37, and 42°C to determine the temperature range for growth. In all cases the presence of growth was checked during 1 week.

For the catalase test, bacterial cells growing in an TSA agar plate were collected and drops of 30% H_2O_2 were added over them to detect the formation of bubbles after 5 min, indicating a positive result. The oxidase test was performed following the protocol described by Kovacs (1956).

Finally, the phenotypic characterization of strain IA19^T and the type strains of the closest related species was completed using the API20NE and API50CH (bioMerieux) systems.

Screening for Antimicrobial Production

The antimicrobial activity screen using the 61 isolates analyzed in this study was carried out on 6 indicator strains: A Gram-negative bacterium, *Klebsiella oxytoca*, a Gram-positive bacterium, *Arthrobacter phenanthrenivorans* two yeast-like fungi, *Candida humilis* and *Pichia fermentans* and two filamentous fungi, *Aspergillus* sp. and *Fusarium* sp.

To analyze the capability of each of the isolates to inhibit the different bacteria and yeast-like indicator strains, cross streak method was used. Each of our isolates was seeded by a single streak in the center of a NA plate. After 5 days of incubation at 28°C, the plates were seeded with the indicator microorganisms by single streaks perpendicular to the central one. After additional 48h incubation, the growth of each indicator strain was analyzed. To study the inhibition of filamentous fungi, our isolates were seeded by a single streak in the center of a NA plate. After 5 days of pre-incubation at 28°C, fungi were inoculated 2 cm far from the central strike by 0.5 cm mycelia discs. After further incubation under suitable conditions for the fungal strains tested (25°C), the diameters of fungal growth in control and sample plates were measured, and the antifungal effect was evaluated.

Draft Genome Sequencing and Annotation

The genomic DNA for genome sequencing was obtained from bacterial cells of strains IA19^T, A2-NA12, and A2-NA13 grown on NA plates and collected after 24 h at 28°C, using the ZR Fungal/Bacterial DNA MiniPrep (Zymo Research).

The draft genome sequences of the selected isolates were obtained by shotgun sequencing on an Illumina MiSeq platform via a paired-end run (2 × 251 bp). The sequence data were assembled using Velvet 1.2.10 (Zerbino and Birney, 2008) and a draft genome was obtained. Gene calling and annotation was performed using RAST 2.0 (Rapid Annotation using Subsystem Technology) (Aziz et al., 2008). The SEED-viewer framework (Overbeek et al., 2014) was used for a first mining of genes related to antimicrobial production genes. Moreover, a more specific and detailed analysis of the presence of gene clusters related to antimicrobial substances productions and other secondary

metabolites was performed using antiSMASH 3.0 (Weber et al., 2015).

RESULTS

Bacterial Isolation and Identification

The list of bacterial isolates analyzed in this work, as well as their identification based on their partial (≈1,400 bp) 16S rRNA sequence, is presented in **Table 1**. The list includes the isolates from *I. acuminatus* and *P. bidentatus*, obtained in this study, and the ones from the bark beetles *C. piceae* and *P. pityographus* obtained in a study that tested their capacity to degrade plant cells (Fabryová et al., 2018).

Phylogenetic Analyses and Average Nucleotide Identity (ANI) Comparison

Analysis of the 16S rRNA gene sequence of strain IA19^T suggested that this isolate was a member of the genus *Pseudomonas* but could also belong to a new species within this genus. The most closely related species were *P. lutea* DSM 17257^T, (98.6% identity), *P. graminis* DSM 11363^T, (98.5% identity), *P. abietaniphila* ATCC 700689^T, (98.4% identity) and *P. alkylphenolica* KL28^T, (98.4% identity).

ML and NJ trees including all related species within the genus *Pseudomonas* showed a similar 16S rRNA sequence phylogenetic clustering, in which strain IA19^T clustered with *P. alkylphenolica* KL28^T in a broader cluster that also contained *P. lutea* OK2^T and *P. graminis* DSM 11363^T (**Figure 1** and **Supplementary Figure 1**). Nevertheless, the phylogenetic distances between strain IA19^T and its closest related species were broader than those distances among several other different species belonging to this genus, suggesting its classification into a different species. However, the limitations of the analysis based on 16S rRNA gene sequences to discriminate the genus *Pseudomonas* sufficiently at the inter-species level have been described (Ramírez-Bahena et al., 2014). Therefore, the Mutli-Locus Sequence Analysis (MLSA) based on the three housekeeping genes *gyr*B, *rpo*B, and *rpo*D was used for species classification in *Pseudomonas* (Ait Tayeb et al., 2005; Mulet et al., 2009, 2010, 2012; Ramos et al., 2013; Toro et al., 2013; Ramírez-Bahena et al., 2014; Menéndez et al., 2015). Both the NJ and ML analysis of the concatenated housekeeping *gyr*B, *rpo*B and *rpo*D genes sequences (**Figure 2** and **Supplementary Figure 2**) showed that the strain IA19^T clustered with *P. abitaniphila* DSM 17554^T, but the phylogenetic distance between both strains clearly indicated that they belonged to different species. The sequence similarities between strains IA19^T and *P. abietaniphila* DSM 17554^T housekeeping genes *rpo*D, *rpo*B, and *gyr*B were 89.3, 92.4, and 87.5%, respectively. These values were below those commonly found among different but closely related species of the genus *Pseudomonas* (Ramírez-Bahena et al., 2014).

Finally, authenticity of the novel species was confirmed by ANI comparison between strain IA19^T and the genome sequence of the type strains of its closest related strains *P. abietaniphila* ATCC 700689^T (GenBank access number: FNCO00000000.1, draft genome), *P. graminis* DSM 11363^T (GenBank access number: NZ_FOHW01000000.1, draft genome), *P. alkylphenolica* KL28^T (GenBank access number: CP009048.1, complete genome) and *P. lutea* DSM 17257^T (GenBank access number: FOEV01000000.1, draft genome) showed values of were 85.31, 80.21, 79.0, and 72.15%, respectively (**Table 2**). Considering that a threshold value of 95–96% ANI has been established for defining a bacterial species (Richter and Rosselló-Móra, 2009), our results indicated that strain IA19^T represented a distinct species of the genus *Pseudomonas*.

Colony and Cellular Morphology

The strain IA19^T formed round, bright, clear beige and convex colonies with entire border on TSA medium, which were visible after 24 h of incubation at 28°C; these colonies grew to a size of 1–3 mm after 72 h of growth. Also, growth was observed in the temperature range of 4–37°C, where the optimum temperature was at 28°C, and at a pH range of 6.5 and 8, with 7 being the optimum. Cells were Gram-negative, rod-shaped and motile by means of a polar flagellum (**Figure 3**).

Phenotypic and Chemotaxonomic Characterization

The analysis of isopropenoid quinones showed that the isolate IA19^T contained ubiquinone-9 (Q9) (97%) as the main respiratory quinone, which is one of the most common quinone systems in those strains belonging to the genus *Pseudomonas* (Oyaizu and Komagata, 1983), and a small amount of ubiquinone-8 (Q8) (3%).

As for its closest related species, the major fatty acids of strain IA19^T were 16:0 iso (30.5%) and in summed feature 3 (21.1%), being the whole FAME composition detailed in **Table 3**.

The main phenotypic features observed for strain IA19^T are detailed in the species description, and the main differences found between this strain and its closest related species, as well as the type species of the genus, *Pseudomonas aeruginosa*, are listed in **Table 4**.

In Vitro Detection of the Production of Antimicrobial Substances

From the total of 61 isolates, all showed antimicrobial activity against at least one of the indicator strains (**Table 1**). With the only exception of two strains, the analyzed isolates from this study inhibited the yeast's growth, 4.9% isolates inhibited the growth of *K. oxytoca* and 13.1% isolates inhibited the growth of *A. phenanthrenivorans*. Moreover, the strain from *Aspergillus* sp. was inhibited by all the strains of this study (**Table 1**).

Several of the isolates obtained in this study appeared to be potential candidates to produce antimicrobial substances, being several strains identified as *Pseudomonas* some of the best antimicrobial producers, according to our screening. This genus is known to produce antimicrobials and other interesting bioactive compounds (Laine et al., 1996; Stintzi et al., 1996; Raaijmakers et al., 1997; Marinho et al., 2009; Bauer et al., 2015; Nishanth Kumar et al., 2016; Ganne et al., 2017). Based on that, strains *Pseudomonas* IA19^T, *Pseudomonas* A2-NA12, *Pseudomonas* A2-NA13 were selected for further analysis of their genetic potential to produce pharmaceutically interesting molecules by sequencing and mining their genomes.

TABLE 1 | List of the bacterial strains analyzed in this study and their capability to inhibit reference microbial strains.

						Inhibition of:					
Isolate	**Origin**	**Beetle**	**Tree**	**Closest related species based on 16S rRNA sequence similarity**	**Similarity (%)**	***K. oxytoca***	***C. humilis***	***P. fermentans***	***A. phenanthrenivorans***	***Aspergillus* sp.**	***Fusarium* sp.**
P3-NA1	Ref	PP	Ps	*Yersinia ruckeri* ATCC 29473	97.4	–	+	–	–	+	–
P3-NA2	Ref	PP	Ps	*Yersinia ruckeri* ATCC 29473	97.5	–	+	–	–	w	–
A1-NA1	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.7	–	+	+	–	+	–
A1-NA3	Ref	CP	Aal	*Pantoea* sp.	–	–	+	–	–	+	–
A1-NA4	Ref	CP	Aal	*Curtobacterium plantarum* CIP 108988	98.3	–	–	+	–	+	–
A2-NA4	Ref	CP	Aal	*Curtobacterium plantarum* CIP 108988	98.3	–	+	–	–	w	–
A2-NA5	Ref	CP	Aal	*Curtobacterium plantarum* CIP 108988	98.3	–	+	–	–	+	–
A2-NA8	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.7	–	–	+	–	+	–
A2-NA12	Ref	**CP**	**Aal**	***Pseudomonas arsenicoxydans* VC-1**	**99.9**	–	**+**	**+**	**w**	**w**	**+**
A2-NA13	Ref	**CP**	**Aal**	***Pseudomonas arsenicoxydans* VC-1**	**99.9**	–	**+**	**+**	**+**	**+**	**+**
A3-NA1	Ref	CP	Aal	*Yersinia ruckeri* ATCC 29473	97.5	–	–	–	–	+	–
A3-NA4	Ref	CP	Aal	*Yersinia ruckeri* ATCC 29473	97.5	–	+	–	–	+	–
P2-1	Ref	PP	Ps	*Staphylococcus pasteuri* ATCC 51129	99.9	–	–	+	–	+	–
P2-2	Ref	PP	Ps	*Staphylococcus epidermidis* ATCC 14990	100.0	–	–	+	–	+	–
P2-3	Ref	PP	Ps	*Staphylococcus pasteuri* ATCC 51129	99.7	–	–	+	–	+	–
P2-4	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.7	–	+	–	–	+	–
P2-5	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.9	–	–	+	–	+	–
P2-6	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.9	–	–	–	–	w	–
P2-7	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.7	–	–	+	–	+	–
P3-1	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.9	–	+	–	–	+	–
P3-2	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.5	–	+	–	–	+	–
P3-4	Ref	PP	Ps	*Curtobacterium plantarum* CIP 108988	99.9	–	+	–	–	+	–
P4-2	Ref	PP	Ps	*Staphylococcus epidermidis* ATCC 14990	99.9	–	+	w	–	+	–
P4-3	Ref	PP	Ps	*Staphylococcus epidermidis* ATCC 14990	99.9	–	–	+	–	+	–
P4-4	Ref	PP	Ps	*Staphylococcus epidermidis* ATCC 14990	100.0	–	+	–	–	+	–
P5-3	Ref	PP	Ps	*Staphylococcus pasteuri* ATCC 51129	99.9	–	+	–	–	+	–
P5-4	Ref	PP	Ps	*Staphylococcus epidermidis* ATCC 14990	99.9	–	–	+	–	w	–
P5-5	Ref	PP	Ps	*Staphylococcus epidermidis* ATCC 14990	100.0	–	–	+	–	+	–
P5-11	Ref	PP	Ps	*Erwinia billingiae* Eb661	99.7	–	–	–	–	+	–
A1-1	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.6	–	–	–	–	+	–
A1-2	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.4	–	+	–	–	+	–
A1-3	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.6	–	+	–	–	w	–
A2-2	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.7	–	+	–	–	+	–
A2-5	Ref	CP	Aal	*Curtobacterium plantarum* CIP 108988	98.3	–	+	–	–	w	–
A3-2	Ref	CP	Aal	*Yersinia ruckeri* ATCC 29473	97.4	–	–	+	–	+	–
A3-3	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.7	–	–	+	–	+	–
A3-7	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.6	–	+	+	–	+	–
A3-8	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.5	–	+	–	–	+	–
A3-9	Ref	CP	Aal	*Erwinia billingiae* Eb661	99.5	–	+	–	–	+	–
A4-1	Ref	CP	Aal	*Staphylococcus epidermidis* ATCC 14990	99.9	–	–	+	–	+	–
IA2	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	–	+	+	w	w	w
IA3	N	IA	Ps	*Micrococcus yunnanensis* YIM 65004 (T)	99.7	–	+	+	–	+	–

(Continued)

TABLE 1 | Continued

Isolate	Origin	Beetle	Tree	Closest related species based on 16S rRNA sequence similarity	Similarity (%)	Inhibition of: *K. oxytoca*	*C. humilis*	*P. fermentans*	*A. phenanthrenivorans*	*Aspergillus* sp.	*Fusarium* sp.
IA4	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	–	+	w	–	+	–
IA5	N	IA	Ps	*Pseudomonas koreensis* AF 468452	99.4	–	+	w	–	+	+
IA6	N	IA	Ps	*Pseudomonas yamanorum* 8H1 (T)	99.9	–	+	+	–	+	+
IA7	N	IA	Ps	*Arthrobacter globiformis* NBRC 12137 (T)	99.5	–	+	+	–	w	–
IA9	N	IA	Ps	*Pseudomonas yamanorum* 8H1 (T)	99.9	–	+	+	–	w	–
IA12	N	IA	Ps	*Pseudomonas gessardii* DSN 17152 (T)	99.9	–	+	+	–	+	+
IA13	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	w	+	+	–	w	–
IA14	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	–	+	+	w	+	–
IA15A	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	–	+	+	w	+	–
IA15B	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	–	+	+	w	+	–
IA16	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	w	+	+	–	w	–
IA17	N	IA	Ps	*Pseudomonas gessardii* DSN 17152 (T)	99.8	–	+	+	–	+	+
IA18	N	IA	Ps	*Pseudomonas lutea* DSM 17257(T)	98.6	–	+	+	w	w	w
IA19	**N**	**IA**	**Ps**	***Pseudomonas lutea* DSM 17257(T)**	**98.6**	**+**	**+**	**+**	**+**	**+**	**+**
PB2	N	PB	Ps	*Acidovorax radicis* N35 (T)	99.5	–	–	–	–	+	–
PB4	N	PB	Ps	*Pseudomonas agarici* NCPPB 2298 (T)	98.7	–	+	+	–	+	–
PB6	N	PB	Ps	*Erwinia billingiae* CIP 106121 (T)	100.0	–	+	–	–	+	–
PB8	N	PB	Ps	*Pseudomonas cichoriii* ATCC 10857 (T)	99.2	–	+	+	–	+	–
PB9	N	PB	Ps	*Erwinia typography* DSM 22678 (T)	98.2	–	+	–	–	w	–

Ref, Isolated previously in Fabryová et al. (2018); N, Novel isolates; PP, Pityophthorus pityographus; CP, Cryphalus piceae; IA, Ips acuminatus; PB, Pityogenes bidentatus; Ps, Pinus sylvestris; Aal, Abies alba; +, It does exist inhibition; –, It does not exist inhibition; w, weakly inhibition. Bold values are for those strains selected for genome sequencing.

Genomic Properties

The size of the genome of the isolate IA19^T was estimated to be 6.487 Mb with 5,961 predicted coding sequences, A2-NA12 was estimated to be 5.933 Mb with 5,148 predicted coding sequences and A2- NA13 was estimated to be 5.938 Mb with 5,172 predicted coding sequences. The GC content was predicted to be 59.5 for IA19^T and 59.7 for A2-NA12 and A2-NA13. The features of all three genomes are summarized in **Table 5**.

Draft genome sequences of strains *Pseudomonas bohemica* IA19^T, A2-NA12 and A2-NA13 were deposited in GenBank under the accession numbers NKHL00000000, PEGA00000000 and PEGB00000000, respectively.

Genome Mining of Biosynthetic Gene Clusters With Potential Interest in the Pharmaceutical Industry

Using the SEED viewer, based on the automatic annotation of the bacterial genomes performed with RAST, it was found that each of the genomes of the three bacteria IA19^T, A2-NA12, and A2-NA13 presented a cluster of genes implicated in the production of the peptide antibiotic colicin V. Also, all strains showed genes implicated in the synthesis, reception and transport of siderophores, molecules implicated not only in the acquisition of iron, but also in microbial inhibition (Becerra et al., 2003). Strains A2-NA12 and A2-NA13 appeared to possess clusters of genes implicated in the synthesis of the siderophores pyoverdine and enterobactin. A gene from a cluster related to achromobactin siderophore was annotated in the draft genome sequence of strain IA19^T.

The prediction of gene clusters involved in secondary metabolite biosynthesis by antiSMASH version 3.0.5 (Weber et al., 2015) suggested that the genome of the three bacterial isolates contained several potential biosynthetic gene clusters encoding the production of antimicrobial substances, as well as other compounds with potential interest in the pharmaceutical industry, which are listed in **Table 6**.

Specifically, the draft genome of isolate IA19^T included 49 gene clusters related to the synthesis of secondary metabolites. Out of these, 13 were predicted to be related to already described metabolic pathways involved in antimicrobials production: 2 polyketide synthases (PKSs), 2 non-ribosomal peptide synthetases (NRPS), a terpene-siderophore hybrid, 2 bromophenols (caryoynecins), 5 saccharides and a fatty acid.

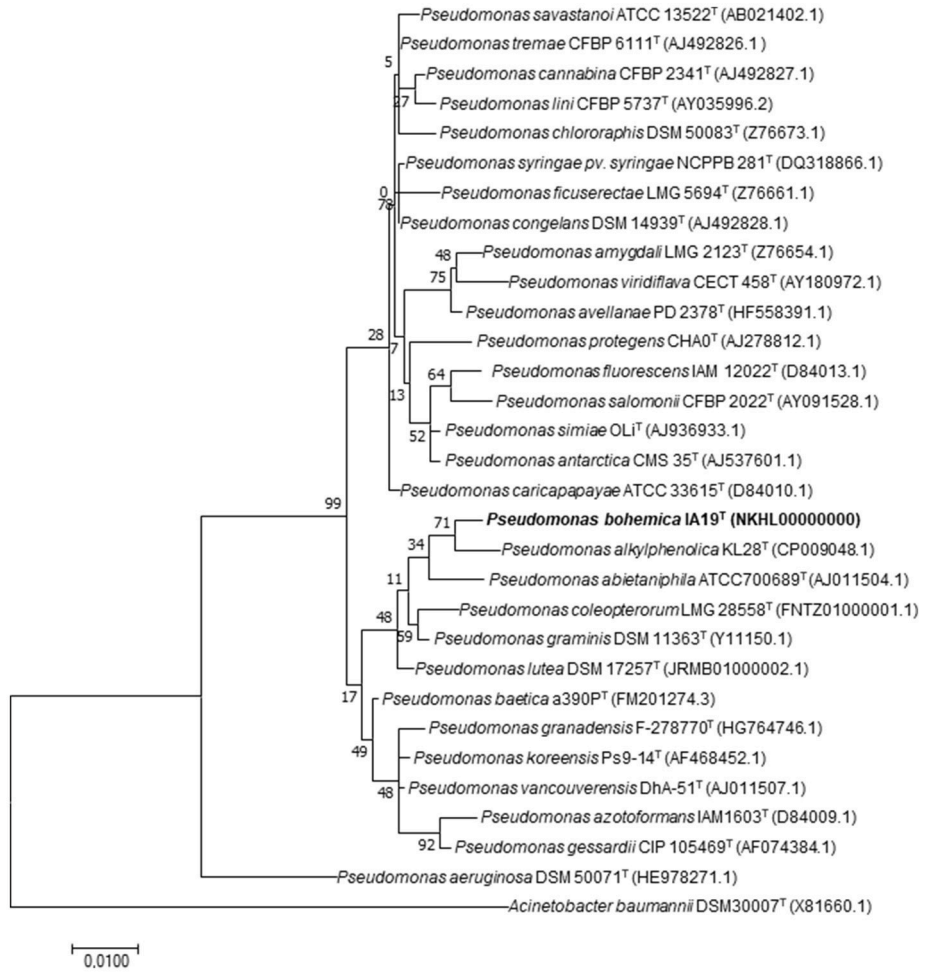


FIGURE 1 | Maximum-likelihood phylogenetic tree based on nearly complete (1,400 bp) 16S rRNA gene sequences of all *Pseudomonas* species closely related to *P. bohemica* IA19^T and the species *Acinetobacter baumannii* DSM30007 ^{T}, which was included as an outgroup. Bootstrap values (expressed as percentages of 1,000 replications) are shown at the branching points. Scale bar = 1 nucleotide (nt) substitutions per 100 nt. Accession numbers of the sequences are indicated in parentheses.

In the case of strain A2-NA12, its draft genome contained 37 clusters encoding secondary metabolites, of which 10 are related to described clusters of genes implicated in the synthesis of enzymes related to the production of antimicrobial compounds: 2 NRPS, a PKS, a bacteriocin, 3 saccharides, a NRPS-PKS hybrid, a NRP-PKS-saccharide hybrid and a fatty acid-saccharide hybrid.

Finally, the draft genome of strain A2-NA13 included 40 clusters encoding the synthesis of secondary metabolites, where 10 of them were associated to known metabolic pathways for antimicrobials synthesis. These included 2 PKS, a lipopolysaccharide, a bacteriocin, 2 NRPS, a NRPS-PKS hybrid, a saccharide, a fatty acid-saccharide hybrid and a NRPS-PKS-saccharide hybrid.

DISCUSSION

Several recent studies have suggested the implication of some bacteria associated with bark beetles in the protection of the bark beetle holobiont. This occurs through the inhibition of antagonists of the beetle itself or the beetle's symbionts, or the detoxification of the bark beetle environment (García-Fraile, 2018). Indeed, all bacterial isolates from this study were able to reduce or inhibit *Aspergillus*, a fungus which has been shown to greatly reduced the number of larvae in the mountain pine beetle (Therrien et al., 2015) and the spruce bark beetle (Cardoza et al., 2006).

In this study, it has been shown that many bacterial isolates from bark beetles, such as *I. acuminatus*, *C. piceae*, *P. pityographus*, and *P. bidentatus*, have the potential to produce antimicrobial compounds. Therefore, the genetic potential of some of these isolates to produce antibiotics, as well as other bioactive compounds was further analyzed. Bacteria from the genus *Pseudomonas* are frequently isolated from bark beetles of different species and life cycle stages (Adams et al., 2009, 2013; Morales-Jiménez et al., 2012; Menéndez et al., 2015; Fabryová et al., 2018; García-Fraile, 2018). Members of

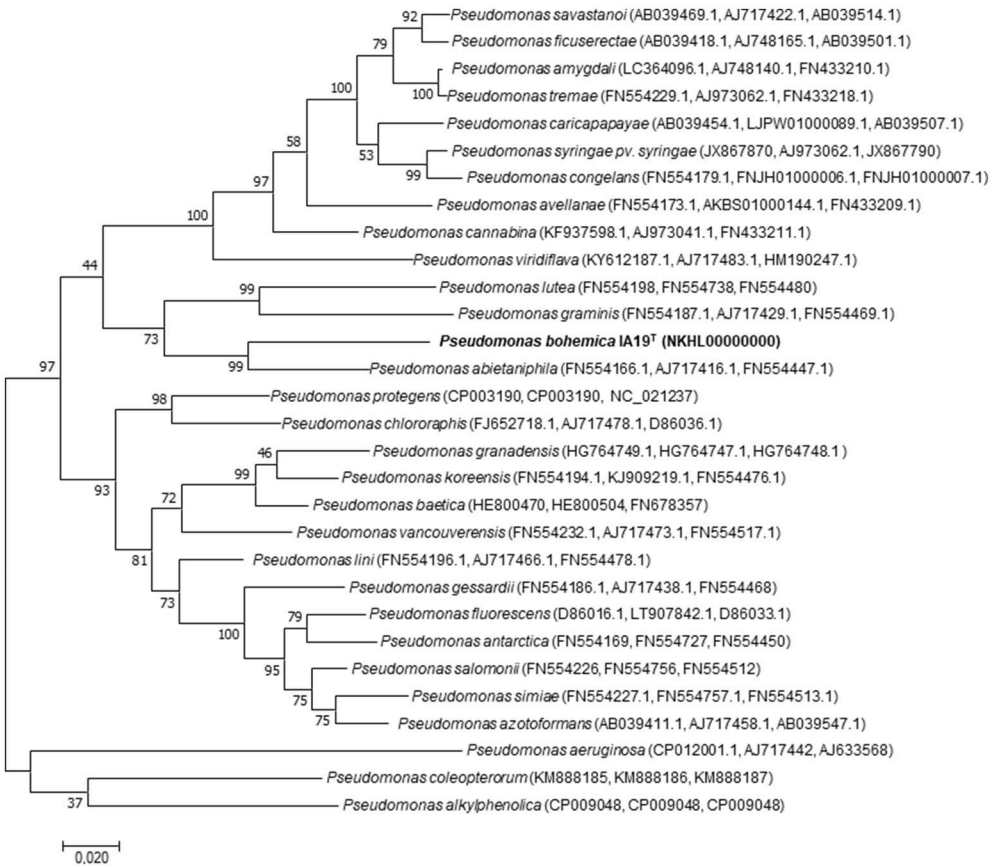


FIGURE 2 | Maximum-likelihood phylogenetic tree based on concatenated partial *gyr*B, *rpo*B, *rpo*D gene sequences of strain *P. bohemica* IA19^T and closely related species of the genus *Pseudomonas*. Bootstrap values (expressed as percentages of 1,000 replications) are shown at the branching points. Scale bar = Bar, 2 nucleotide (nt) substitutions per 100 nt. Accession numbers of the sequences are indicated in parentheses.

the genus *Pseudomonas* have been broadly studied for their capability to produce several different secondary metabolites with potential biotechnological applications. In several articles, different strains of *Pseudomonas* are described as biological control agents of plant diseases (Haas and Keel, 2003) and bioactive substances described from *Pseudomonas* strains are diverse, and include quinolines, pyrroles, pseudopeptide pyrrolidinediones, pseudopyronines, siderophores, phthalates, phenazine, phloroglucinol, benzaldehyde, phenanthren, moiramides, andrimid, zafrin, caryoynencin, and bushrin (Pierson et al., 1994; Sarniguet et al., 1995; Schnider et al., 1995; Yamaguchi et al., 1995; Raaijmakers et al., 1997; Isnansetyo and Kamei, 2009; Marinho et al., 2009; Bauer et al., 2015; Ganne et al., 2017). Therefore, these bacteria may be implicated in the protection of the bark beetle by inhibiting microbial antagonists. In this sense, it was observed that bacteria belonging to this genus were among the best microbial inhibitors in our study. Thus, the genomes of those isolates identified as *Pseudomonas* were selected and sequenced, and the exploration of their metabolic capacity to produce bioactive compounds revealed the presence of numerous clusters that are potentially implicated in the synthesis of important bioactive metabolites (**Table 6**).

A cluster related to the synthesis of 2-amino-4-methoxy-3-butenoic acid, which is a compound also found in *P. aeruginosa* that has been described as being antitumoral (Tisdale, 1980) and to have antiparasitic activity against *Acantamoeba castellani* and antimicrobial activity against *Erwinia amylovora*, *Bacillus* spp. and *Escherichia coli* (Lee et al., 2010, 2012, 2013), was predicted in strain IA19^T.

A cluster of genes for the synthesis of PM100117 and PM100118, two bioactive polyhydroxyl macrolide lactones, which have been isolated from the culture broth of the marine-derived *Streptomyces caniferus* (Pérez et al., 2016), was discovered in the genome sequence of strain IA19^T. It has been shown that these molecules have antitumoral activity as well as slight antifungal activity against *Candida albicans* (Pérez et al., 2016).

A cluster related to the synthesis of caryoynencin, an antibiotic found in liquid cultures of the plant pathogen *Pseudomonas caryophylli* (Yamaguchi et al., 1995) and in a *Burkholderia* sp. isolate from a beetle (Flórez et al., 2017), showing potent antimicrobial activities against Gram-positive and Gram-negative bacteria, as well as antifungal activity, has been predicted in the genome sequence of strain IA19^T. The spectrum of activity of the described caryoynencin

TABLE 2 | Average nucleotide identity (ANI) comparison between strain IA19^T and the genome sequence of its closest relatives.

ANI	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
1	100				
2	72.15	100			
3	80.21	69.83	100		
4	79.00	72.17	75.76	100	
5	85.31	71.69	79.53	77.39	100

1, P. bohemica IA19^T (NKHL00000000); 2, P. lutea LMG 21974 ^{T} (FOEV01000000.1); 3, P. graminis DSM 11363^T (NZ_FOHW01000000.1); 4, P. alkylpenolica KL28^T (CP009048.1); 5, P. abietaniphila ATCC 700689^T (FNCO00000000.1).

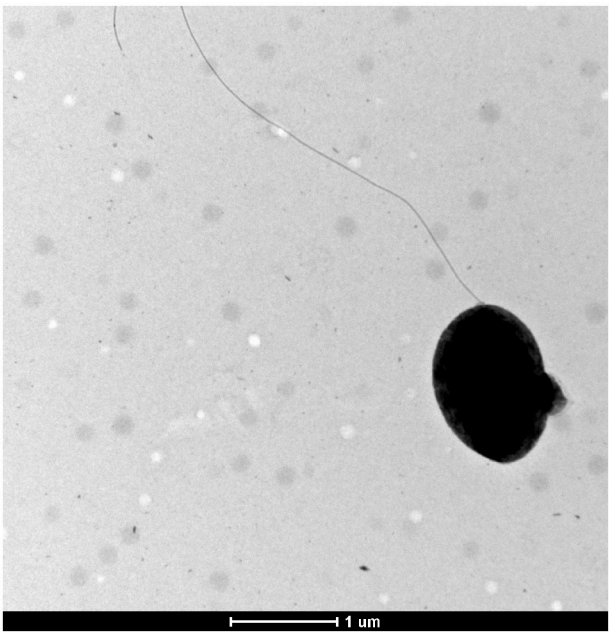


FIGURE 3 | Electron micrograph showing the morphology and flagellation type of *Pseudomonas bohemica* IA19^T.

TABLE 3 | Cellular fatty acid composition (%) of *Pseudomonas bohemica* IA19^T and its closest related species.

Fatty acid	1	2	3	4	5	6
$C_{10:0}$ 3OH	2.6	2.2	1.7	1.5	7.8	3.6
$C_{12:0}$	5.7	5.2	4.6	4.9	3.2	4.8
$C_{12:0}$ 2OH	4.0	3.5	2.9	2.8	7.1	3.7
$C_{12:0}$ 3OH	5.7	4.0	4.1	3.5	7.8	4.5
Sum In Feature 3¶	20.1	27.2	38.7	37.2	26.7	20.0
$C_{16:0}$	30.5	24.7	27.7	25.6	27.8	20.5
$C_{17:0}$ cyclo	14.6	10.3	1.3	1.7	6.0	ND
$C_{17:0}$	0.3	1.9	0.3	0.6	ND	TR
Sum In Feature 5¥	1.5	0.4	ND	ND	ND	20.0
Sum In Feature 8§	11.4	17.3	17.4	21.2	10.5	38.9
$C_{19:0}$ Cyclo w8c	1.3	0.5	ND	ND	ND	ND

Strains: 1, P. bohemica IA19^T; 2, P. abietaniphila DSM 17554^T; 3, P. graminis DSM 11363^T; 4, P. lutea OK2^T; 5, P. alkylphenolica KL28^T (Mulet et al., 2015); 6, P. aeruginosa ATCC 10145^T (Menéndez et al., 2015). Values are percentages of total fatty acids. Those values under 1 per cent for both the strains are not included.
¶ Sum in Feature 3: 16:1 w7c/16:1 w6c.
¥ Sum in Feature 5: 18:2 w6,9c/18:0 ante.
§ Sum in Feature 8: 18:1 w7c/: 18:1 w6c.
Values are percentages of total fatty acids.
Those values under 1 per cent for both the strains are not included.
ND, No Data; TR, Traces.

includes: methicillin-resistant *Staphylococcus aureus*, *Bacillus subtilis*, *Enterococcus faecalis*, *Escherichia coli*, "*Salmonella enteritidis*," *Klebsiella pneumoniae*, *Serratia marcescens*, *Proteus vulgaris*, *Shigella flexneri*, *Enterobacter cloacae*, *P. aeruginosa*, *Candida albicans*, "*Cryptococcus neoformans*," *Mucor mucedo*, *Aspergillus fumigatus*, *Microsporum gypseum*, *Trichophyton mentagrophytes*, *Trichophyton interdigitale*, and *Trichophyton rubrum* (Yamaguchi et al., 1995).

In the genome sequences of the three isolates, a cluster related to the synthesis of a lipopolysaccharide compound, also found in *Escherichia coli* and described as a potent inhibitor of HIV-1 replication in T lymphocytes and macrophages (Verani et al., 2002), was predicted.

A cluster for the synthesis of a carotenoid was predicted in the genome sequence of strain IA19^T. Carotenoids have been reported to have many health benefits, such as prevention of cancer, improvement of visual function and enhancement of immune responses (Sedkova et al., 2005).

A cluster for the synthesis of an alginate, which has been described to improve the efficacy of some antibiotics (Onsoyen et al., 2010), has been predicted in the genome sequence of strains IA19^T and A2-NA12.

In the genome sequence of strain IA19^T, a cluster for the synthesis of pseudopyronines A and B was predicted. Both substances have antibacterial activity based on selective membrane disruption and inhibition of fatty-acid synthase against *Mycobacterium tuberculosis*, *Bacillus subtilis*, *Pseudomonas savastanoi*, methicillin-resistant *Staphylococcus aureus*, *Moraxella catarrhalis*, and vancomycin-resistant Enterococci. This strain also displayed moderate inhibition of other *Firmicutes* (*Listeria welshimeri*) and *Actinobacteria* (*Micrococcus luteus*, *Arthrobacter crystallopoietes*, and *Corynebacterium xerosis*), as well as anti-leishmanial and algaecide activities (Bauer et al., 2015). Moreover, pseudopyronine B has also shown antitumoral activity (Nishanth Kumar et al., 2016). These compounds have been identified in different species of the genus *Pseudomonas* and in the genus *Alteromonas* (Bauer et al., 2015).

In the genome sequences of the three bacterial isolates, clusters related to the synthesis of pyoverdine, a siderophore and an iron chelating agent produced also be *P. aeruginosa* (Ganne et al., 2017), which can act as antibacterial (Becerra et al., 2003), have been predicted. This molecule has also been described as producer of oxidative stress in leukocytes (Becerra et al., 2003). Moreover, when pyoverdina is conjugated with antibiotics it facilitates them to overcome the bacterial membrane (Kinzel et al., 1998; Kinzel and Budzikiewicz, 1999), which could be used against antibiotic resistance.

TABLE 4 | Phenotypic differences between *Pseudomonas bohemica* IA19^T and its closest related species.

	1	2	3	4	5	6
Arginine dihydrolase	−	+	+	−	+	+
Urease	−	+	+	−	−	d
Hydrolysis of esculin	+	−	+	+	+	−
Asimilation of:						
L-arabinose	+	+	+	+	−	−
Glycerol	+	+	−	+	nd	+
L (+) fructose	+	+	−	nd	+	+
Rhamnose	−	+	−	−	nd	−
Sorbitol	−	+	+	−	nd	−
Sucrose	−	+	+	−	−	−
D-maltose	−	+	−	−	−	−
Capric acid	−	+	+	+	+	nd
Malic acid	−	+	+	+	+	nd
Adonitol	−	+	w	+	−	−
Galactose	−	+	w	+	−	−
Mannose	+	+	−	+	−	−
Meso-inositol	−	+	w	nd	−	−
Manitol	−	+	w	+	−	+
Amygdalin	+	−	−	−	−	nd
Arbutine	+	−	−	−	nd	nd
Salicin	+	−	−	−	nd	nd
Melobiose	−	−	+	+	nd	−
D (-) trehalose	−	+	−	−	−	−
Xylitol	−	+	+	+	−	−
D-Fucose	−	+	+	+	nd	nd
L-Fucose	−	+	+	+	−	−
D-Arabitol	−	+	+	+	−	−
L-Arabitol	−	+	+	+	nd	nd
Gluconate	−	+	+	+	+	−
2-ketogluconate	−	+	+	+	nd	+
5-ketogluconate	−	−	+	−	nd	nd

Strains: 1, P. bohemica IA19^T; 2, P. abietaniphila DSM 17554^T; 3, P. graminis DSM 11363^T; 4, P. lutea OK2^T; 5, P. alkylphenolica KL28^T (Mulet et al., 2015; Frasson et al., 2017); 6, P. aeruginosa ATCC 10145^T (Palleroni, 2005; Clark et al., 2006; Xiao et al., 2009).

+, The bacteria can growth; −, The bacteria cannot growth; w, the bacteria growths weakly; nd, No data; d, variable.

TABLE 5 | Features of the draft genomes of the three *Pseudomonas* isolates sequenced in this study.

Characteristic	IA19^T	A2-NA12	A2-NA13
Genome size (bp)	6,487,705	5,933,382	5,938,884
GC content	59.5	59.7	59.7
Number of contigs	127	84	80
Predicted coding sequences	5,961	5,148	5,172
Subsystems	534	524	531
Number of RNAs	70	68	71
Glycosyl hydrolases (GH)	38	32	34
Polysaccharide lyases (PL)	3	4	4
Carbohydrate esterases (CE)	32	30	32
Carbohydrate-binding modules (CBM)	15	11	10
Auxiliary activities (AA)	22	10	13
Glycosyl transferases (GT)	39	37	38

A cluster of genes predicted to be encoding the synthesis of the bromopyrrole pentabromopseudilin was found within the genome sequence of strain IA19^T. Pentabromopseudilin was isolated for the first time from *Alteromonas luteoviolacea* (Laatsch et al., 1995). It is the most active member in a group of more than 20 pyrrole antibiotics that effectively interferes with the macromolecular syntheses in Gram-positive and Gram-negative bacteria, and also has antifungal activity, as well as the activity against the biosynthesis of cholesterol. In addition, it shows pronounced *in vitro* activity against experimental leukemia and melanoma cell lines (Laatsch et al., 1995).

In the genome sequences of isolates A2-NA12 and A2-NA13, a cluster related to the synthesis of 5′-hydroxystreptomycin was predicted. This compound is an aminoglycoside antibiotic that can be biosynthesized by different *Streptomyces* species (Beyer et al., 1998).

The genome sequences of isolates IA19^T show a cluster of genes for the biosynthesis of amphotericin, a polyene macrolide produced by *Streptomyces nodosus*, which is a potent antifungal compound and also has activity against some viruses, protozoans and prions (Caffrey et al., 2001).

A cluster of genes related to that encoding the biosynthesis of streptolydigin, a tetramic acid also produced by *Streptomyces lydicus*, has been predicted in the genome sequences of strains A2-NA12 and A2-NA13. This compound is a potent antibiotic which inhibits the bacterial RNA polymerase. It can be also used in acute limphoblastic leukemia or even in acute and chronic myelocytic leukemia (Olano et al., 2009).

The genomes of strains A2-NA12 and A2-NA13 both contain a group of genes that have been related to clusters of genes from *Streptomyces*, which have been described to encode the synthesis of lankacidin, a macrocyclic antibiotic and antitumor against leukemia (Arakawa et al., 2005).

Also, in the genome sequences of strains A2-NA12 and A2-NA13, there are cluster of genes related to other genes implicated in the biosynthesis of 9-methylstreptimidone. This substance is an antifungal (Allen et al., 1976) and antiviral agent which inhibits the growth of polyvirus, vesicular stomatitis virus (VSV) and Newcastle disease virus (NDV) (Saito et al., 1974) and also inhibit NF-κB (nuclear factor-κB) (Ishikawa et al., 2009).

Clusters related to the synthesis of bacillomycin have been predicted in strains A2-NA12 and A2-NA13. This antibiotic, described in *Bacillus subtilis*, possesses antifungal activity against practically all the important dermatophytes and systemic infectious fungi (Landy et al., 1948).

A cluster related to the synthesis of meilingmycin, which possesses potent, broad-spectrum anthelmintic, insecticidal and acaricidal activities and is produced by *Streptomyces nanchangensis* (Zhuang et al., 2006), has been found in the genome sequence of the strain A2-NA13.

As shown above, the selected isolates encoded a large number of clusters involved in the synthesis of secondary metabolites.

TABLE 6 | Cluster of genes predicted to encode the synthesis of bioactive compounds in the genome sequences of the strains of this study based on the analysis of genome sequences with AntiSMASH 3.0 program.

Cluster	Type	Size (bp)	Most similar known cluster (% of similarity)	MIB – ID*
***Pseudomonas* IA19^T**				
Cluster 3	–	13,912	Bromophenols and Bromopyrroles biosynthetic genes (20%)	BGC0000891_c1
Cluster 6	–	6,919	Lipopolysaccharide biosynthetic genes (8%)	BGC0000774_c1
Cluster 11	Fatty acid	19,493	Pyoverdine biosynthetic genes (4%)	BGC0000413_c1
Cluster 12	–	19,861	Amphotericin biosynthetic genes (15%)	BGC0000015_c2
Cluster 14	–	23,611	Pyoverdine biosynthetic genes (1%)	BGC0000413_c1
Cluster 22	Saccharide	31,526	Pyoverdine biosynthetic genes (3%)	BGC0000413_c1
Cluster 24	Saccharide	40,812	Pseudopyronine A and Pseudopyronine B biosynthetic genes (62%)	BGC0001285_c1
Cluster 25	Terpene-Siderophore	43,656	Carotenoid biosynthetic genes (85%)	BGC0000642_c1
Cluster 32	Saccharide	50,218	Lipopolysaccharide biosynthetic genes (36%)	BGC0000776_c1
Cluster 37	–	17,662	PM100117 and 100118 biosynthetic genes (10%)	BGC0001359_c1
Cluster 41	–	24,157	2-amino-4-methoxy-trans-3-butenoic acid biosynthetic genes (40%)	BGC0000287_c1
Cluster 42	–	9,467	Caryoynencin biosynthetic genes (16%)	BGC0000892_c1
Cluster 47	–	13,041	Alginate biosynthetic genes (100%)	BGC0000726_c1
***Pseudomonas* A2-NA12**				
Cluster 1	Saccharide, Fatty acid	34,524	5′-hydroxystreptomycin biosynthetic genes (9%)	BGC0000690_c1
Cluster 2	–	23,963	Polysaccharide B biosynthetic gene (6%)	BGC0001411_c1
Cluster 4	–	28,529	Streptolydigin biosynthetic gene (5%)	BGC0001046_c1
Cluster 14	–	25,995	Lankacidin biosynthetic gene (13%)	BGC0001100_c1
Cluster 16	–	12,110	Alginate biosynthetic genes (100%)	BGC0000726_c1
Cluster 22	–	14,085	9-methylstreptimidone biosynthetic gene (9%)	BGC0000171_c1
Cluster 23	NRPS	52,992	Pyoverdine biosynthetic genes (22%)	BGC0000413_c1
Cluster 26	Saccharide	4,662	Lipopolysaccharide biosynthetic genes (36%)	BGC0000776_c1
Cluster 28	Bacteriocin	10,839	Bacillomycin_biosynthetic_gene_cluster (20%)	BGC0001090_c1
Cluster 34	NRPS	129,458	Pyoverdine biosynthetic gene (18%)	BGC0000413_c1
***Pseudomonas* A2-NA13**				
Cluster 3	–	10,565	Lipopolysaccharide biosynthetic genes (5%)	BGC0000774_c1
Cluster 5	–	23,657	Polysaccharide B biosynthetic genes (6%)	BGC0001411_c1
Cluster 6	Bacteriocin	10,839	Bacillomycin B biosynthetic genes (20%)	BGC0001090_c1
Cluster 7	NRPS	1,333,658	Pyoverdine biosynthetic gene (21%)	BGC0000413_c1
Cluster 13	–	18,593	9-methylstreptomidone biosynthetic gene (9%)	BGC0000171_c1
Cluster 21	–	17,344	Meilingmycin biosynthetic gene (3%)	BGC0000093_c1
Cluster 27	–	25,977	Lankacidin biosynthetic gene (13%)	BGC0001100_c1
Cluster 28	NRPS	52,992	Pyoverdine biosynthetic gene (21%)	BGC0000413_c1
Cluster 31	Saccharide	46,647	Lipopolysaccharide biosynthetic gene (36%)	BGC0000776_c1
Cluster 38	Fatty acid, Saccharide	32,344	5′-hydroxystreptomycin biosynthetic gene (9%)	BGC0000690_c1
Cluster 40	–	24,316	Streptolydigin biosynthetic gene (5%)	BGC0001046_c1

**Minimum Information about a Biosynthetic Gene Cluster repository.*

The strains encoded in their genomes 37 (A2-NA12), 40 (A2-NA13), and 49 (IA19^T) of these clusters. These values are higher than the average of clusters found in the genomes of other *Pseudomonas* strains isolated from different origins. For example, recently published genomes of *Pseudomonas* strains isolated from soil or rhizosphere encode 9 (Hennessy et al., 2015), 13 (Vida et al., 2017) or 16 (Adam et al., 2015) of these clusters; in the same way, *Pseudomonas* strains isolated from plant tissues present 4 (Wemheuer et al., 2017) or 6 (Maggini et al., 2017) of these clusters. Most of these new genomes have revealed new potentially bioactive compounds in this genus, but importantly, many of the substances predicted in the genome sequences of the strains of this study have never been identified in other *Pseudomonas* strains. Moreover, the low similarity found between some of the described predicted clusters and those of the microorganisms in the AntiSMASH database seem to indicate that several of those substances are potential new chemical compounds, which increases the interest of these bacteria as potential producers of new medical drugs. These results are in accordance with previous findings in the genomes of bacterial

symbionts in other insects (Arnam et al., 2018), demonstrating the large variety of cryptic metabolites encoded by these strains, of which only a small quantity have been characterized so far. Therefore, the genomic information and the clusters of genes predicted to be involved in secondary metabolite biosynthesis of the bacterial isolates of this study will require further analysis regarding the structure and function of the bioactive compounds encoded in their genomes.

Our results show that several *Pseudomonas* strains associated with bark beetles possess the genetic potential to produce several antimicrobial substances, as well as other chemical compounds with pharmaceutical interest. Although the bacterial strains identified as *Pseudomonas* were selected and their genome sequences studied, many other bark beetle isolates screened in this study for the capacity to produce antimicrobials seem to have great potential as producers of bioactive compounds. The genome sequencing and mining of other isolates, as well as the production, purification and identification of the predicted bioactive compounds *in silico* are necessary and could have a potential positive impact on the availability of new compounds for the development of drugs.

Regarding the taxonomic study of the strain IA19^T, the analysis of its 16S rRNA sequence supports its classification within the genus *Pseudomonas*. Nevertheless, the phylogenetic analysis of this gene sequence, as well as the sequences of the housekeeping genes, and those genes of related species within the genus, suggests that IA19^T is a new species of the genus *Pseudomonas*. Furthermore, ANI values with the type strains of its closest related species confirm that this strain belongs to a new species. On the other hand, our isolate differs from its closest related species in several phenotypic characters. Upon considering all phenotypic, chemotaxonomic and phylogenetic data of this study, the strain IA19^T appears to represent a novel species within the genus *Pseudomonas*, for which the name *Pseudomonas bohemica* sp. nov. is proposed.

Description of *Pseudomonas bohemica* sp. nov.

Pseudomonas bohemica (bo.he'mi.ca. M.L.adj. related to Bohemia, the region in the Czech Republic were the type strain was isolated). Temperature for growth ranges between 4 and 37, growth pH ranges between 6.5 and 8. Optimal growth occurs at 25°C and pH 7–8. Able to grow with up to 2.5% NaCl in TSB. The respiratory ubiquinones are Q9 (97%) and Q8 (3%). C16:0 (30.5%) and in summed feature 3 (20.1%) are the main fatty acids. Oxidase- and catalase-positive. In the API20NE system, aesculin hydrolysis, assimilation of D-glucose, L-arabinose, D-mannose, D-mannitol, potassium gluconate and trisodium citrate are positive whereas reduction of nitrates, glucose fermentation, production of gelatinase, urease, indole arginine dihydrolase and ß-galactosidase and assimilation of adonitol, methyl-xyloside, N-acetyl glucosamine, D-maltose, caprate, adipate, malate and phenylacetate are negative. In API 50CH, it produces acid from D-glucose, glycerol, ribose, L-xylose, L-fructose, D-mannose, L-arabinose, amigdalin, arbutine and salicin.

The type strain, IA19^T (=CECT 9403^T =LMG 30182^T), was isolated from a bark beetle from the species *Ips acuminatus* in the Czech Republic. The DNA G+C content of the type strain is 59.5 mol%.

AUTHOR CONTRIBUTIONS

ZS-S, AJ-G, AD-M, JI performed the experiments. MK made the sampling. RL-M, TV, ZS-S, AJ-G, PG-F performed the bioinformatic analysis of the data. PG-F and RR designed the research project. PG-F, ZS-S, RL-M, RR, EV analyzed the data. PG-F, ZS-S, AJ-G, EV interpreted the results. ZS-S and PG-F contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

This work was supported by the Czech Science Foundation (GACR) under the project number 16-15293Y. Authors thank A. Fabryová and R. Martínez for their collaboration in the strains isolation and Emma J. Keck for English language edition.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00913/full#supplementary-material

Supplementary Figure 1 | Neighbor-joining phylogenetic tree based on nearly complete (1,400 bp) 16S rRNA gene sequences of all *Pseudomonas* species closely related to *P. bohemica* IA19^T and the species *Acinetobacter baumannii* DSM30007 ^{T}, which was included as an outgroup. Bootstrap values (expressed as percentages of 1,000 replications) are shown at the branching points. Scale bar = 2 nucleotides (nt) substitutions per 100 nt. Accession numbers of the sequences are indicated in parentheses.

Supplementary Figure 2 | Neighbor-joining phylogenetic tree based on concatenated partial *gyr*B, *rpo*B, *rpo*D and gene sequences of strain *P. bohemica* IA19^T and closely related species of the genus *Pseudomonas*. Bootstrap values (expressed as percentages of 1,000 replications) are shown at the branching points. Scale bar = 2 nucleotides (nt) substitutions per 100 nt. Accession numbers of the sequences are indicated in parentheses.

REFERENCES

Adam, Z., Chen, Q., Xu, R., Diange, A. E., Bromfield, E. S. P., and Tambong, J. T. (2015). Draft genome sequence of *Pseudomonas simiae* strain 2-36, an *in vitro* antagonist of *Rhizoctonia solani* and *Gaeumannomyces graminis*. *Genome Announc.* 3, e01534–e01514. doi: 10.1128/genomeA.01534-14

Adams, A. S., Aylward, F. O., Adams, S. M., Erbilgin, N., Aukema, B. H., Currie, C. R., et al. (2013). Mountain pine beetles colonizing historical and naïve host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. *Appl. Environ. Microbiol.* 79, 3468–3475. doi: 10.1128/AEM.00068-13

Adams, A. S., Currie, C. R., Cardoza, Y., Klepzig, K. D., and Raffa, K. F. (2009). Effects of symbiotic bacteria and tree chemistry on the growth and reproduction of bark beetle fungal symbionts. *Can. J. For. Res.* 39, 1133–1147. doi: 10.1139/X09-034

Ait Tayeb, L., Ageron, E., Grimont, F., and Grimont, P. A. D. (2005). Molecular phylogeny of the genus *Pseudomonas* based on rpoB sequences and application for the identification of isolates. *Res. Microbiol.* 156, 763–773. doi: 10.1016/j.resmic.2005.02.009

Allen, M. S., Becker, A. M., and Rickards, R. W. (1976). The glutarimide antibiotic 9-methylstreptimidone: structure, biogenesis and biological activity. *Aust. J. Chem.* 29, 673–679. doi: 10.1071/ch9760673

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

Arakawa, K., Sugino, F., Kodama, K., Ishii, T., and Kinashi, H. (2005). Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* 12, 249–256. doi: 10.1016/j.chembiol.2005.01.009

Arnam, E. B. V., Currie, R. C., and Clardy, J. (2018). Defense contracts: molecular protection in insect-microbe symbioses. *Chem. Soc. Rev.* 47, 1638–1651. doi: 10.1039/C7CS00340D

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

Balamurugan, R., Dekker, F. J., and Waldmann, H. (2005). Design of compound libraries based on natural product scaffolds and protein structure similarity clustering (PSSC). *Mol. Biosyst.* 1, 36–45. doi: 10.1039/B503623B

Bauer, J. S., Ghequire, M. G. K., Nett, M., Josten, M., Sahl, H.-G., De Mot, R., et al. (2015). Biosynthetic origin of the antibiotic Pseudopyronines a and b in *Pseudomonas putida* BW11M1. *ChemBioChem* 16, 2491–2497. doi: 10.1002/cbic.201500413

Becerra, M. C., Eraso, A. J., and Albesa, I. (2003). Comparison of oxidative stress induced by ciprofloxacin and pyoverdin in bacteria and in leukocytes to evaluate toxicity. *Luminescence* 18, 334–340. doi: 10.1002/bio.742

Beyer, S., Mayer, G., and Piepersberg, W. (1998). The StrQ protein encoded in the gene cluster for 5′-hydroxystreptomycin of *Streptomyces glaucescens* GLA.0 is a α- D-glucose-1-phosphate cytidylyltransferase (CDP- D-glucose synthase). *Eur. J. Biochem.* 258, 1059–1067. doi: 10.1046/j.1432-1327.1998.2581059.x

Boone, C. K., Keefover-Ring, K., Mapes, A. C., Adams, A. S., Bohlmann, J., and Raffa, K. F. (2013). Bacteria associated with a tree-killing insect reduce concentrations of plant defense compounds. *J. Chem. Ecol.* 39, 1003–1006. doi: 10.1007/s10886-013-0313-0

Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2001). Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* 8, 713–723. doi: 10.1016/S1074-5521(01)00046-1

Cardoza, Y. J., Klepzig, K. D., and Raffa, K. F. (2006). Bacteria in oral secretions of an endophytic insect inhibit antagonistic fungi. *Ecol. Entomol.* 31, 636–645. doi: 10.1111/j.1365-2311.2006.00829.x

Clark, L. L., Dajcs, J. J., McLean, C. H., Bartell, J. G., and Stroman, D. W. (2006). *Pseudomonas otitidis* sp. nov., isolated from patients with otic infections. *Int. J. Syst. Evol. Microbiol.* 56, 709–714. doi: 10.1099/ijs.0.63753-0

Cragg, G. M., and Newman, D. J. (2009). Biodiversity: a continuing source of novel drug leads. *Pure Appl. Chem.* 77, 7–24. doi: 10.1351/pac200577010007

Dias, D. A., Urban, S., and Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites* 2, 303–336. doi: 10.3390/metabo2020303

Doetsch, R. N. (1981). "Determinative methods of light microscopy," in *Manual of Methods for General Bacteriology*, eds P. Gerdhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg, and G. B. Phillips (Washington, DC: American Society for Microbiology), 21–33.

Drewry, D. H., and Macarron, R. (2010). Enhancements of screening collections to address areas of unmet medical need: an industry perspective. *Curr. Opin. Chem. Biol.* 14, 289–298. doi: 10.1016/j.cbpa.2010.03.024

Fabryová, A., Kostovčík, M., Díez-Méndez, A., Jiménez-Gómez, A., Celador-Lera, L., Saati-Santamaría, Z., et al. (2018). On the bright side of a forest pest-the metabolic potential of bark beetles' bacterial associates. *Sci. Total Environ.* 619–620, 9–17. doi: 10.1016/j.scitotenv.2017.11.074

Flórez, L. V., Scherlach, K., Gaube, P., Ross, C., Sitte, E., Hermes, C., et al. (2017). Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nat. Commun.* 8:15172. doi: 10.1038/ncomms15172

Frasson, D., Opoku, M., Picozzi, T., Torossi, T., Balada, S., Smits, T. H. M., et al. (2017). *Pseudomonas wadenswilerensis* sp. nov. and *Pseudomonas reidholzensis* sp. nov., two novel species within the *Pseudomonas putida* group isolated from forest soil. *Int. J. Syst. Evol. Microbiol.* 67, 2853–2861. doi: 10.1099/ijsem.0.002035

Ganne, G., Brillet, K., Basta, B., Roche, B., Hoegy, F., Gasser, V., et al. (2017). Iron release from the siderophore pyoverdine in *Pseudomonas* aeruginosa involves three new actors: FpvC, FpvG, and FpvH. *ACS Chem. Biol.* 12, 1056–1065. doi: 10.1021/acschembio.6b01077

García-Fraile, P. (2018). Roles of bacteria in the bark beetle holobiont – how do they shape this forest pest? *Ann. Appl. Biol.* 172, 111–125. doi: 10.1111/aab.12406

García-Fraile, P., Chudíčková, M., Benada, O., Pikula, J., and Kolařík, M. (2015). *Serratia myotis sp*. nov. and *Serratia vespertilionis* sp. nov., isolated from bats hibernating in caves. *Int. J. Syst. Evol. Microbiol.* 65, 90–94. doi: 10.1099/ijs.0.066407-0

Haas, D., and Keel, C. (2003). Regulation of antibiotic production in root-colonizing pseudomonas spp. and relevance for biological control of plant disease. *Annu. Rev. Phytopathol.* 41, 117–153. doi: 10.1146/annurev.phyto.41.052002.095656

Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* 14, 111–129. doi: 10.1038/nrd4510

Hennessy, R. C., Glaring, M. A., Michelsen, C. F., Olsson, S., and Stougaard, P. (2015). Draft genome sequence of pseudomonas sp. strain in5 isolated from a greenlandic disease suppressive soil with potent antimicrobial activity. *Genome Announc.* 3:e01251–15. doi: 10.1128/genomeA.01251-15

Hoffman, S. J., Outterson, K., Røttingen, J.-A., Cars, O., Clift, C., Rizvi, Z., et al. (2015). An international legal framework to address antimicrobial resistance. *Bull. World Health Organ.* 93, 66–66. doi: 10.2471/BLT.15.152710

Hu, X., Yu, J., Wang, C., and Chen, H. (2014). Cellulolytic bacteria associated with the gut of dendroctonus armandi larvae (coleoptera: curculionidae: *scolytinae*). *Forests* 5, 455–465. doi: 10.3390/f5030455

Ishikawa, Y., Tachibana, M., Matsui, C., Obata, R., Umezawa, K., and Nishiyama, S. (2009). Synthesis and biological evaluation on novel analogs of 9-methylstreptimidone, an inhibitor of NF-KB. *Bioorg. Med. Chem. Lett.* 19, 1726–1728. doi: 10.1016/j.bmcl.2009.01.107

Isnansetyo, A., and Kamei, Y. (2009). Bioactive substances produced by marine isolates of Pseudomonas. *J. Ind. Microbiol. Biotechnol.* 36, 1239–1248. doi: 10.1007/s10295-009-0611-2

Kennedy, J., Marchesi, J. R., and Dobson, A. D. W. (2007). Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl. Microbiol. Biotechnol.* 75, 11–20. doi: 10.1007/s00253-007-0875-2

Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., et al. (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* 62, 716–721. doi: 10.1099/ijs.0.038075-0

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Kinzel, O., and Budzikiewicz, H. (1999). Synthesis and biological evaluation of a pyoverdin-β-lactam conjugate: a new type of arginine-specific cross-linking in aqueous solution. *J. Pept. Res.* 53, 618–625. doi: 10.1034/j.1399-3011.1999.00053.x

Kinzel, O., Tappe, R., Gerus, I., and Budzikiewicz, H. (1998). The synthesis and antibacterial activity of two pyoverdin-ampicillin conjugates, entering *Pseudomonas aeruginosa* via the pyoverdin-mediated iron uptake pathway. *J. Antibiot* 51, 499–507. doi: 10.7164/antibiotics.51.499

Kovacs, N. (1956). Identification of Pseudomonas pyocyanea by the oxidase reaction. *Nature* 178, 703. doi: 10.1038/178703a0

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

Laatsch, H., Renneberg, B., Hanefeld, U., Kellner, M., Pudleiner, H., Hamprecht, G., et al. (1995). Structure-activity relationships of phenyl- and benzoylpyrroles. *Chem. Pharm. Bull.* 43, 537–546. doi: 10.1248/cpb.43.537

Laine, M. H., Karwoski, M. T., Raaska, L. B., and Mattila-Sandholm, T.-M. (1996). Antimicrobial activity of Pseudomonas spp. against food poisoning bacteria and moulds. *Lett. Appl. Microbiol.* 22, 214–218. doi: 10.1111/j.1472-765X.1996.tb01146.x

Landy, M., Warren, G. H., Rosenman, M., S. B., and Colio, L. G. (1948). Bacillomycin: an antibiotic from *Bacillus subtilis* active against pathogenic fungi*. *Proc. Soc. Exp. Biol. Med.* 67, 539–541. doi: 10.3181/00379727-67-16367

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lee, X., Azevedo, M. D., Armstrong, D. J., Banowetz, G. M., and Reimmann, C. (2013). The *Pseudomonas aeruginosa* antimetabolite L-2-amino-4-methoxy-trans-3-butenoic acid inhibits growth of Erwinia amylovora and acts as a seed germination-arrest factor. *Environ. Microbiol. Rep.* 5, 83–89. doi: 10.1111/j.1758-2229.2012.00395.x

Lee, X., Fox, Á., Sufrin, J., Henry, H., Majcherczyk, P., Haas, D., et al. (2010). Identification of the biosynthetic gene cluster for the pseudomonas aeruginosa antimetabolite l-2-amino-4-methoxy-trans-3-butenoic acid. *J. Bacteriol.* 192, 4251–4255. doi: 10.1128/JB.00492-10

Lee, X., Reimmann, C., Greub, G., Sufrin, J., and Croxatto, A. (2012). The *Pseudomonas aeruginosa* toxin L-2-amino-4-methoxy-trans-3-butenoic acid inhibits growth and induces encystment in *Acanthamoeba* castellanii. *Microbes Infect.* 14, 268–272. doi: 10.1016/j.micinf.2011.10.004

Levy, S. B., and Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med. N. Y.* 10(Suppl. 1), S122–S129. doi: 10.1038/nm1145

Maggini, V., Presta, L., Miceli, E., Fondi, M., Bosi, E., Chiellini, C., et al. (2017). Draft genome sequence of pseudomonas sp. strain ep r1 isolated from echinacea purpurea roots and effective in the growth inhibition of human opportunistic pathogens belonging to the burkholderia cepacia complex. *Genome Announc.* 5:e00351-17. doi: 10.1128/genomeA.00351-17

Marinho, P. R., Moreira, A. P., Pellegrino, F. L., Muricy, G., Bastos Mdo C., Santos, K. R., et al. (2009). Marine *Pseudomonas putida*: a potential source of antimicrobial substances against antibiotic-resistant bacteria. *Mem. Inst. Oswaldo Cruz* 104, 678–682. doi: 10.1590/S0074-02762009000500002

Menéndez, E., Ramírez-Bahena, M. H., Fabryová, A., Igual, J. M., Benada, O., Mateos, P. F., et al. (2015). *Pseudomonas coleopterorum sp*. nov., a cellulase-producing bacterium isolated from the bark beetle Hylesinus fraxini. *Int. J. Syst. Evol. Microbiol.* 65, 2852–2858. doi: 10.1099/ijs.0.000344

Mishra, B. B., and Tiwari, V. K. (2011). Natural products: an evolving role in future drug discovery. *Eur. J. Med. Chem.* 46, 4769–4807. doi: 10.1016/j.ejmech.2011.07.057

Morales-Jiménez, J., Zúñiga, G., Ramírez-Saad, H. C., and Hernández-Rodríguez, C. (2012). Gut-associated bacteria throughout the life cycle of the bark beetle *Dendroctonus rhizophagus* thomas and bright (*Curculionidae*: scolytinae) and their cellulolytic activities. *Microb. Ecol.* 64, 268–278. doi: 10.1007/s00248-011-9999-0

Mulet, M., Bennasar, A., Lalucat, J., and García-Valdés, E. (2009). An rpoD-based PCR procedure for the identification of Pseudomonas species and for their detection in environmental samples. *Mol. Cell. Probes* 23, 140–147. doi: 10.1016/j.mcp.2009.02.001

Mulet, M., Gomila, M., Lemaitre, B., Lalucat, J., and García-Valdés, E. (2012). Taxonomic characterisation of Pseudomonas strain L48 and formal proposal of *Pseudomonas entomophila* sp. nov. *Syst. Appl. Microbiol.* 35, 145–149. doi: 10.1016/j.syapm.2011.12.003

Mulet, M., Lalucat, J., and García-Valdés, E. (2010). DNA sequence-based analysis of the Pseudomonas species. *Environ. Microbiol.* 12, 1513–1530. doi: 10.1111/j.1462-2920.2010.02181.x

Mulet, M., Sánchez, D., Lalucat, J., Lee, K., and García-Valdés, E. (2015). *Pseudomonas alkylphenolica sp*. nov. a bacterial species able to form special aerial structures when grown on p-cresol. *Int. J. Syst. Evol. Microbiol.* 65, 4013–4018. doi: 10.1099/ijsem.0.000529

Nishanth Kumar, S., Aravind, S. R., Jacob, J., Gopinath, G., Lankalapalli, R. S., Sreelekha, T. T., et al. (2016). Pseudopyronine B: A potent antimicrobial and anticancer molecule isolated from a *pseudomonas mosselii*. *Front. Microbiol.* 7:1307. doi: 10.3389/fmicb.2016.01307

Olano, C., Gómez, C., Pérez, M., Palomino, M., Pineda-Lucena, A., Carbajo, R. J., et al. (2009). Deciphering biosynthesis of the RNA polymerase inhibitor streptolydigin and generation of glycosylated derivatives. *Chem. Biol.* 16, 1031–1044. doi: 10.1016/j.chembiol.2009.09.015

Olorunleke, F. E., Kieu Phuong, N., and Höfte, M. (2015). "Recent advances in Pseudomonas biocontrol," in *Bacteria-Plant Interactions: Advanced Research and Future Trends*, eds J. Murillo, B. A. Vinatzer, R. W. Jackson, and D. L. Arnold (Poole: Caister Academic Press), 167–198.

Onsoyen, E., Myrvold, R., Dessen, A., Thomas, D., and Walsh, T. R. (2010). *Treatment of Acinetobacter with Alginate Oligomers and Antibiotics*. Patent number: WO2010139956.

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

Oyaizu, H., and Komagata, K. (1983). Grouping of Pseudomonas species on the basis of cellular fatty acid composition and the quinone system with special reference to the existence of 3-hydroxy fatty acids. *J. Gen. Appl. Microbiol.* 29, 17–40. doi: 10.2323/jgam.29.17

Palleroni, N. (2005). Genus I Pseudomonas Migula 1894, 237. *Bergeys Man. Syst. Bacteriol.* 2, 323–378.

Peix, A., Ramírez-Bahena, M.-H., and Velázquez, E. (2018). The current status on the taxonomy of Pseudomonas revisited: an update. *Infect. Genet. Evol.* 57, 106–116. doi: 10.1016/j.meegid.2017.10.026

Pérez, M., Schleissner, C., Fernández, R., Rodríguez, P., Reyes, F., Zuñiga, P., et al. (2016). PM100117 and PM100118, new antitumor macrolides produced by a marine Streptomyces caniferus GUA-06-05-006A. *J. Antibiot.* 69, 388–394. doi: 10.1038/ja.2015.121

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

Pierson, L. S., Keppenne, V. D., and Wood, D. W. (1994). Phenazine antibiotic biosynthesis in *Pseudomonas aureofaciens* 30-84 is regulated by PhzR in response to cell density. *J. Bacteriol.* 176, 3966–3974. doi: 10.1128/jb.176.13.3966-3974.1994

Premanandh, J., Samara, B. S., and Mazen, A. N. (2016). Race against antimicrobial resistance requires coordinated action – an overview. *Front. Microbiol.* 6:1536. doi: 10.3389/fmicb.2015.01536

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.

Ramírez-Bahena, M.-H., Cuesta, M. J., Flores-Félix, J. D., Mulas, R., Rivas, R., Castro-Pinto, J., et al. (2014). *Pseudomonas helmanticensis sp*. nov., isolated from forest soil. *Int. J. Syst. Evol. Microbiol.* 64, 2338–2345. doi: 10.1099/ijs.0.063560-0

Ramos, E., Ramírez-Bahena, M.-H., Valverde, A., Velázquez, E., Zúñiga, D., Velezmoro, C., et al. (2013). *Pseudomonas punonensis* sp. nov., isolated from straw. *Int. J. Syst. Evol. Microbiol.* 63, 1834–1839. doi: 10.1099/ijs.0.042119-0

Ravelo, A. G., and Braun, A. E. (2009). Relevancia de los productos naturales en el descubrimiento de nuevos fármacos en el s. XXI. *Rev. Real Acad. Cienc. Exactas Físicas Nat.* 103, 409–420.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rivas, R., García-Fraile, P., Mateos, P., f., Martínez-Molina, E., and Velázquez, E. (2007). Characterization of xylanolytic bacteria present in the bract phyllosphere of the date palm Phoenix dactylifera. *Lett. Appl. Microbiol.* 44, 181–187. doi: 10.1111/j.1472-765X.2006.02050.x

Rogers, J. S., Swofford, D. L., and Cannatella, D. (1998). A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. *Syst. Biol.* 47, 77–89. doi: 10.1080/106351598261049

Saito, N., Kitame, F., Kikuchi, M., and Ishida, N. (1974). Studies on a new antiviral antibiotic, 9-methylstreptimidone. I. Physicochemical and biological properties. *J. Antibiot.* 27, 206–214. doi: 10.7164/antibiotics.27.206

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

Sarniguet, A., Kraus, J., Henkels, M. D., Muehlchen, A. M., and Loper, J. E. (1995). The sigma factor sigma s affects antibiotic production and biological control

activity of *Pseudomonas fluorescens* Pf-5. *Proc. Natl. Acad. Sci.* 92, 12255–12259. doi: 10.1073/pnas.92.26.12255

Sasser, M. (1990). *Identification of Bacteria by Gas Chromatography of Cellular Fatty Acids, MIDI Technical Note 101*. Newark, DE: MIDI Inc.

Schnider, U., Keel, C., Blumer, C., Troxler, J., Défago, G., and Haas, D. (1995). Amplification of the housekeeping sigma factor in *Pseudomonas fluorescens* CHA0 enhances antibiotic production and improves biocontrol abilities. *J. Bacteriol.* 177, 5387–5392. doi: 10.1128/jb.177.18.5387-5392.1995

Sedkova, N., Tao, L., Rouvière, P. E., and Cheng, Q. (2005). Diversity of carotenoid synthesis gene clusters from environmental enterobacteriaceae strains. *Appl. Environ. Microbiol.* 71, 8141–8146. doi: 10.1128/AEM.71.12.8141-8146.2005

Six, D. L. (2012). Ecological and evolutionary determinants of bark beetle —fungus symbioses. *Insects* 3, 339–366. doi: 10.3390/insects3010339

Stintzi, A., Cornelis, P., Hohnadel, D., Meyer, J.-M., Dean, C., Poole, K., et al. (1996). Novel pyoverdine biosynthesis gene(s) of *Pseudomonas aeruginosa* PAO. *Microbiology* 142, 1181–1190. doi: 10.1099/13500872-142-5-1181

Therrien, J., Mason, C. J., Cale, J. A., Adams, A., Aukema, B. H., Currie, C. R., et al. (2015). Bacteria influence mountain pine beetle brood development through interactions with symbiotic and antagonistic fungi: implications for climate-driven host range expansion. *Oecologia* 179, 467–485. doi: 10.1007/s00442-015-3356-9

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Tindall, B. J. (1990a). A comparative study of the lipid composition of halobacterium saccharovorum from various sources. *Syst. Appl. Microbiol.* 13, 128–130. doi: 10.1016/S0723-2020(11)80158-X

Tindall, B. J. (1990b). Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiol. Lett.* 66, 199–202. doi: 10.1111/j.1574-6968.1990.tb03996.x

Tisdale, M. J. (1980). The effect of the methionine antagonist l-2-amino-4-methoxy-trans-3-butenoic acid on the growth and metabolism of walker carcinosarcoma *in vitro*. *Biochem. Pharmacol.* 29, 501–508. doi: 10.1016/0006-2952(80)90369-X

Toro, M., Ramírez-Bahena, M.-H., Cuesta, M. J., Velázquez, E., and Peix, A. (2013). *Pseudomonas guariconensis* sp. nov., isolated from rhizospheric soil. *Int. J. Syst. Evol. Microbiol.* 63, 4413–4420. doi: 10.1099/ijs.0.051193-0

Verani, A., Sironi, F., Siccardi, A. G., Lusso, P., and Vercelli, D. (2002). Inhibition of CXCR4-tropic HIV-1 infection by lipopolysaccharide: evidence of different mechanisms in macrophages and T Lymphocytes. *J. Immunol.* 168, 6388–6395. doi: 10.4049/jimmunol.168.12.6388

Vida, C., de Vicente, A., and Cazorla, F. M. (2017). Draft genome sequence of the rhizobacterium *Pseudomonas chlororaphis* PCL1601, displaying biocontrol against soilborne phytopathogens. *Genome Announc.* 5, e00130–e00117. doi: 10.1128/genomeA.00130-17

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

Wemheuer, F., Hollensteiner, J., Poehlein, A., Granzow, S., Daniel, R., Vidal, S., et al. (2017). Draft genome sequence of pseudomonas putida strain GM4FR, an endophytic bacterium isolated from *Festuca rubra* L. *Genome Announc.* 5, e00086–e00017. doi: 10.1128/genomeA.00086-17

Xiao, Y.-P., Hui, W., Wang, Q., Roh, S. W., Shi, X.-Q., Shi, J.-H., et al. (2009). *Pseudomonas caeni* sp. nov., a denitrifying bacterium isolated from the sludge of an anaerobic ammonium-oxidizing bioreactor. *Int. J. Syst. Evol. Microbiol.* 59, 2594–2598. doi: 10.1099/ijs.0.005108-0

Xu, L., Lou, Q., Cheng, C., Lu, M., and Sun, J. (2015). Gut-associated bacteria of dendroctonus valens and their involvement in verbenone production. *Microb. Ecol.* 70, 1012–1023. doi: 10.1007/s00248-015-0625-4

Xu, L.-T., Lu, M., and Sun, J.-H. (2016). Invasive bark beetle-associated microbes degrade a host defensive monoterpene. *Insect Sci.* 23, 183–190. doi: 10.1111/1744-7917.12255

Yamaguchi, M., Park, H. J., Ishizuka, S., Omata, K., and Hirama, M. (1995). Chemistry and antimicrobial activity of caryoynencins analogs. *J. Med. Chem.* 38, 5015–5022. doi: 10.1021/jm00026a008

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

Zhuang, Y. P., Chen, B., Chu, J., and Zhang, S. (2006). Medium optimization for meilingmycin production by *Streptomyces nanchangensis* using response surface methodology. *Process Biochem.* 41, 405–409. doi: 10.1016/j.procbio.2005.07.004

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 Saati-Santamaría, López-Mondéjar, Jiménez-Gómez, Díez-Méndez, Větrovský, Igual, Velázquez, Kolarik, Rivas and García-Fraile. 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.2018.00369

Comparative Genomics and Biosynthetic Potential Analysis of Two Lichen-Isolated *Amycolatopsis* Strains

*Marina Sánchez-Hidalgo, Ignacio González, Cristian Díaz-Muñoz, Germán Martínez and Olga Genilloud**

Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores de Andalucía, Granada, Spain

Actinomycetes have been extensively exploited as one of the most prolific secondary metabolite-producer sources and continue to be in the focus of interest in the constant search of novel bioactive compounds. The availability of less expensive next generation genome sequencing techniques has not only confirmed the extraordinary richness and broad distribution of silent natural product biosynthetic gene clusters among these bacterial genomes, but also has allowed the incorporation of genomics in bacterial taxonomy and systematics. As part of our efforts to isolate novel strains from unique environments, we explored lichen-associated microbial communities as unique assemblages to be studied as potential sources of novel bioactive natural products with application in biotechnology and drug discovery. In this work, we have studied the whole genome sequences of two new *Amycolatopsis* strains (CA-126428 and CA-128772) isolated from tropical lichens, and performed a comparative genomic analysis with 41 publicly available *Amycolatopsis* genomes. This work has not only permitted to infer and discuss their taxonomic position on the basis of the different phylogenetic approaches used, but has also allowed to assess the richness and uniqueness of the biosynthetic pathways associated to primary and secondary metabolism, and to provide a first insight on the potential role of these bacteria in the lichen-associated microbial community.

Keywords: *Amycolatopsis*, secondary metabolites, phylogeny, whole genome sequence, biosynthetic gene clusters

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Martha E. Trujillo, Universidad de Salamanca, Spain
Henk Bolhuis, Royal Netherlands Institute for Sea Research (NIOZ), Netherlands

****Correspondence:***
Olga Genilloud
olga.genilloud@medinaandalucia.es

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

Received: *09 October 2017*
Accepted: *16 February 2018*
Published: *13 March 2018*

Citation:
Sánchez-Hidalgo M, González I, Díaz-Muñoz C, Martínez G and Genilloud O (2018) Comparative Genomics and Biosynthetic Potential Analysis of Two Lichen-Isolated Amycolatopsis Strains. Front. Microbiol. 9:369. doi: 10.3389/fmicb.2018.00369

INTRODUCTION

The class *Actinobacteria* was defined for Gram-positive bacteria with high genomic G+C content (over 55%), among which are included major families of actinomycetes that produce almost 75% of all known secondary metabolites, many of them of high relevance for human health and biotechnology industry (Barka et al., 2016). These compounds have been shown to include a wide range of industrial and medical applications, as drugs (i.e., antifungals, antibacterials, antitumorals, or immunosuppresors), herbicides and plant growth promoting agents among others (Genilloud et al., 2011; Sharma et al., 2014; Genilloud, 2017a). Within the actinomycetes, the genus *Streptomyces* is the most prolific and most studied producer of secondary metabolites, but members of the families *Pseudonocardiaceae* and *Micromonosporaceae* have also shown to produce a broad diversity of bioactive molecules. Among *Pseudonocardiaceae*, members of the genus *Amycolatopsis*

produce several relevant secondary metabolites, such as balhimycin, vancomycin, avoparcin, ristomycin, chelocardin, chloroeremomycin, ECO-0501 and rifamycin (Chen et al., 2016; Kumari et al., 2016). More recently, other antibiotics have been described from some *Amycolatopsis* strains such as macrotermycins A–D (Beemelmanns et al., 2017), pargamicins B–D (Hashizume et al., 2017) and rifamorpholines A–E (Xiao et al., 2017). In addition, the importance of *Amycolatopsis* strains in industrial processes such as bioremediation (heavy metal immobilization, herbicide and polymers biodegradation) and bioconversion (wuxistatin and vanillin production) (Dávila Costa and Amoroso, 2014), has been clearly demonstrated.

Members of the genus *Amycolatopsis* were originally misidentified as *Streptomyces*, then as *Nocardia* to be finally recognized as belonging to a new genus with species lacking mycolic acid in their cell wall (Lechevalier et al., 1986). A total of 70 *Amycolatopsis* species have been recognized so far (http://www.bacterio.net/amycolatopsis.html) and isolated from a broad diversity of environments ranging from soils, plants, and ocean sediments to clinical sources. Most species of *Amycolatopsis* belong to two major subclades: *A. methanolica* (AMS) and *A. orientalis* (AOS) (Tang et al., 2016). Several AOS species have been shown to synthesize antibiotics, while AMS strains show a biotechnological potential for the overproduction of aromatic amino acids and bioremediation (Tang et al., 2016).

Recent advances in the field of next generation sequencing (NGS), have allowed an exponential increase of bacterial genome sequences available in public databases, being the genomes of *Streptomyces* spp. the most intensively studied (Loman and Pallen, 2015; Chen et al., 2016; Kumari et al., 2016). In the case of *Amycolatopsis,* as of June 2017, 54 sequencing genome projects had been assembled (www.ncbi.nlm.nih.gov/assembly) from which 30 belong to type strains. The 9 genome projects that have been completely sequenced have shown that *Amycolatopsis* strains contain comparatively large genomes (from 5 to 10 Mb) in the form of a circular chromosome. The analysis of these genomes with improved algorithms has revealed that *Amycolatopsis* strains harbor many more cryptic biosynthetic gene clusters (BGCs) than previously estimated. Eleven BGCs from *Amycolatopsis* have been characterized and compiled in the MIBiG (Minimum Information about a Biosynthetic Gene cluster) Repository (**Table 1**) (http://mibig.secondarymetabolites.org, Medema et al., 2015). Moreover, this increased availability of new genome sequences in public databases is allowing a deeper characterization of microbial biosynthetic potential using genomic data as well as a new standardization of new taxonomic approaches based on full genome sequence information (Colston et al., 2014).

The emergence of antimicrobial resistance against frequently used antibiotics has brought to light the urgent need of novel antibacterial compounds, and the necessity to look for alternative isolation sources and new drug discovery strategies to identify novel chemical classes of compounds (Adamek et al., 2017). As part of our integrated antibiotic discovery programs, we selected two bioactive *Amycolatopsis* strains (CA-128772 and CA-126428) previously isolated in our laboratory from lichens collected respectively in tropical areas from Hawaii and Reunion islands (González et al., 2005). Lichens are symbiotic associations of a fungal mycobiont, one or more algal or cyanobacterial photobionts and a diverse community of associated microbes, and represent important sources of natural products, mostly produced by the mycobiont. The *Alphaproteobacteria* are the predominant lichen-associated bacteria, but *Actinobacteria, Firmicutes, Betaproteobacteria, Deltaproteobacteria*, and *Gammaproteobacteria* have been also identified (Cardinale et al., 2006; Aschenbrenner et al., 2016). These lichen-associated bacterial communities have been suggested to play important roles in the symbiosis, being of special interest the populations of the orders *Burkholderiales* and *Actinomycetales*, well known as prolific secondary metabolite producers (Calcott et al., 2017). Recent studies have confirmed that lichen-associated bacteria produce new bioactive substances, especially among *Streptomyces* species (González et al., 2005; Cardinale et al., 2006; Parrot et al., 2015; Calcott et al., 2017; Liu et al., 2017). However, few actinomycetes belonging to the genus *Amycolatopsis* have been isolated so far from these sources (González et al., 2005; Liu et al., 2017), what triggered our interest to study their biosynthetic potential and suggest a role in this unique environment.

In this work, we have established the taxonomic position and mined the draft genomes of two new strains of *Amycolatopsis,* CA-126428 and CA-128772. The study has permitted to perform a comparative analysis with all publicly available complete or draft *Amycolatopsis* genomes, with a specific focus on the richness and diversity of their BGCs.

MATERIALS AND METHODS

DNA Extraction and Whole Genome Next Generation Sequencing

Genomic DNAs from strains CA-126428 and CA-128772 were extracted and purified as previously described (Kieser et al., 2000) from strains grown in ATCC-2 liquid medium [0.5% yeast extract (Difco, Franklin Lakes, NJ, USA), 0.3% beef extract (Difco), 0.5% peptone (Difco), 0.1% dextrose (Difco), 0.2% starch from potato (Panreac, Barcelona, Spain), 0.1% CaCO3 (E. Merck, Darmstadt, Germany), and 0.5% NZ amine E (Sigma, St Louis, MO, USA)].

Genomes of strains CA-126428 and CA-128772 were sequenced *de novo* by Macrogen (Seoul, Korea; http://www.macrogen.com/) and Service XS (Leiden, the Netherlands; http://www.servicexs.com), respectively, using the Illumina HiSeq 2500 platform. Paired-end libraries were created using the NEBNext Ultra DNA library prep kit (New England Biolabs). The quality and yield after sample preparation was measured with the Fragment Analyzer (AATI), and the size of the resulting product was consistent with the expected size of 500–700 bp. Clustering and DNA sequencing was performed according to manufacturer's protocol. A concentration of 15.0 pM DNA was used. Image analysis, base calling and quality check was performed with the Illumina data analysis pipeline RTA v1.18.64 and Bcl2fastq v2.17. A dataset of at least 1.3 Gb per sample was delivered. Prior to assembly, the reads were trimmed for adapter sequences and filtered for sequence quality. Presumed adapter

TABLE 1 | Biosynthetic gene clusters from *Amycolatopsis* strains described in the MIBiG database.

Organism	Main product	Biosynthetic class	MIBiG accession	Full/partial MIBiG info
A. alba DSM 44262^T	Albachelin	NRP	BGC0001211	Partial
A. balhimycina DSM 44591^T	Balhimycin	NRP	BGC0000311	Partial
A. japonica DSM 44213^T	Ristomycin A	NRP/Saccharide	BGC0000419	Full
A. lurida DSM 43134^T	Ristocetin	NRP	BGC0000418	Partial
A. mediterranei S699	Rifamycin	Polyketide	BGC0000136	Partial
A. orientalis DSM 40040^T	Quartromicin	Polyketide	BGC0000133	Partial
A. orientalis A82846	Chloroeremomycin	NRP	BGC0000322	Partial
A. orientalis HCCB10007	Vancomycin	NRP	BGC0000455	Partial
A. orientalis subsp. *vinearia* BA-07585	BE-7585A	Polyketide	BGC0000203	Partial
Amycolatopsis sp. SANK 60206	A-102395	Other	BGC0001191	Partial
A. sulphurea DSM 46092^T	Chelocardin	Polyketide	BGC0000208	Partial

sequences were removed from the read when the bases matched a sequence in the adapter sequence set (TruSeq adapters) with 2 or less mismatches and an alignment score of at least 12. Bases with phred scores below Q22 were removed from the reads. Glimmer v3.2 (Aggarwal and Ramaswamy, 2002) was used for annotation.

16S Ribosomal RNA (rRNA) Gene Amplification and Sequencing

Since the sequences of the 16S rRNA genes were incomplete in the genomes, PCR primers FD1 and RP2 were used to amplify the nearly full-length 16S rRNA genes of the strains CA-126428 and CA-128772 (Weisburg et al., 1991). PCR products were sequenced by Secugen (Madrid, Spain; http://www.secugen.es/) with primers FD1, RP2, 1100R, and 926F (Lane, 1991). Partial sequences were assembled and edited using the Assembler contig editor component of Bionumerics 5.10 analysis software (Applied Maths NV, Sint-Martens-Latem, Belgium).

The identification of the closest match sequences was performed at the EzBiocloud server (http://www.ezbiocloud.net/identify) (Yoon et al., 2017).

Strains and Sequences

Strains CA-128772 and CA-126428 belong to MEDINA's microbial collection and were previously isolated from arboricolous lichens collected in humid tropical forests from Hawaii and dry tropical forests from Reunion islands, respectively, as previously described (González et al., 2005). In brief, each lichen sample (300–500 mg) was washed twice with sterile water and homogenized in 30 ml of sterile water using a blender. Serial dilutions were plated on selective actinomycete isolation media (González et al., 2005). Individual colonies were isolated and grown at 28°C on YME agar medium (0.4% yeast extract, 1% malt extract, 0.4% glucose and 0.2% Bacto-agar).

The complete 16S rRNA sequences of 70 *Amycolatopsis* type strains were downloaded from the List of Prokaryotic Names with Standing in Nomenclature (LPSN) database (http://www.bacterio.net/amycolatopsis.html) (Supplementary Table 1). The sequences ranged from 1,351 to 1,530 bp. A set of publicly available 41 complete and draft genomes from *Amycolatopsis* strains and the 16S rRNA gene sequences of *Amycolatopsis* non-type strains were downloaded from NCBI (https://www.ncbi.nlm.nih.gov) (Supplementary Table 1).

Phylogenetic Analysis

16S rRNA Analysis

The 16S rRNA sequences were aligned with the MEGA 7.0.26 package (http://www.megasoftware.net) (Kumar et al., 2016) using Clustawl. The pairwise deletion Neighbor-Joining method of the MEGA 7.0.26 package (http://www.megasoftware.net) (Kumar et al., 2016), corrected with Jukes-Cantor algorithms and a bootstrap of 1,000 replicates, was used to construct a phylogenetic tree based on the complete 16S rRNA sequences. *Micromonospora chalcea* DMS 43026^T (Foulerton, 1905) 16S rRNA gene was used as outgroup. Genomic distances were calculated with the Kimura-2 parameter model included also in the MEGA software.

Multi-Locus Sequence Analysis (MLSA)

Four single-copy housekeeping genes present in all the genomes were used for MLSA analysis: *atpD* (ATP synthase F1, beta subunit), *dnaK* (Hsp70 chaperone), *recA* (recombinase A) and *rpoB* (RNA polymerase beta subunit) (Sentausa and Fournier, 2013; Glaeser and Kämpfer, 2015). The sequences of each of these genes were extracted from the *Amycolatopsis* genomes using the Multi-tBlastN algorithm (https://blast.ncbi.nlm.nih.gov) and were subsequently concatenated with Geneious 9.1.8 software (Biomatters, www.geneious.com, (Kearse et al., 2012), generating a sequence of approximately 7.8 Kb. A phylogenetic tree was constructed using the MEGA 7.0.26. pairwise deletion Neighbor-Joining and Jukes-Cantor algorithms methods and a bootstrap of 1,000 replicates (Kumar et al., 2016). *Micromonospora chalcea* DMS 43026^T (Foulerton, 1905) concatenated housekeeping genes were used as outgroup. Genomic distances were calculated with the Kimura-2 parameter model included also in the MEGA software.

Primary Metabolism Analysis

The genomes were functionally characterized in the KEGG Database (http://www.kegg.jp) (Kanehisa et al., 2016a). Ortholog

K numbers were assigned by the BlastKOALA sequence similarity tool (http://www.kegg.jp/blastkoala/) (Kanehisa et al., 2016b). KEGG mapper tool (http://www.kegg.jp/kegg/tool/map_pathway.html) was used to reconstruct and compare the metabolic pathways.

Secondary Metabolite Pathways Analysis

The presence of BGCs in the genomes was analyzed using antiSMASH 4.0, (http://antismash.secondarymetabolites.org) (Blin et al., 2017) including Clusterfinder with a minimum probability of 60%.

Genome Comparisons

Genome comparisons were performed on both complete and draft genomes. The contigs of draft genomes were concatenated with Geneious 9.1.8 software (Biomatters, www.geneious.com), (Kearse et al., 2012) to facilitate analysis.

Gene synteny analysis was assessed with progressiveMauve (http://darlinglab.org/mauve/), (Darling et al., 2010). To compare and visualize genomic regions from Mauve results we used the R-package genoPlotR (http://genoplotr.r-forge.r-project.org/) (Guy et al., 2011).

The genome sequence similarity between *Amycolatopsis* strains was evaluated with the Genome-to-Genome Distance Calculator (GGDC) 2.1 online software (http://ggdc.dsmz.de), (Auch et al., 2010; Meier-Kolthoff et al., 2013, 2014).

The Average Nucleotide Identity (ANI) and the orthology of genome sequences were calculated with OrthoANI, (http://www.ezbiocloud.net/sw/oat), (Lee et al., 2016).

RESULTS AND DISCUSSION

Whole Genome Sequencing

The genomes of strains CA-126428 and CA-128772 were sequenced, assembled and annotated by external providers and their characteristics are summarized in **Table 2**. A total of 2 × 10,113,591 paired reads for CA-128772 and 2 × 6,420,119 paired reads for CA-126428 were generated with theoretical coverage >100 ×. All reads were assembled to a draft genome of 10.5 Mb for CA-126428 and 10.2 Mb for CA-128772, with a G+C content of 71.4 and 71.8%, respectively. These G+C content values are similar to those of some *Amycolatopsis* complete genomes, such as the strains of *A. mediterranei* U32 and S699 (Kumari et al., 2016). A total of 188 contigs and 9626 coding DNA sequences (CDSs) were obtained for CA-126428, while 121 contigs and 10986 CDSs were assembled for CA-128772. The inconsistency between the number of CDS and the length of the genomes can be explained by the N50 and the average length of the contigs. As it is shown in **Table 2**, the genome sequence from strain CA-128772 presents a higher N50 value (186,760) and a higher average contig length (83,989 bp) than strain CA-126428 (N50 = 139,227, average length = 55,808 bp) suggesting that the CA-126428 genome is more fragmented and probably more CDSs are missed.

These Whole Genome Shotgun projects were deposited at DDBJ/ENA/GenBank under the accession numbers PPHF00000000 (strain CA-126428) and PPHG00000000 (strain CA-128772). The versions described in this paper are versions PPHF01000000 and PPHG01000000 (Supplementary Table 1).

Taxonomic Identification and Phylogenetic Analysis of Strains CA-126428 and CA-128772

The molecular identification of strains CA-128772 and CA-126428 was based on the comparison of their 16S rRNA gene sequences to reference type strains sequences using the EzBiocloud server. Both strains showed to contain only one 16S rRNA copy. The 16S rRNA gene sequences were deposited at GenBank under the accession numbers MG800320 (strain CA-126428) and MG799844 (strain CA-128772) (Supplementary Table 1). Strain CA-128772 showed the highest similarity values with *A. pretoriensis* DSM 44654^T (99.79%), *A. rifamycinica* DSM 46095^T (99.58%), *A. lexingtoniensis* DSM 44653^T (99.58%) and *A. tolypomycina* DSM 44544^T (99.50%). With respect to isolate CA-126428, the highest similarity values were obtained with *A. mediterranei* DSM 43304^T (99.51%), *A. kentuckyensis* DSM 44652^T (99.44%), *A. rifamycinica* DSM 46095^T (99.44%) and *A. pretoriensis* DSM 44654^T (99.37%).

To confirm the taxonomic assignment and the existing phylogenetic relationships between our isolates and other reference and non-type *Amycolatopsis* strains (Supplementary Table 1), a phylogenetic tree was built based on the complete 16S rRNA gene sequences, using *Micromonospora chalcea* DSM 43026^T as outgroup (**Figure 1**). *Amycolatopsis* species group into three major subclades including the mesophilic or moderately thermophilic *A. orientalis* (AOS) subclade, the thermophilic *A. methanolica* (AMS) subclade, and the mesophilic *A. taiwanensis* (ATS) subclade (**Figure 1**) (Tang et al., 2016). The AOS subclade includes previously defined groups A-E and G (Everest and Meyers, 2009, 2011) and groups H, I, and J (Tang et al., 2016). The ATS and AMS subclades include the F group (Everest and Meyers, 2009, 2011), that was latter divided in F1 (ATS) and F2 (AMS) (Tang et al., 2016).

The phylogenetic tree associates both strains CA-126428 and CA-128772 to the AOS subclade, containing 36 type species from which three have a full genome sequence, as well as three non-type strains of *A. mediterranei* (RB, S699 and

TABLE 2 | NGS statistic data of CA-126428 and CA-128772 draft genomes.

Strain	# Contigs	Contigs sum (bp)	N50	Longest contig (bp)	Shortest contig (bp)	Average length (bp)	% GC content	# CDS
CA-126428	188	10,492,049	139,227	724,738	1,147	55,808	71.39	9,626
CA-128772	121	10,162,752	186,760	371,657	231	83,989	71.80	10,896

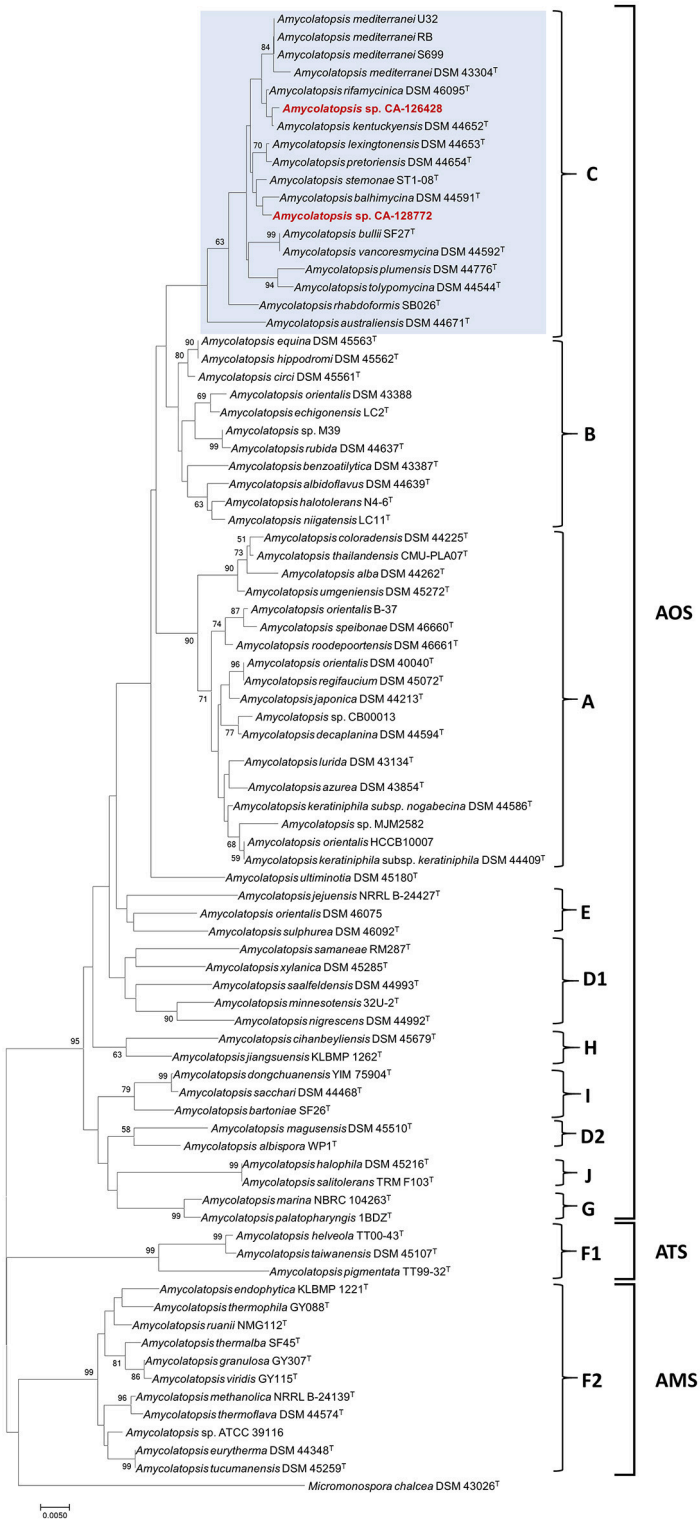


FIGURE 1 | Phylogenetic NJ tree based on 16S rRNA gene sequences from 82 *Amycolatopsis* strains. The scale bar indicates 5 nucleotide substitutions per 1,000 nucleotides, and the node numbers are percentage bootstrap values based on 1,000 resampled datasets. Bootstrap values below 50% are not shown. The AOS, ATS and AMS subclades are indicated, as well as the groups A-J. The strains grouped with CA-126428 and CA-128772 isolates (group C) are blue-shaded.

U32), five non-type strains of *A. orientalis* (DSM 43388, DSM 46075, HCCB10007, CB00013, and B-37) and two strains of *Amycolatopsis* sp. (M39 and CB00013). Both CA-126428 and CA-128772 strains, together with the three non-type *A. mediterranei* strains, are included in a branch containing 13 type strains, identified as group C (**Figure 1**). Within the AMS subclade, another non-type genome-sequenced strain, *Amycolatopsis* sp. ATCCC 39116, was included (**Figure 1**). In general, the topology of the groups A, B, C, F1, F2, G, H, and I is conserved in our tree despite the additional species that were not previously analyzed (**Figure 1**). The only exceptions are the group D, that is divided in two groups (D1 and D2), and the strain *A. ultiminotia* DSM 45180^T, that is no longer included in group E (**Figure 1**).

The closest genetic distances were determined for strain CA-128772 with *A. pretoriensis* DSM 44654^T (0.003) and for strain CA-126428 with *A. kentuckyensis* DSM 44652^T (0.001) (Supplementary Figure 1). The distances among the strains in group C ranged between 0.001 and 0.02, and no distance was observed among the non-type *A. mediterranei* strains. Strains *A. helveola* TT00-43^T, *A. taiwanensis* DSM 45107^T and *A. pigmentata* TT99-32^T showed the highest evolutionary distances (0.07–0.09) with the rest of strains. Nonetheless, these values are very low and show that these species are very closely-related.

Two new 16S rRNA- and MLSA-based phylogenetic trees were constructed only including the 43 genome-sequenced strains (**Figure 2**). The MLSA tree was constructed from four-gene (*atpD*, *dnaK*, *recA*, and *rpoB*) concatenated nucleotide sequences. As shown in **Figure 2**, the three AOS, ATS and AMS subclades are present in both trees, but some differences are observed in the organization of the groups in AOS subclade. The same A, B and C groups are defined in both trees, whereas in the MLSA tree, group B includes strains from group E and group A includes strains from group D1. In the 16S rRNA-based tree strains from groups E, G, and J clustered together, as well as strains from D1 and I groups, whereas in the MLSA tree, strains from groups G and J are clustered.

The composition of group C, that contains strains CA-126428 and CA-128772, is conserved in both 16S rRNA and MLSA trees, except for the strain *A. saalfeldensis* DSM 44993^T, which is not included in the group in the MLSA tree. However, the relative position of the strains differs in both trees. In the 16S rRNA tree strain CA-126428 is closely related to the strains *A. rifamycinica* DSM 46095^T, *A. kentuckyensis* DSM 44652^T and *A. mediterranei*, and strain CA-128772 is closely related to *A. balhimycina* DSM 44591^T, *A. lexingtoniensis* DSM 44653^T and *A. pretoriensis* DSM 44654^T. In contrast, in the MLSA tree, both strains cluster together and are closely related to *A. rifamycinica* DSM 46095^T and the *A. mediterranei* strains. Nevertheless none of these relationships were supported by the bootstrap values, suggesting that both strains could represent novel species within the genus. Bootstrap values of group C (100% in MLSA tree and >50% in 16S rRNA tree) support the significance of the MLSA-based phylogeny. The genomic distances based on both 16S rRNA and MLSA phylogenetic analyses are shown in **Figure 3**. The 16S rRNA sequence distances ranged from 0 to 0.096 (mean 0.039), while MLSA distances present a broader range between 0 and 0.582 (mean 0.071).

Previous whole genome sequence studies (Tang et al., 2016) have shown that some *Amycolatopsis* strains present multiple copies of 16S rRNA genes, which is not the case of strains CA-126428 and CA-128772. The 16S rRNA sequence identities for some inter-species pairs are higher than those of the corresponding intra-species pairs and this fact may influence the structure of *Amycolatopsis* 16S rRNA gene phylogenetic trees (Tang et al., 2016). The lack of resolution at the species level of 16S rRNA gene-based phylogenies can be overcome by MLSA methods (Glaeser and Kämpfer, 2015). The observed inconsistency between MLSA and 16S rRNA gene phylogenies reinforce the idea that MLSA allows a much more precise species delineation within bacteria (Stackebrandt et al., 2002; Konstantinidis and Tiedje, 2007; Thompson et al., 2013). However, given that not all housekeeping gene sequences are available for all species, some improvements are needed to make MLSA more generally applicable. The selection of an universal set of genes for all the prokaryotes, the standardization of the number and length of genes to be used or a low time consuming calculation method have been recently proposed (Glaeser and Kämpfer, 2015).

In the case of the strains of *Amycolatopsis*, it has been shown that a 315 bp variable fragment of the *gyrB* gene has a higher resolution power than the 16S rRNA gene (Everest and Meyers, 2009). This partial *gyrB* gene sequence, together with other housekeeping genes, has been proposed as a good candidate to be used for MLSA of this genus. Unfortunately, we were not able to use the *gyrB* sequence as a MLSA marker, since the sequences were partial or not present in some genomes.

Genome Comparisons

The overall similarity of the *Amycolatopsis* genomes was analyzed using several approaches.

First, we performed an overall analysis using the Genome-to-Genome Distance Calculator (GGDC) (Auch et al., 2010; Meier-Kolthoff et al., 2013, 2014), that calculates digital (*in silico*) DNA-DNA hybridization (dDDH) from the intergenomic distances and quantifies the G+C content. No G+C content differences were found among *A. mediterranei* strains. The highest G+C content differences were found between *A. halophila* DSM 45216^{T}and the strains from group C (about 4% difference) (Supplementary Figure 2). In the case of strain CA-128772, the maximum and minimal G+C content differences were found with *A. balhimycina* DSM 44591^T (1.07%), and with *A. rifamycinica* DSM 46095^T (0.1%). In strain CA-126428, the maximum G+C content difference was also observed with *A. balhimycina* DSM 44591^T (0.62%), whereas the minimum difference (0.15%) was with *A. lexingtonensis* DSM 44653^T. According to Meier-Kolthoff et al. (2014), within-species differences in the G+C content are almost exclusively below 1%. The observed differences below 1% between different species are probably due to the incompleteness of most of the genomes analyzed and/or inaccuracies of the species descriptions (Meier-Kolthoff et al., 2014). dDDH analysis showed that, in general, the re-association values among *Amycolatopsis* strains were in the range of 20%. However, re-association values higher than 70% were observed between some of the genomes analyzed corresponding to non-type species, indicating that they may

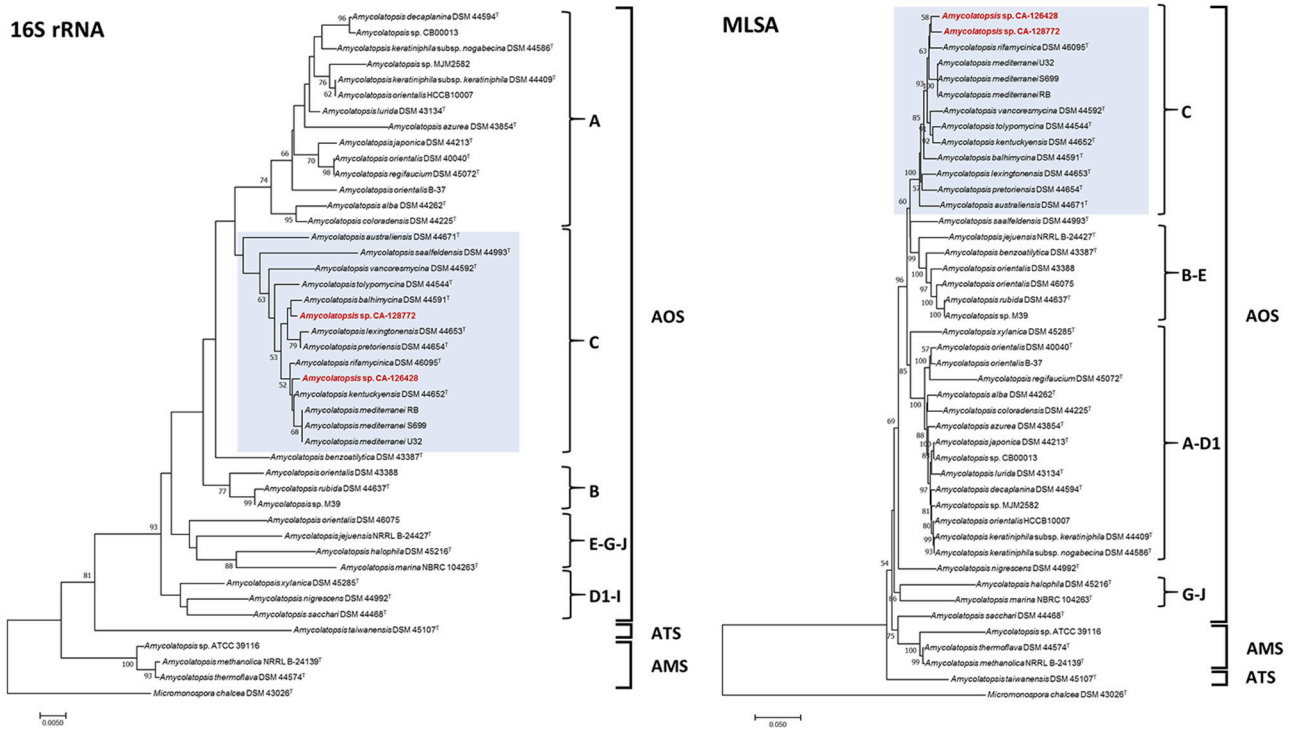


FIGURE 2 | Phylogenetic NJ trees based on 16S rRNA gene sequences **(Left)** and four concatenated sequences (*atpD, dnaK, recA* and *rpoB*) (MLSA, **Right**) from 43 genome-sequenced *Amycolatopsis* strains. The scale bars indicate the percentage of difference in the nucleotide sequences, and the node numbers are percentage bootstrap values based on 1,000 resampled datasets. Bootstrap values below 50% are not shown. The AOS, ATS and AMS subclades are indicated, as well as the groups A–J. The strains grouped with CA-126428 and CA-128772 isolates (group C) are blue-shaded.

belong to the same species (Supplementary Figure 3). This is the case of *A. orientalis* B-37 and *A. orientalis* DSM 40040^T with a 72.1% of re-association; *A. orientalis* HCCB10007 showed a 72.5% of re-association with *A. keratiniphila* subsp. *keratiniphila* DSM 44409^T and a 71.9% with *A. keratiniphila* subsp. *nogabecina* DSM 44586^T; *A. japonica* DSM 44213^T showed a 72.7% of re-association with *Amycolatopsis* sp. MJM2582 and a 88.3% with *Amycolatopsis* sp. CB00013 suggesting both strains to belong to *A. japonica; Amycolatopsis* sp. M39 and *A. rubida* DSM 44637^T had a 93.3% of re-association; *A. keratiniphila* subsp. *nogabecina* DSM 44586^T and *A. keratiniphila* subsp. *keratiniphila* DSM 44409^T showed a 90.1%; *Amycolatopsis* sp. CB00013 had a 72.6% of re-association with *Amycolatopsis* sp. MJM2582. The rest of the dDDH obtained were below 70%, indicating that these strains belong to different species.

The re-association values between strains of group C ranged from 23 to 54%, except for the three species strains of *A. mediterranei* (U32, S699, and RB), with a 100% of re-association. In the case of the two new strains of our study, strain CA-128772 showed the maximum re-association value (45.3%) with *A. vancoresmycina* DSM 44592^T, and strain CA-126428 had the maximum value (54.4%) with the three *A. mediterranei* strains, well below the threshold defined for strains of the same species. These results suggest that both CA-128772 and CA-126428 isolates may represent new *Amycolatopsis* strains.

Another approach used to study the similarity of the *Amycolatopsis* genomes was to perform an analysis of the Average Nucleotide Identity (ANI) and the orthology (OrthoANI) (Lee et al., 2016). The comparison of CA-128772 and CA-126428 genomes with the rest of *Amycolatopsis* genomes showed that the maximum ANI and OrthoANI values were obtained with strains included in group C. However, none of them reach the ANI threshold range (95–96%) for species delineation (**Table 3**). Again, as previously observed with the dDDH values, strain CA-128772 showed the highest similarity (92.0 and 91.35%) with *A. vancoresmycina* DSM 44592^T, while strain CA-126428 showed the highest similarity (93.97 and 93.03%) with *A. mediterranei* U32. These results suggest again that CA-128772 and CA-126428 isolates may represent new *Amycolatopsis* species.

Several studies have confirmed that GGDC analysis yields higher correlations with classical DDH than ANI softwares (Auch et al., 2010; Meier-Kolthoff et al., 2013, 2014). Moreover, dDDH calculation is independent of genome length and is thus robust against the use of incomplete draft genomes (Auch et al., 2010).

To compare the relative organization of the genomes, we also performed a gene synteny progressiveMauve analysis

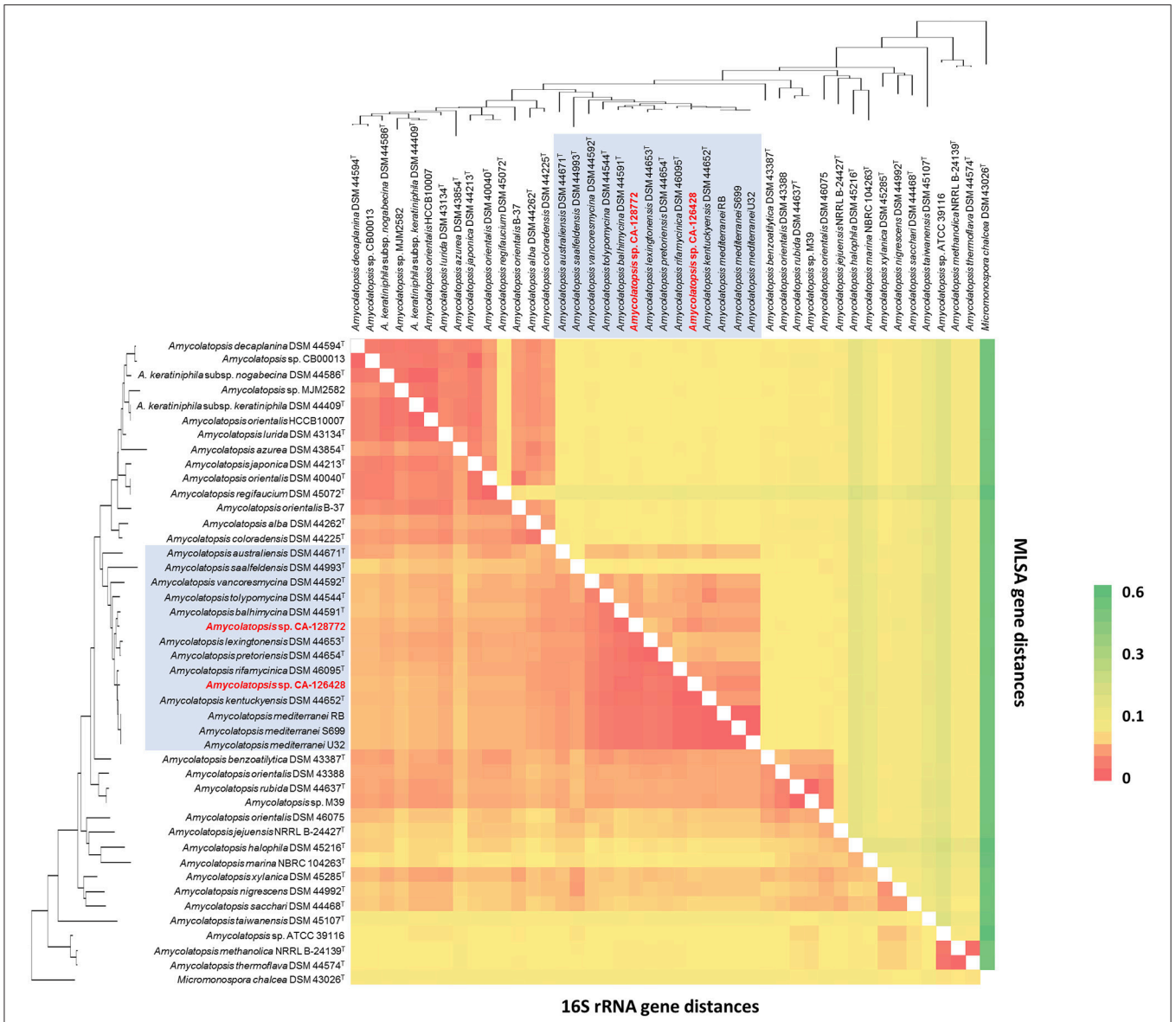
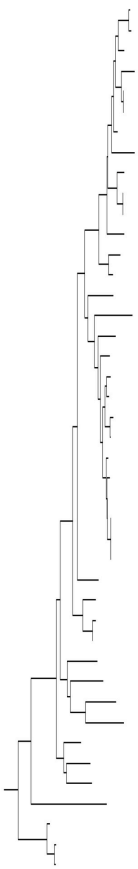


FIGURE 3 | 16S rRNA (Bottom left) vs. MLSA (Top right) genomic distances heatmap using the Kimura 2-parameter model. Self-genome comparisons occur on a diagonal line stretching from the top left to the bottom right corners. The strains are ordered in the same way as the 16S phylogenetic tree in **Figure 2**, which has been placed upon both axes for orientation. The strains belonging to group C have been shaded in blue for clarification. The heatmap legend is shown on the right.

(Darling et al., 2010) of the concatenated draft genomes of strains CA-126428 and CA-128772 with nine *Amycolatopsis* complete genomes (**Figure 4**). The complete genomes were linearized at position 1 of their sequences. Once again, the high similarity of the three *A. mediterranei* strains was clearly observed. As previously described (Tang et al., 2016), highly conserved core regions were detected in the left and right arms of all the genomes, except for *A. lurida* DSM 43134^T, which has a different genome rearrangement, and strains CA-126428 and CA-128772. Since the latter are draft concatenated contigs and not complete genomes, the order of the regions is altered, but homologous regions can be observed in the alignment (**Figure 4**). Other central regions of the chromosomes are also conserved among the genomes, especially in the *A. orientalis, A. mediterranei, A. japonica* and *A. keratiniphila* strains.

Due to the lack of resolution at the species level of the 16S rRNA gene-based phylogenies, and the difficulty to design an universal set of primers for MLSA analysis, our results derived from whole-genome content comparisons arise, as previously reported for other strains (Colston et al., 2014), as the most valuable tool to discriminate between bacterial species. As more bacterial genomes become available, the use of whole genome sequences opens new opportunities for the characterization of bacterial species.

TABLE 3 | OrthoANI and ANI calculations of CA-128772 and CA-126428 genomes against other *Amycolatopsis* genomes.



Strain	CA-128772			CA-126428		
	OrthoANI (%)	ANI (%)	dDDH (%)	OrthoANI (%)	ANI (%)	dDDH (%)
A. decaplanina DSM 44594^T	78.50	77.18	22.1	78.41	77.19	22
Amycolatopsis sp. CB00013	78.46	77.22	22.1	78.51	77.24	22.2
A. keratiniphila subsp. *nogabecina* DSM 44586^T	78.50	77.25	22.2	78.34	77.22	22.3
Amycolatopsis sp. MJM2582	78.54	77.25	22.2	78.32	77.18	22.1
A. keratiniphila subsp. *keratiniphila* DSM 44409^T	78.43	77.26	22.1	78.41	77.21	22.2
A. orientalis HCCB10007	78.55	77.30	22.3	78.58	77.26	22.2
A. lurida DSM 43134^T	78.43	77.22	22.3	78.38	77.11	22.2
A. azurea DSM 43854^T	78.55	77.37	22.2	78.57	77.24	22.2
A. japonica DSM 44213^T	78.53	77.35	22.2	78.51	77.31	22.3
A. orientalis DSM 40040^T	78.57	77.30	22.2	78.47	77.21	22.2
A. regifaucium DSM 45072^T	78.54	77.42	22.2	78.42	77.23	22.2
A. orientalis B-37	78.55	77.27	22.2	78.35	77.10	22.2
A. alba DSM 44262^T	78.29	77.04	22.1	78.25	77.00	22
A. coloradensis DSM 44225^T	78.17	77.04	22.1	78.11	76.97	22.1
A. australiensis DSM 44671^T	87.59	86.53	33.5	87.53	86.40	33.7
A. saalfeldensis DSM 44993^T	80.02	78.70	23.3	79.83	78.60	23.5
A. vancoresmycina DSM 44592^T	92.05	91.35	45.3	89.98	88.97	39.4
A. tolypomycina DSM 44544^T	89.75	88.68	38.3	90.59	89.53	41.1
A. balhimycina DSM 44591^T	88.55	87.20	36	89.14	87.75	37.6
***Amycolatopsis* sp. CA-128772**	100.00	100.00	100.00	89.80	88.79	38.7
A. lexingtonensis DSM 44653^T	88.33	87.44	35.3	88.07	86.80	34.7
A. pretoriensis DSM 44654^T	87.79	86.71	34.2	87.64	86.58	34.2
A. rifamycinica DSM 46095^T	89.82	88.93	38.9	91.14	90.23	43.2
***Amycolatopsis* sp. CA-126428**	89.80	88.79	38.7	100.00	100.00	100
A. kentuckyensis DSM 44652^T	90.37	89.36	40.4	91.81	90.83	45.6
A. mediterranei RB	90.12	89.05	39.8	93.92	93.01	54.4
A. mediterranei S699	90.09	89.04	39.8	93.94	93.03	54.4
A. mediterranei U32	90.06	89.08	39.8	93.97	93.03	54.4
A. benzoatilytica DSM 43387^T	78.69	77.43	22.7	78.69	77.37	22.5
A. orientalis DSM 43388	78.86	77.67	22.5	78.68	77.59	22.6
A. rubida DSM 44637^T	78.49	77.16	22.2	78.34	77.06	22.1
Amycolatopsis sp. M39	78.47	77.01	22.1	78.21	76.98	22.1
A. orientalis DSM 46075	78.60	77.36	22.5	78.54	77.35	22.4
A. jejuensis NRRL B-24427^T	78.28	76.99	22.3	78.19	76.92	22.4
A. halophila DSM 45216^T	73.87	73.04	20.2	73.78	72.96	19.9
A. marina NBRC 104263^T	75.33	74.05	20.4	75.23	73.86	20.6
A. xylanica DSM 45285^T	77.57	76.21	21.5	77.47	76.12	21.3
A. nigrescens DSM 44992^T	76.46	75.14	20.9	76.25	75.02	20.8
A. sacchari DSM 44468^T	76.88	75.53	21.4	76.96	75.46	21.4
A. taiwanensis DSM 45107^T	75.40	74.23	21.4	75.27	74.11	21
Amycolatopsis sp. ATCC 39116	77.07	75.81	21.7	77.09	75.69	21.5
A. methanolica NRRL B-24139^T	77.21	76.17	21.6	77.12	75.93	21.5
A. thermoflava DSM 44574^T	76.86	75.61	21.6	76.88	75.60	21.4

The dDDH values have been included for comparison. Blue: lowest values, yellow: medium values, green: highest values. The strains forming a clade with CA-126428 and CA-128772 are gray-shaded. The strains are ordered in the same way as the 16S phylogenetic tree in ***Figure 2****, which has been placed on the left for orientation.*

Primary Metabolism Analysis

As saprophytic bacteria, actinomycetes have a well-coordinated carbohydrate and nitrogen catabolic systems that allow the adaptation and the efficient utilization of the resources in the nutrient-limited conditions of the environment. These bacteria can produce a large diversity of extracellular enzymes to digest complex polymeric substrates and import the resulting monomers and oligomers to be used as nutrients for catabolism and anabolism and biomass generation (Genilloud, 2017b).

With the aim to analyze the global primary metabolism of strains CA-126428 and CA-128772, their genomes were functionally characterized in the KEGG PATHWAY Database (http://www.genome.jp/kegg/pathway.html) using the BlastKOALA sequence similarity tool (http://www.kegg.jp/blastkoala/). The KEGG BlastKOALA tool mapped 31.8% of the CA-126428 predicted proteins (3063 proteins) to KEGG ortholog groups, while only 19% of the CA-128772 predicted proteins (1882) could be mapped. Supplementary Table 2 presents the distribution of pathways and the number of genes annotated with BlastKOALA, and shows that the content and distribution of different metabolic pathways are very similar between both strains. Although not enough data are available because of the incompleteness of the genomes, some observations on their primary metabolism have been achieved.

A wide variety of extracellular polysaccharide-degrading enzymes have been mapped in the genomes of both *Amycolatopsis* strains including amylases, endoglucanases, xylanases, chitinases, and beta-glucosidases. However, no ligninases, agarases, mannanases, or cellulases have been detected. As in the case of many actinomycetes, some multiple transport systems to uptake specific carbohydrates are present in both genomes: such as ABC permeases and specific phosphoenolpyruvate-dependent phosphotransferase systems (PTS). In most bacteria, the PTS system plays a major role in the carbon catabolite repression (CCR), a global control system that ensures the preferential use of the different carbon sources available. CCR is mediated by the glycolytic enzyme

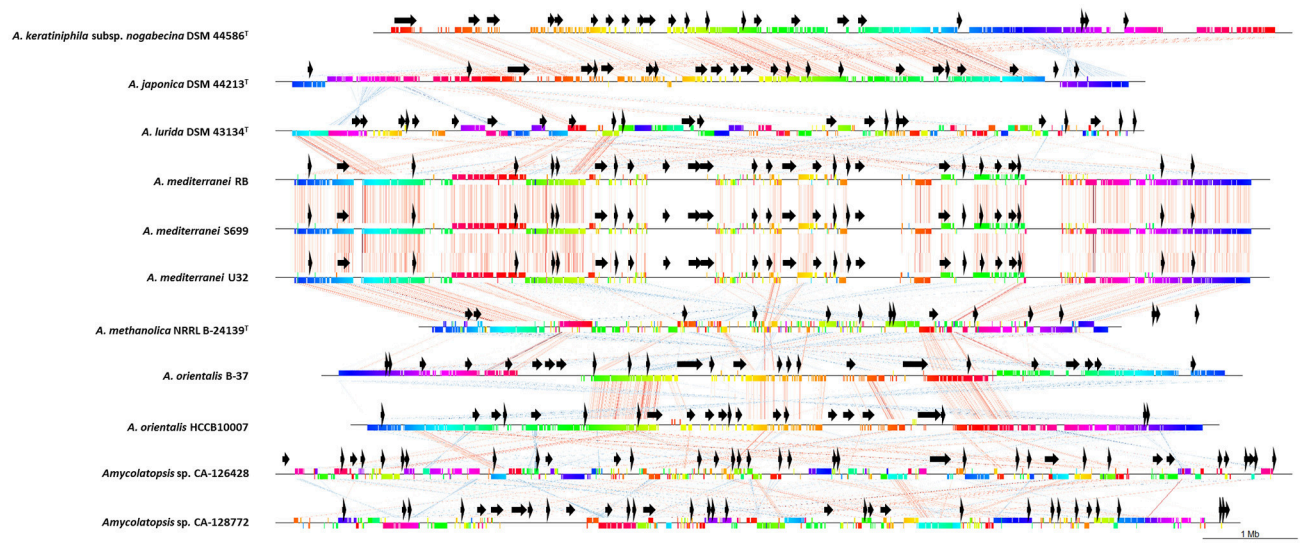


FIGURE 4 | Comparative analyses of CA-126428 and CA-128772 draft genomes with 9 complete *Amycolatopsis* genomes. Horizontal straight lines represent the genomes and vertical colored bars represent different conserved genes. The vertical bars are connected by corresponding colored thin lines. Black arrows represent the position of the secondary metabolite biosynthetic pathways predicted by antiSMASH.

glucose kinase, for which an ortholog has been identified in both strains. Glucose kinase converts glucose to glucose-6-phosphate, a substrate of different enzymes and pathways (glucose phosphate isomerase, glucose-6-phosphate dehydrogenase, pentose phosphate pathway, and glucose-6-phosphatase) and also plays a key role mediating CCR for enzymes involved in primary and secondary metabolism and precursors for natural product synthesis (Kwakman and Postma, 1994).

Actinomycetes carbohydrate primary metabolism is complex and characterized by the presence of multiple isoenzymes involved in single catalytic steps of glycolysis, the pentose phosphate and TCA cycle pathways. Normally, the genes encoding those enzymes do not cluster in operons and are scattered in the chromosome, subjected to different regulation depending on the carbon sources or the development stage (Genilloud, 2017b). The most important glycolytic pathway is the Embden-Meyerhof (EM) pathway (Salas et al., 1984), which is used to generate ATP and provide precursors for secondary metabolism. The EM pathway is regulated at the level of the glycolytic enzyme phosphofructokinase (Pfk) (Genilloud, 2017b), an enzyme that has not been identified so far in the genomes of strains CA-126428 and CA-128772, suggesting a different type of regulation. The enzymes involved in glucose metabolism *via* glycolysis and the pentose phosphate pathway (PPP) (Salas et al., 1984), as well as those involved in the tricarboxylic acid cycle (TCA), essential in the supply of precursors to secondary metabolism, have been mapped as expected in both isolates CA-126428 and CA-128772. The Entner-Doudoroff (ED) pathway is unusual in actinomycetes (Gunnarsson et al., 2004), although some of them contain homologs of 6-phosphogluconate dehydratase gene, as is the case of the strain CA-128772. Whether this pathway is active or not on this strain should be further investigated.

Amycolatopsis, as other actinomycetes, require nitrogen to ensure biomass production and to synthesize a large diversity of secondary metabolites from amino acids that are used as key precursors. Since actinomycetes are frequently isolated from nitrogen-poor environments, these bacteria have developed a complex system to efficiently retrieve nitrogen. A large diversity of extracellular proteases and peptidases are produced, as well as amino acid and oligopeptide transport systems. Some orthologous genes of these systems are also present in the genomes of strains CA-126428 and CA-128772, such as serine protease PepD, leucyl aminopeptidase, methionyl aminopeptidase, D-aminopeptidase, carboxypeptidases, amino acid, and oligopeptide transport system permeases. Different routes are followed to catabolize the families of amino acids introduced by permeases upon action of the extracellular proteases. Strains CA-126428 and CA-128772 have orthologous of most of the genes involved in these reactions, although some differences have been found between them. For example, histidine can be processed *via* formyl glutamic acid in the strain CA-128772, but in the case of CA-126428, it seems that histidine is processed *via* ergothioneine. Proline catabolism involves a proline oxidase and a pyrroline-5-carboxylate dehydrogenase to form glutamate, but the proline oxidase activity has not been mapped so far in any of the genomes. Valine dehydrogenase is responsible of the deamination of valine, leucine, and isoleucine. However, this enzyme has only been detected in strain CA-128772. Another enzyme with the same function, a branched-chain amino acid aminotransferase, is present both in CA-126428 and CA-128772. The enzymes involved in the catabolism of lysine, which usually follows the cadaverine pathway to

generate glutarate, are absent in both strains. Only enzymes lysine N6-hydroxylase and lysine 2,3-aminomutase, involved in the conversion of L-lysine to N6-hydroxylysin and L-β-lysine, respectively, have been located. Enzymes involved in alanine catabolism have not been found either.

The enzymes of the shikimate pathway have been mapped in the isolates CA-126428 and CA-128772. An important number of secondary metabolites are derived from intermediates of the shikimate pathway, which is responsible of the formation of chorismate, a precursor of tryptophan, phenylalanine and tyrosine, as well as prephenate, anthranilate and p-aminobenzoate.

Interestingly, both strains possess genes involved in the degradation of aromatic compounds such as toluene, benzoate, fluorobenzoate or xylene, a trait that could be very useful for their application in bioremediation processes. So far, the genus *Amycolatopsis* has not been deeply studied in the field of bioremediation, with the exception of *Amycolatopsis tucumanensis* DSM45259^T, which is the only species of this genus found to be resistant to copper (Albarracin et al., 2010). Another *Amycolatopsis* strain, M3-1, is able to degrade the herbicide ZJ0273, and several *Amycolatopsis* strains possess the capacity to degrade polylactic acid (PLA) (Dávila Costa and Amoroso, 2014).

In addition, we identified in both strains genes involved in antimicrobial resistance to vancomycin, beta-lactam antibiotics and cationic antimicrobial peptides. The presence of vancomycin resistance genes has been applied to the discovery of glycopeptide-producing strains using different screening approaches (Thaker et al., 2013; Truman et al., 2014), as it was the case of the ristomycin A producer *Amycolatopsis* sp. MJM2582 (Truman et al., 2014). The analysis of these genes, together with other genome mining approaches, highlights the potential to produce secondary metabolites by strains CA-128772 and CA-126428.

The former analysis of the different primary metabolism pathways present in the genomes in study supports the ability of both *Amycolatopsis* strains to utilize a broad variety of sources to ensure the supply of basic nutrients. Particularly in lichens, symbiotic partners contributions allow to colonize extreme environments and to tolerate harsh conditions. Despite the lack of information about the potential role of these strains as part of the microbial community associated to the lichen, there is also accumulating evidence that the bacterial counterparts may also contribute given their metabolic capabilities, providing carbohydrates, nitrogen sources and secondary metabolites to the microbial consortia (Scherlach and Hertweck, 2017). Future research in the physiology and primary metabolism regulation of *Amycolatopsis* spp., as well as its influence on lichen symbiosis and secondary metabolism, will benefit from the new genomic approaches and genomic-scale metabolic analyses.

Secondary Metabolite Biosynthetic Gene Cluster Analysis

Primary and secondary metabolisms are deeply interconnected in a complex network of regulatory signals that sense the environment and ensure the survival and adaptation of the microbial community (Genilloud, 2017b). The capacity to produce multiple secondary metabolites depends on the use of the available precursors and building blocks provided by the primary metabolism. In contrast to what has been observed in primary metabolism, the genes related with the production of secondary metabolites are frequently clustered, and their expression is modulated by transcriptional regulators (Genilloud, 2017b).

We used the antiSMASH algorithm (Blin et al., 2017) to search for putative BGCs clusters in all the *Amycolatopsis* genomes described in Supplementary Table 1. The secondary metabolite classes examined cover all common secondary metabolites in actinomycetes (butyrolactone, ectoine, fatty acid, indole, NRPS, RiPP, saccharide, siderophore, PKS-I, PKS-II, PKS-III, and terpene) and are shown in **Figure 5.** As it might be expected, the number of predicted secondary metabolite biosynthetic gene clusters depends on the completeness of the genomes and the size of the contigs. Overall, the number of polyketide (PKS) and non-ribosomal peptide (NRPS) BGCs is similar among all the *Amycolatopsis* strains and correlates with the completeness of the genomes. In the case of *A. kentuckyensis* DSM 44652^T and *A. lexingtoniensis* DSM 44653^T, a great number of putative PKS-I and NRPS pathways were predicted; however, these are very fragmented genomes and thus the number of predicted ORFs and BGCs increases, especially in the case of repeating, multi-modular proteins such as PKS and NRPS (Klassen and Currie, 2012).

The antiSMASH analysis of *Amycolatopsis* sequenced genomes detected many secondary metabolite BGCs when only a few metabolites have been reported to be produced by these strains, suggesting that even these well-studied species have the potential to produce new molecules (Chen et al., 2016). In the case of strains CA-126428 and CA-128772, antiSMASH predicted as many as 140 pathways for both strains given the high fragmentation level (Supplementary Table 3). In spite of the high number of predicted BGCs, only 70 clusters from CA-126428 and 55 clusters from CA-128772 show some degree of homology with known BGCs in MIBiG, and among them, only five clusters from each strain show equal or more than 50% of homology (Supplementary Table 3) suggesting the high new biosynthetic potential encoded in the genomes of the isolates CA-126428 and CA-128772.

Strain CA-128772 is richer than strain CA-126428 in saccharide pathways (**Figure 5**, Supplementary Table 3). A predominance of saccharide gene clusters among microbial genomes was previoulsy found in a global analysis of prokaryotic biosynthetic gene clusters (Cimermancic et al., 2014). Cell wall-associated saccharides play key roles in microbe-host and microbe-microbe interactions, while other diffusible saccharides have antibacterial activity (Cimermancic et al., 2014). The functions of many of the putative saccharide BGCs are still unknown since they are not closely related to any known gene cluster. In the case of strain CA-128772, approximately 40% of the predicted saccharide clusters do not show homology to any known cluster, in contrast to the 27% in strain CA-126428 (Supplementary Table 3).

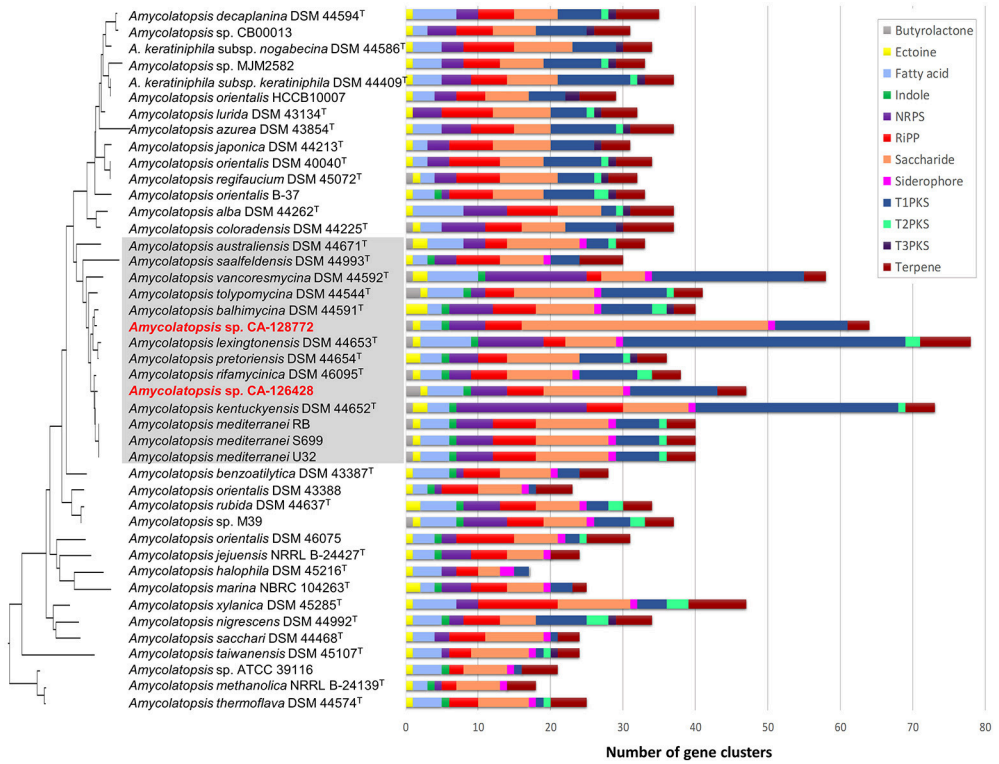


FIGURE 5 | Most frequent secondary metabolite biosynthetic gene clusters predicted by antiSMASH in the genomes of *Amycolatopsis* strains. The strains are ordered in the same way as the 16S rRNA phylogenetic tree in **Figure 2**, which has been placed on the left axis for orientation. The strains of group C have been shaded in gray for clarification. The length of each horizontal bar corresponds with the number of BGCs, and the color code is indicated on the right.

We focused our analyses on the predicted BGC with more than 70% homology with known BGCs from the MIBiGC database (**Figure 6**) (Medema et al., 2015). **Figure 6** shows the pathways predicted in each strain (yellow boxes), as well as the secondary metabolites that have been detected in culture fermentations (green or red boxes if the pathway has been or not predicted, respectively). Interestingly, strains CA-126428 and CA-128772 show only two and three BGCs, respectively, with more than 70% homology with known pathways (**Figure 6**). The rest of pathways (**Figure 5**, Supplementary Table 3) show low homology with known BGCs. As stated above, this fact reflects the biosynthetic potential of the strains, which may host novel BGCs encoding new secondary metabolites.

All the strains, including CA-126428 and CA-128772, were predicted to produce ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid), a compatible solute with a considerable biotechnological importance that acts as stress protectant and stabilizes macromolecules against severe environmental conditions (Hamedi et al., 2013). The production of ectoine has been widely described in salt-tolerant *Streptomyces* strains (Nett et al., 2009; Zhao et al., 2016), and the antiSMASH database (https://antismash-db.secondarymetabolites.org) shows the presence of the ectoine BGC in a high number of strains belonging to the *Proteobacteria* and *Actinobacteria*. This fact suggests the capacity of adaptation of the members of this genus to different environments, since the production of ectoine may act as a protectant against desiccation or drought periods. Nevertheless, we have no further evidences about the potential role of the production of ectoine by *Amycolatopsis* in the lichen-associated microbial community.

In addition, our strains were only predicted to produce three compounds. The strain CA-128772, was predicted to produce 2-methylisoborneol, a widespread odorous-terpenoid compound (Nett et al., 2009) also shown in another 23 strains. Surprisingly, the production of this compound was not identified in any of the strains from subclades ATS and AMS. This strain was also predicted to synthetize the siderophore schabichelin (Kodani et al., 2013) as well as 9 of the 14 strains of group C (**Figure 6**). Strain CA-126428, as well as another 21 strains, contained a cluster encoding thiolactomycin, a thiotetronate antibiotic first described in 1982 with a broad antibacterial activity against a wide spectrum of Gram-positive and Gram-negative bacteria (Yurkovich et al., 2017). Interestingly, the production of thiolactomycin was predicted in nearly all the strains from group A except for *A. lurida* DSM 43134^T, whereas within group C only 4 of the 14 strains may produce this compound, as well as *A. thermoflava* DSM 44574^T, a strain from the AMS subclade (**Figure 6**).

An additional group of 21 strains were predicted to produce the RiPP class III lantibiotic erythreapeptin (Völler et al.,

FIGURE 6 | Biosynthetic gene clusters predicted in the genomes of *Amycolatopsis* with more than 70% homology with known BGCs. Yellow: predicted pathways whose encoded metabolites have not been detected in culture; green: predicted pathways whose encoded metabolites have been detected in culture; red: compounds detected in culture but not predicted by antiSMASH. The strains are ordered in the same way as the 16S rRNA phylogenetic tree in **Figure 2**, which has been placed on the left axis for orientation. The strains of group C have been shaded in gray for clarification.

2012), from which only strains *A. australiensis* DSM 44671^T, *A. balhimycina* DSM 44591^T and *A. pretoriensis* DSM 44654^T belonged to group C. As in the case of thiolactomycin, the production of erythreapeptin was predicted in nearly all the strains conforming group A (**Figure 6**), with the exception of *A. lurida* DSM 43134^T.

Interestingly, only five strains showed a BGC with more than 70% homology with the rifamycin pathway of the ansamycin class of antibiotics, and another five strains with the vancomycin glycopeptide pathway (Chen et al., 2016). Curiously, the strains predicted to contain PKS pathways associated to rifamycin belong exclusively to group C, while strains predicted to contain a vancomycin related pathways are only present in group A.

Other biosynthetic pathways have been predicted exclusively in specific groups. These are the cases of the clusters encoding the polyketides amphotericin (Caffrey et al., 2001), BE-7585A (Sasaki et al., 2010), halstoctacosanolide (Tohyama et al., 2006), micromonolactam (Skellam et al., 2013) and

the non-ribosomal peptide tomaymycin (Li et al., 2009) that were also detected only in individual strains from group C (**Figure 6**). The production of the glycopeptide antibiotic ristomycin A (Spohn et al., 2014) was predicted only in 7 of 14 strains from group A. Other predicted clusters are involved in the production of several siderophores such as mirubactin (Giessen et al., 2012), predicted in 11 strains from group A and in another 4 strains conforming group B, abachelin (Kodani et al., 2015), predicted in 12 strains from group A and some strains from groups D1, G, and E, and amychelin (Seyedsayamdost et al., 2011), predicted in two strains of the AMS subclade and in two strains of another AOS subgroups.

The relative position of the BGC homologous to known pathways is shown in **Figure 4** (black arrows). In strains CA-126428 and CA-128772, the observed BGC distribution is not the order in the genome given that they correspond to draft concatenated and not complete genomes. Our comparison shows that strains *A. keraniphila* subsp. *nogabecina* DSM 44586^T and *A. japonica* DSM 44213^T, and strains *A. mediterranei,* respectively, share similar BGCs organization, with most of the BGCs located in the non-core regions in agreement with Xu et al. (2014).

Despite the large number of BGCs detected in our strains in study, only four of them can be highly correlated to known families of compounds, with a high prevalence of BGCs showing low homology to any annotated cluster.

CONCLUDING REMARKS

The sequencing and comparative analysis of *Amycolatopsis* CA-126428 and CA-128772 genomes has contributed to improve our understanding of the taxonomic diversity and metabolic potential of the species of this genus. Both MLSA and 16S rRNA phylogenies consistently show that strains CA-126428 and CA-128772 belong to the group C of the AOS subclade. The different relative position of the strains in the 16S rRNA and MLSA phylogenies, as well as the differences observed from the genomic comparisons, suggest that strains CA-126428 and CA-128772 may represent new species, requiring further investigations to explore the uniqueness and the role of these new strains in the lichen associated microbial community.

All the *Amycolatopsis* strains analyzed have shown a large genomic potential to produce different classes of specialized metabolites restricted to few groups of species. In addition, the results of our analysis support previous reports suggesting actinomycetes as a still untapped source of novel compounds and the relevance in modern natural products drug discovery of the application of genome-based mining approaches of these species to foster the discovery of new natural products encoded by cryptic or poorly expressed BGCs.

AUTHOR CONTRIBUTIONS

MS-H and OG designed the experiments; MS-H, IG, and CD-M performed the experiments; GM configured the bioinformatic software; and MS-H and OG wrote the manuscript.

FUNDING

This study was supported by Fundación MEDINA, Centro de Excelencia en Medicamentos Innovadores en Andalucía, Granada, Spain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00369/full#supplementary-material

REFERENCES

Adamek, M., Spohn, M., Stegmann, E., and Ziemert, N. (2017). "Mining bacterial genomes for secondary metabolite gene clusters," in *Antibiotics, Methods in Molecular Biology*, ed P. Sass (New York, NY: Humana Press), 23–47.

Aggarwal, G., and Ramaswamy, R. (2002). *Ab initio* gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *J. Biosci*. 27, 7–14. doi: 10.1007/BF02703679

Albarracin, V. H., Alonso-Vega, P., Trujillo, M. E., Amoroso, M. J., and Abate, C. M. (2010). *Amycolatopsis tucumanensis* sp. nov., a copper-resistant actinobacterium isolated from polluted sediments. *Int. J. Syst. Evol. Microbiol.* 60, 397–401. doi: 10.1099/ijs.0.010587-0

Aschenbrenner, I. A., Cernava, T., Berg, G., and Grube, M. (2016). Understanding microbial multi-species symbioses. *Front. Microbiol.* 7:180. doi: 10.3389/fmicb.2016.00180

Auch, A. F., von Jan, M., Klenk, H. P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-vaillant, N., Jacquard, C., Klenk, H. P., et al. (2016). Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev.* 80, 1–43. doi: 10.1128/MMBR.00019-15

Beemelmanns, C., Ramadhar, T. R., Kim, K. H., Klassen, J. L., Cao, S., Wyche, T. P., et al. (2017). Macrotermycins A–D, glycosylated macrolactams from a termite-associated *Amycolatopsis* sp. M39. *Org. Lett*. 19, 1000–1003. doi: 10.1021/acs.orglett.6b03831

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., et al. (2017). antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 1–6. doi: 10.1093/nar/gkx319

Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2001). Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* 8, 713–723. doi: 10.1016/S1074-5521(01)00046-1

Calcott, M. J., Ackerley, D. F., Knight, A., Keyzers, R. A., and Owen, J. G. (2017). Secondary metabolism in the lichen symbiosis. *Chem. Soc. Rev.* doi: 10.1039/c7cs00431a. [Epub ahead of print].

Cardinale, M., Puglia, A. M., and Grube, M. (2006). Molecular analysis of lichen-associated bacterial communities. *FEMS Microbiol. Ecol.* 57, 484–495. doi: 10.1111/j.1574-6941.2006.00133.x

Chen, S., Wu, Q., Shen, Q., and Wang, H. (2016). Progress in understanding the genetic information and biosynthetic pathways behind *Amycolatopsis* antibiotics, with implications for the continued discovery of novel drugs. *ChemBioChem.* 17, 119–128. doi: 10.1002/cbic.201500542

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., et al. (2014). Insights into secondary metabolism from a

global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158, 412–421. doi: 10.1016/j.cell.2014.06.034

Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Peter Gogarten, J., and Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. *MBio* 5, 1–13. doi: 10.1128/mBio.02136-14

Darling, A. E., Mau, B., and Perna, N. T. (2010). ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLOS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147

Dávila Costa, J. S., and Amoroso, M. J. (2014). Current biotechnological applications of the genus *Amycolatopsis*. *World J. Microbiol. Biotechnol*. 30, 1919–1926. doi: 10.1007/s11274-014-1622-3

Everest, G. J., and Meyers, P. R. (2009). The use of *gyrB* sequence analysis in the phylogeny of the genus *Amycolatopsis*. *Antonie Van Leeuwenhoek* 95, 1–11. doi: 10.1007/s10482-008-9280-9

Everest, G. J., and Meyers, P. R. (2011). Evaluation of the antibiotic biosynthetic potential of the genus *Amycolatopsis* and description of *Amycolatopsis circi* sp. nov., *Amycolatopsis equina* sp. nov. and *Amycolatopsis hippodromi* sp. nov. *J. Appl. Microbiol*. 111, 300–311. doi: 10.1111/j.1365-2672.2011.05058.x

Foulerton, A. (1905). New species of *Streptothrix* isolated from the air. *Lancet* 1, 1199–1200.

Genilloud, O. (2017a). Actinomycetes: still a source of novel antibiotics. *Nat. Prod. Rep*. 34, 1203–1232. doi: 10.1039/C7NP00026J

Genilloud, O. (2017b). "Physiology of actinobacteria," in *Biology and Biotechnology of Actinobacteria,* eds J. Wink, F. Mohammadipanah, and J. Hamedi (Cham: Springer), 151–180.

Genilloud, O., González, I., Salazar, O., Martín, J., Tormo, J. R., and Vicente, F. (2011). Current approaches to exploit actinomycetes as a source of novel natural products. *J. Ind. Microbiol. Biotechnol.* 38, 375–389. doi: 10.1007/s10295-010-0882-7

Giessen, T. W., Franke, K. B., Knappe, T. A., Kraas, F. I., Bosello, M., Xie, X., et al. (2012). Isolation, structure elucidation, and biosynthesis of an unusual hydroxamic acid ester-containing siderophore from *Actinosynnema mirum*. *J. Nat. Prod.* 75, 905–914. doi: 10.1021/np300046k

Glaeser, S. P., and Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst. Appl. Microbiol*. 38, 237–245. doi: 10.1016/j.syapm.2015.03.007

González, I., Ayuso-Sacido, A., Anderson, A., and Genilloud, O. (2005). Actinomycetes isolated from lichens: evaluation of their diversity and detection of biosynthetic gene sequences. *FEMS Microbiol. Ecol*. 54, 401–415. doi: 10.1016/j.femsec.2005.05.004

Gunnarsson, N., Mortensen, U. H., Sosio, M., and Nielsen, J. (2004). Identification of the Entner–Doudoroff pathway in an antibiotic-producing actinomycete species. *Mol. Microbiol.* 52, 895–902. doi: 10.1111/j.1365-2958.2004.04028.x

Guy, L., Kultima, J. R., Andersson, S. G. E., and Quackenbush, J. (2011). GenoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 27, 2334–2335. doi: 10.1093/bioinformatics/btq413

Hamedi, J., Mohammadipanah, F., and Ventosa, A. (2013). Systematic and biotechnological aspects of halophilic and halotolerant actinomycetes. *Extremophiles* 17, 1–13. doi: 10.1007/s00792-012-0493-5

Hashizume, H., Sawa, R., Yamashita, K., Nishimura, Y., and Igarashi, M. (2017). Structure and antibacterial activities of new cyclic peptide antibiotics, pargamicins B, C and D, from *Amycolatopsis* sp. ML1-hF4. *J. Antibiot. (Tokyo)*. 70, 699–704. doi: 10.1038/ja.2017.34

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016a). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., and Morishima, K. (2016b). 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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). "Preparation and analysis of genomic and plasmid DNA," in *Practical Streptomyces Genetics*, Vol. 1, eds T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood (Norwich: John Innes Foundation), 161–210.

Klassen, J. L., and Currie, C. R. (2012). Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. *BMC Genomics* 13:14. doi: 10.1186/1471-2164-13-14

Kodani, S., Bicz, J., Song, L., Deeth, R. J., Ohnishi-Kameyama, M., Yoshida, M., et al. (2013). Structure and biosynthesis of scabichelin, a novel tris-hydroxamate siderophore produced by the plant pathogen *Streptomyces scabies* 87.22. *Org. Biomol. Chem*. 11, 4686–4694. doi: 10.1039/c3ob40536b

Kodani, S., Komaki, H., Suzuki, M., Hemmi, H., and Ohnishi-Kameyama, M. (2015). Isolation and structure determination of new siderophore albachelin from *Amycolatopsis alba*. *Biometals* 28, 381–389. doi: 10.1007/s10534-015-9842-z

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr. Opin. Microbiol*. 10, 504–509. doi: 10.1016/j.mib.2007.08.006

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

Kumari, R., Singh, P., and Lal, R. (2016). Genetics and genomics of the genus *Amycolatopsis*. *Indian J. Microbiol*. 56, 233–246. doi: 10.1007/s12088-016-0590-8

Kwakman, J. H., and Postma, P. W. (1994). Glucose kinase has a regulatory role in carbon catabolite repression in *Streptomyces coelicolor*. *J. Bacteriol.* 176, 2694–2698. doi: 10.1128/jb.176.9.2694-2698.1994

Lane, D. J. (1991). "16S/23S rRNA sequencing," in *Nucleic Acid Techniques in Bacterial Systematics*, eds E. Stackebrandt and M. Goodfellow (New York, NY: John Wiley & Sons), 115–175.

Lechevalier, M. P., Prauser, H., Labeda, D. P., and Ruan, J. S. (1986). Two new genera of nocardioform actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov. *Int. J. Syst. Bacteriol.* 36, 29–37. doi: 10.1099/00207713-36-1-29

Lee, I., Kim, Y. O., Park, S. C., and Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol*. 66, 1100–1103. doi: 10.1099/ijsem.0.000760

Li, W., Chou, S., Khullar, A., and Gerratana, B. (2009). Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an SJG-136 monomeric analog. *Appl. Environ. Microbiol.* 75, 2958–2963. doi: 10.1128/AEM.02325-08

Liu, C., Jiang, Y., Wang, X., Chen, D., Chen, X., Wang, L., et al. (2017). Diversity, antimicrobial activity, and biosynthetic potential of cultivable actinomycetes associated with lichen symbiosis. *Microb. Ecol*., 74, 570–584. doi: 10.1007/s00248-017-0972-4

Loman, N. J., and Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* 13, 1–9. doi: 10.1038/nrmicro3565

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol*. 11, 625–631. doi: 10.1038/nchembio.1890

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Klenk, H. P., and Göker, M. (2014). Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol*. 64, 352–356. doi: 10.1099/ijs.0.056994-0

Nett, M., Ikeda, H., and Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep*. 26, 1362–1384. doi: 10.1039/b817069j

Parrot, D., Antony-Babu, S., Intertaglia, L., Grube, M., Tomasi, S., and Suzuki, M. T. (2015). Littoral lichens as a novel source of potentially bioactive actinobacteria. *Sci. Rep*. 5:15839. doi: 10.1038/srep15839

Salas, J. A., Quirós, L. M., and Hardisson, C. (1984). Pathways of glucose catabolism during germination of *Streptomyces* spores. *FEMS Microbiol. Lett.* 22, 229–233. doi: 10.1111/j.1574-6968.1984.tb00732.x

Sasaki, E., Ogasawara, Y., and Liu, H. W. (2010). A biosynthetic pathway for BE-7585A, a 2-thiosugar-containing angucycline-type natural product. *J. Am. Chem. Soc.* 132, 7405–7417. doi: 10.1021/ja1014037

Scherlach, K., and Hertweck, C. (2017). Mediators of mutualistic microbe–microbe interactions. *Nat. Prod. Rep*. doi: 10.1039/C7NP00035A. [Epub ahead of print].

Sentausa, E., and Fournier, P. E. (2013). Advantages and limitations of genomics in prokaryotic taxonomy. *Clin. Microbiol. Infect*. 19, 790–795. doi: 10.1111/1469-0691.12181

Seyedsayamdost, M. R., Traxler, M. F., Zheng, S. L., Kolter, R., and Clardy, J. (2011). Structure and biosynthesis of amychelin, an unusual mixed-ligand siderophore from *Amycolatopsis* sp. AA4. *J. Am. Chem. Soc*. 133, 11434–11437. doi: 10.1021/ja203577e

Sharma, M., Dangi, P., and Choudhary, M. (2014). Actinomycetes: source, identification, and their applications. *Int. J. Curr. Microbiol. App. Sci*. 3, 801–832. Available online at: https://www.ijcmas.com/Archives-15.php

Skellam, E. J., Stewart, A. K., Strangman, W. K., and Wright, J. L. (2013). Identification of micromonolactam, a new polyene macrocyclic lactam from two marine *Micromonospora* strains using chemical and molecular methods: clarification of the biosynthetic pathway from a glutamate starter unit. *J. Antibiot. (Tokyo)*. 66, 431–441. doi: 10.1038/ja.2013.34

Spohn, M., Kirchner, N., Kulik, A., Jochim, A., Wolf, F., Muenzer, P., et al. (2014). Overproduction of ristomycin A by activation of a silent gene cluster in *Amycolatopsis japonicum* MG417-CF17. *Antimicrob. Agents Chemother*. 58, 6185–6196. doi: 10.1128/AAC.03512-14

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kampfer, P., Maiden, M. C. J., et al. (2002). Report of the *ad hoc* comittee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol*. 52, 1043–1047. doi: 10.1099/ijs.0.02360-0.02360

Tang, B., Xie, F., Zhao, W., Wang, J., Dai, S., Zheng, H., et al. (2016). A systematic study of the whole genome sequence of *Amycolatopsis methanolica* strain 239^T provides an insight into its physiological and taxonomic properties which correlate with its position in the genus. *Synth. Syst. Biotechnol*. 1, 169–186. doi: 10.1016/j.synbio.2016.05.001

Thaker, M. N., Wang, W., Spanogiannopoulos, P., Waglechner, N., King, A. M., Medina, R., et al. (2013). Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat. Biotechnol*. 31, 922–927. doi: 10.1038/nbt.2685

Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., and Thompson, F. L. (2013). Microbial genomic taxonomy. *BMC Genomics* 14:913. doi: 10.1186/1471-2164-14-913

Tohyama, S., Kakinuma, K., and Eguchi, T. (2006). The complete biosynthetic gene cluster of the 28-membered polyketide macrolactones, halstoctacosanolides, from *Streptomyces halstedii* HC34. *J. Antibiot. (Tokyo)* 59, 44–52. doi: 10.1038/ja.2006.7

Truman, A. W., Kwun, M. J., Cheng, J., Yang, S. H., Suh, J. W., and Hong, H. J. (2014). Antibiotic resistance mechanisms inform discovery: identification and characterization of a novel *Amycolatopsis* strain producing ristocetin. *Antimicrob. Agents Chemother*. 58, 5687–5695. doi: 10.1128/AAC.03349-14

Völler, G. H., Krawczyk, J. M., Pesic, A., Krawczyk, B., Nachtigall, J., and Süssmuth, R. D. (2012). Characterization of new class III lantibiotics-erythreapeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea, Streptomyces avermitilis* and *Streptomyces griseus* demonstrates stepwise N-terminal leader processing. *ChemBioChem* 13, 1174–1183. doi: 10.1002/cbic.201200118

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S Ribosomal DNA amplification for phylogenetic study. *J. Biotechnol*. 173, 697–703. doi: 10.1128/jb.173.2.697-703.1991

Xiao, Y. S., Zhang, B., Zhang, M., Guo, Z. K., Deng, X. Z., Shi, J., et al. (2017). Rifamorpholines A–E, potential antibiotics from locust-associated actinobacteria *Amycolatopsis* sp. Hca4. *Org. Biomol. Chem*. 15, 3909–3916. doi: 10.1039/C7OB00614D

Xu, L., Huang, H., Wei, W., Zhong, Y., Tang, B., Yuan, H., et al. (2014). Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*. *BMC Genomics* 15:363. doi: 10.1186/1471-2164-15-363

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol*. 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Yurkovich, M. E., Jenkins, R., Sun, Y., Tosin, M., and Leadlay, P. F. (2017). The polyketide backbone of thiolactomycin is assembled by an unusual iterative polyketide synthase. *Chem. Commun*. 53, 2182–2185. doi: 10.1039/C6CC09934C

Zhao, F., Qin, Y. H., Zheng, X., Zhao, H. W., Chai, D. Y., Li, W., et al. (2016). Biogeography and adaptive evolution of *Streptomyces* strains from saline environments. *Sci. Rep*. 6:32718. doi: 10.1038/srep32718

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 Sánchez-Hidalgo, González, Díaz-Muñoz, Martínez and Genilloud. 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: 05 January 2018
doi: 10.3389/fmicb.2017.02613

Phylogeny of *Vibrio vulnificus* from the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy

*Francisco J. Roig[1,2,3], Fernando González-Candelas[4,5], Eva Sanjuán[1,2], Belén Fouz[1,2], Edward J. Feil[6], Carlos Llorens[3], Craig Baker-Austin[7], James D. Oliver[8,9], Yael Danin-Poleg[10], Cynthia J. Gibas[11], Yechezkel Kashi[10], Paul A. Gulig[12], Shatavia S. Morrison[11] and Carmen Amaro[1,2]**

[1] Estructura de Investigación Interdisciplinar en Biotecnología y Biomedicina BIOTECMED, University of Valencia, Valencia, Spain, [2] Departmento de Microbiología y Ecología, Universidad de Valencia, Valencia, Spain, [3] Biotechvana, Parc Cientific, Universitat de Valencia, Valencia, Spain, [4] Joint Research Unit on Infection and Public Health FISABIO-Salud Pública and Universitat de Valencia-I2SysBio, Valencia, Spain, [5] CIBEResp, National Network Center for Research on Epidemiology and Public Health, Instituto de Salud Carlos III, Valencia, Spain, [6] Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom, [7] Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom, [8] Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States, [9] Duke University Marine Lab, Beaufort, NC, United States, [10] Faculty of Biotechnology and Food Engineering, Technion–Israel Institute of Technology, Haifa, Israel, [11] Department of Bioinformatics and Genomics, the University of North Carolina at Charlotte, Charlotte, NC, United States, [12] Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Javier Pascual, German Collection of Microorganisms and Cell Cultures (LG), Germany
Karla Satchell, Feinberg School of Medicine, Northwestern University, United States

***Correspondence:**
Carmen Amaro
carmen.amaro@uv.es

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

Received: *13 October 2017*
Accepted: *14 December 2017*
Published: *05 January 2018*

Citation:
Roig FJ, González-Candelas F, Sanjuán E, Fouz B, Feil EJ, Llorens C, Baker-Austin C, Oliver JD, Danin-Poleg Y, Gibas CJ, Kashi Y, Gulig PA, Morrison SS and Amaro C (2018) Phylogeny of Vibrio vulnificus from the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy. Front. Microbiol. 8:2613. doi: 10.3389/fmicb.2017.02613

Vibrio vulnificus (Vv) is a multi-host pathogenic species currently subdivided into three biotypes (Bts). The three Bts are human-pathogens, but only Bt2 is also a fish-pathogen, an ability that is conferred by a transferable virulence-plasmid (pVvbt2). Here we present a phylogenomic analysis from the core genome of 80 Vv strains belonging to the three Bts recovered from a wide range of geographical and ecological sources. We have identified five well-supported phylogenetic groups or lineages (L). L1 comprises a mixture of clinical and environmental Bt1 strains, most of them involved in human clinical cases related to raw seafood ingestion. L2 is formed by a mixture of Bt1 and Bt2 strains from various sources, including diseased fish, and is related to the aquaculture industry. L3 is also linked to the aquaculture industry and includes Bt3 strains exclusively, mostly related to wound infections or secondary septicemia after farmed-fish handling. Lastly, L4 and L5 include a few strains of Bt1 associated with specific geographical areas. The phylogenetic trees for ChrI and II are not congruent to one another, which suggests that inter- and/or intra-chromosomal rearrangements have been produced along Vv evolution. Further, the phylogenetic trees for each chromosome and the virulence plasmid were also not congruent, which also suggests that pVvbt2 has been acquired independently by different clones, probably in fish farms. From all these clones, the one with zoonotic capabilities (Bt2-Serovar E) has successfully spread worldwide. Based on these results, we propose a new updated classification of the species based on phylogenetic lineages rather than on Bts, as well as the inclusion of all Bt2 strains in a pathovar with the particular ability to cause fish vibriosis, for which we suggest the name "piscis."

Keywords: microbial evolution, pathogens, SNP, *Vibrio vulnificus*, core genome, virulence plasmid, pathovar, biotype

INTRODUCTION

Vibrio vulnificus is an emerging zoonotic pathogen that inhabits brackish water ecosystems from temperate and tropical areas and whose geographical distribution has recently extended to Northern countries due to global warming (Baker-Austin et al., 2013, 2017). The pathogen survives in water, either associated with the mucous surfaces of algae and aquatic animals or as a free-living bacterium that can be concentrated by filtering organisms such as oysters (Oliver, 2015).

V. vulnificus was defined as a bacterial species in 1976 (Farmer, 1979) and was later split into three biotypes (Bt) on the basis of differences in genotypic and phenotypic traits as well as in host range (Tison et al., 1982; Bisharat et al., 1999). The three Bts are opportunistic human pathogens, but Bt2 is also pathogenic for aquatic animals. The zoonotic strains belong to the same serovar (Ser) and are classified as Bt2-SerE (Biosca et al., 1997).

The various diseases caused by this species are known as vibriosis. Human vibriosis presents as two main forms depending on the pathogen's route of entry into the body, ingestion or contact (Oliver, 2015). In the first case, the pathogen is ingested with raw seafood, colonizes the intestine and causes gastroenteritis and/or primary septicemia. In the second case, the pathogen crosses the skin barrier during an injury or directly colonizes a preexisting wound causing local but severe necrosis and/or secondary septicemia. The main factor that predisposes to death by sepsis is a high level of serum iron as a consequence of multiple pathologies (e.g., hemochromatosis, diabetes, cirrhosis, and viral hepatitis) (Oliver, 2015).

Epidemiological data on human vibriosis from the Centers for Disease Control and Prevention (CDC) estimate that around 80,000 people are infected by *Vibrio* spp. each year in the USA and that, of these, *V. vulnificus* is responsible for most of the fatal cases (case fatality rate >50% for septicemia; Jones and Oliver, 2009). Thus, *V. vulnificus* is responsible for over 95% of seafood-related deaths (Jones and Oliver, 2009), the highest fatality rate of any known food-borne pathogen (Rippey, 1994). In addition, increasing incidents of infections are occurring globally, with cases reported in Europe and Asia (Hlady and Klontz, 1996; Baker-Austin et al., 2013; Lee et al., 2013b). Crucially, infections currently appear to be increasing in both the USA and in Europe (Newton et al., 2012; Baker-Austin et al., 2013).

Regarding fish vibriosis, eels seem to be the most susceptible host, especially under farming conditions (Amaro et al., 2015). The pathogen colonizes the eel gills, enters the blood and causes death by septicemia, even in healthy individuals (Marco-Noales et al., 2001). Eel vibriosis occurs in farms as epizootics or outbreaks of high mortality that can lead to the closure of the farm if the disease is not controlled promptly. Epizootiological data suggest that outbreaks of eel vibriosis have been registered in all of the countries where eels are cultured in brackish waters, including fish farms located in Northern-European countries (Haenen et al., 2013).

The genetic basis for human virulence is only partially known, although most studies suggest that all strains of the species may have the ability to infect humans regardless of their origin, clinical, or environmental (Gulig et al., 2005). In contrast, the ability to infect fish is dependent on a virulence plasmid (pVvBt2) that is only present in Bt2 strains (Lee et al., 2008; Roig and Amaro, 2009). The plasmid encodes resistance to the innate immunity of eels, and probably other teleost (Lee et al., 2008, 2013a; Pajuelo et al., 2015). Interestingly, pVvbt2 can be transmitted among bacteria aided by a second, conjugative plasmid, which is widespread in the species (Lee et al., 2008).

The aim of this study is to describe the phylogenetic groups within the species and compare them with the current intraspecific groups from the characterization of the core genome of species (CGS), the core genome of plasmid pVvbt2 (CGP) and the core genome of human virulence-related genes (CGV). Our results highlight the importance of aquaculture industry in the recent evolution and epidemic spread of the species and support the intra-specific classification in lineages instead of Bts, as well as the inclusion of a pathovar grouping all fish virulent isolates for which we propose the name "piscis."

MATERIALS AND METHODS

Bacterial Isolates, Culture Conditions, DNA Extraction, and Sequencing

The genomes used in this study and the main features of the corresponding strains are detailed in **Table 1**. The strains whose genome was sequenced are marked in **Table 1**. These strains were routinely grown in Tryptone Soy Broth or agar plus 5 g/l NaCl (TSB-1 or TSA-1, Pronadisa, Spain) at 28°C for 24 h. The strains were maintained both as lyophilized stocks and as frozen stocks at −80°C in marine broth (Difco) plus 20% (v/v) glycerol.

DNA was extracted using GenElute™ Bacterial Genomic DNA (Sigma, Spain) from bacteria grown with shaking at 28°C for 12 h. Samples with a DNA concentration of 10–15 ng/μl were used for sequencing with Illumina Genome Analyzer technology GAII (Illumina MiSeq) flow cell in the Genome Analysis Centre in Norwich (UK) and the SCSIE of the University of Valencia (Spain). To this end, unique index-tagged libraries for each sample (up to 96 strains) were created using TruSeq DNA Sample Preparation for subsequent cluster generation (Illumina cBot), and up to 12 separate libraries were sequenced in each of eight channels in Illumina Genome Analyser GAII cells with 100-base paired-end reads. The index-tag sequence information was used for downstream processing to assign reads to the individual samples (Harris et al., 2010).

Genomes Selected as Reference

The genomes of the Bt1 strain YJ016 (NC_005139 and NC_005140) and the plasmids pR99 (AM293858) and p4602-2 (AM293860) were used as templates for all the genomic analysis. The genome of the strain YJ016 was selected because it is one of the few genomes of *V. vulnificus* that has been accurately closed and annotated (Chen et al., 2003). The YJ016 genome contains 5,097 genes (3,387 in chromosome I and 1,710 in chromosome II), pR99 contains 71 genes and p4602-2 contains 67 genes.

TABLE 1 | Origin, year of isolation, biotype, serovar, virulence-related typing, and genome accession number of *V. vulnificus* strains used in this study.

Strain	Origin	Country	Isolation Year	Biotipe Serovar Clade[a]	vvpdh[b]	vcg[b]	Sequence accession
32	Human blood	Israel	1997?	3/O	+	E	SAMN07739065
162	Human blood	Israel	1997	3/O	+	E	SAMN07739066
2322	Fish pound water	Israel	1997	1	–	E	GCA_000743165.1
491771	Human blood	Israel	1997	3	+	E	GCA_000743115.1
106-2A*; ***(V Harwood lab)	Oyster	U.S.A.	<2011	1	–	C	SAMN07739068
11028***	Human wound	Israel	1997	3/O	+	E	GCA_002074875.1
12***	Health tilapia	Israel	2002	3/O	+	E	GCA_002074885.1
93U204	Diseased tilapia	Taiwan	No reported	1	+	C	GCA_000746665.1
94385* (C Amaro lab)	Human wound	Spain	2001	1	+	E	SAMN07739067
94-8-112*,**,*** (C Amaro lab)	Human wound	Denmark	1994	2/E	+	E	SAMN07739076
94-9-119* (C Amaro lab)	Human wound	Denmark	1994	1	+	E	SAMN07739069
95-8-161*,** (C Amaro lab)	Diseased eel	Denmark	1995	2/I	+	E	SAMN07739070
95-8-6*,** (C Amaro lab)	Diseased eel	Denmark	1995	2/I	–	E	SAMN07739071
95-8-7*,** (C Amaro lab)	Diseased eel	Denmark	1995	2/I	+	E	SAMN07739072
960426-1_4C*,**	Diseased eel	Denmark	1996	2/NT	+	E	SAMN07739073
99-578_DP-B1	Oyster	U.S.A.	1998	1	–	C	GCA_000788325.1
99-796_DP-E7	Oyster	U.S.A.	1998	1	–	E	GCA_000788315.1
A14*,** (C Amaro lab)	Diseased eel	Spain	2002	2/A	–	E	SAMN07739074
AB17-319*,*** (JD Oliver lab)	Oyster	U.S.A.	2005	1	–	C/E	SAMN07739075
ATCC_27562^T (type strain)	Human blood	U.S.A.	No reported	1	–	E	GCA_000299635.1
ATCC_29306	Human wound	U.S.A.	No reported	1	+	E	GCA_001471415.1
ATCC_29307	Human blood	U.S.A.	No reported	1	–	C	GCA_001471465.1
ATCC_33147	Diseased eel	Japan	1979	2/E	+	E	GCA_000764895.1
ATCC_43382	Human blood	U.S.A.	No reported	1	–	E	GCA_001471305.1
ATL_6-1306	Human blood	U.S.A.	1996	1	+	C	GCA_000788335.1
ATL_71503	Human blood	U.S.A.	1996	1	–	E	GCA_000788345.1
B2	Human blood	China	2010	1	–	E	GCA_000303175.1
BAA87	Human wound	Israel	1996	3	+	E	GCA_000576265.1
C7184 (or CDC7184)*,***	Human blood	U.S.A.	1977	1	+	C	SAMN07739077
CECT4604*,**,*** (C Amaro lab)	Diseased eel	Spain	1990	2/E	+	E	SAMN07739078
CECT4606 *,*** (C Amaro lab)	Healthy eel	Spain	1990	1	+	E	LAXL00000000
CECT4608* (C Amaro lab)	Eel tank water	Spain	1990	1	+	C	SAMN07739079
CECT4865*,** (C Amaro lab)	Diseased shrimp	Taiwan	No reported	2/E	+	E	SAMN07739080
CECT4866*,** (C Amaro lab)	Human blood	Australia	1997	2/E	+	E	LABE00000000
CECT4999*, ** (C Amaro lab)	Diseased eel	Spain	1999	2/E	+	E	GCA_002215135.1
CECT5763*,**,*** (C Amaro lab)	Eel tank water	Spain	2002	2/E	+	E	LEAM00000000
CECT5769*,**,*** (C Amaro lab)	Diseased eel	Spain	2004	2/A	–	E	LABF00000000
CECT7030*,** (C Amaro lab)	Diseased eel	Denmark	2004	2/A	–	E	SAMN07739081
CECT898*,** (C Amaro lab)	Diseased eel	Japan	1979	2/E	+	E	SAMN07739082
CG100* (C Amaro lab)	Oyster	Taiwan	1993	1	+	C	SAMN07739083
CG64	Oyster	Taiwan	1993	1	+	C	GCA_000959775.1
CIP8190*, **,*** (C Amaro lab)	Human blood	France	1980	2/E	+	E	LAXM00000000
CMCP6***	Human blood	South Korea	<2003	1	+	C	GCA_000039765.1
E64MW	Human wound	No reported	No reported	1	+	E	GCA_000269745.1
ENV1*,*** (JD Oliver lab)	Oyster	U.S.A.	2005	1	–	E	SAMN07739084
FLA112 (ATL9824)*,*** (A dePaola lab)	Human blood	U.S.A.	1994	1	+	C	SAMN07739085
FLA144 (CDC 90-3095 or ORL 1506)*,*** (A dePaola lab)	Human blood	U.S.A.	1995	1	–	E	SAMN07739086
FORC_009	Stool sample	South Korea	2008	1	–	C	GCA_001433435.1
FORC_016	Human blood	South Korea	2009	1	–	C	GCA_001653775.1

(Continued)

TABLE 1 | Continued

Strain	Origin	Country	Isolation Year	Biotipe Serovar Clade[a]	vvpdh[b]	vcg[b]	Sequence accession
FORC_017	Human blood	South Korea	2014	1	+	C	GCA_001675245.1
JY1305	Oyster	U.S.A.	1999	1	−	E	GCA_000269725.1
JY1701	Oyster	U.S.A.	1999	1	+	E	GCA_000269765.1
LSU1015*,*** (JD Oliver lab)	Human wound	U.S.A.	<1998	1	−	C	PRJNA279176
LSU1657*,*** (JD Oliver lab)	Human wound	U.S.A.	<1998	1	+	E	SAMN07739088
LSU2098*,*** (JD Oliver lab)	Human wound	U.S.A.	<1998	1	−	E	SAMN07739089
MO6-24/O***	Human blood	South Korea	1986	1	+	C	GCA_000186585.1
NB-VV-101	Tilapia	Israel	1997	1	−		GCA_000743155.1
NV1	Seawater	Taiwan	2011	1	+	C	GCA_000959755.1
NV22	Seawater	Taiwan	2011	1	−	C	GCA_000960125.1[c]
PD-2-51*,**,*** (C Amaro lab)	Seawater	Spain	2003	2/E	+	E	SAMN07739090[c]
R02*,**,*** (C Amaro lab)	Diseased eel	Spain	2002	2/E	−	E	SAMN07739091[c]
Ra3*,** (C Amaro lab)	Diseased eel	Spain	2011	2/E	+	E	SAMN07739092[c]
Riu1*	Seawater	Spain	2003	1	−	E	SAMN07739093[c]
S2-22*,*** (V Harwood lab)	Water	U.S.A.	2004	1	+	C	SAMN07739094[c]
S3-16*,*** (V Harwood lab)	Water	U.S.A.	2005	1	+	C	SAMN07739095[c]
SC9613	Crab	South Korea	1996	1	−	E	GCA_000959745.1[c]
SC9629	Seafood	South Korea	1996	1	−	E	GCA_000967055.1[c]
SC9729	Seawater	South Korea	1997	1	−	E	GCA_000959825.1[c]
SC9740	Seawater	South Korea	1997	1	−	E	GCA_000959765.1[c]
SC9761	Oyster	South Korea	1997	1	−	E	GCA_000959835.1[c]
SC9794	Tidal mudflat	South Korea	1997	1	+	C	GCA_000959845.1[c]
SREL119*,*** (JD Oliver lab)	Sediment	U.S.A.	<2006	1	−	E	SAMN07739096[c]
SREL314*,*** (JD Oliver lab)	Water	U.S.A.	<2006	1	−	E	SAMN07739097[c]
SS108-A3A*,*** (JD Oliver lab)	Oyster	U.S.A.	2005	1	−	E	SAMN07739098[c]
V252	Human blood	Israel	2004	1/clade B	−	E	GCA_001277815.1[c]
VV4-03	Human blood	Israel	2003	3	+	E	GCA_000743095.1[c]
VV9-09	Human blood	Israel	1999	3	+	E	GCA_000743105.1
vvyb1***	Healthy Tilapia	Israel	2004	3/O	+	E	GCA_000342305.2[c]
yb158***	Healthy Tilapia	Israel	2005	1/clade A	+	C	GCA_001013325.1
YJ016 ***	Human blood	Taiwan	1993	1	+	C	GCA_000009745.1

**Strains whose genomes were sequenced in this study. The laboratory that purchased the strain is indicated in parenthesis. **Strains used for virulence plasmid analysis. ***Strains used for Vibrio species analysis.*

[a]O-antigen serovar was determined for Bt2 and 3 isolates according to Biosca et al. (1997): Clade A and B were described by Danin-Poleg et al. (2015) and Efimov et al. (2015), respectively. NT, non typable.

[b]vvpdh (V. vulnificus potentially dangerous for humans); pilF polymorphism associated with human virulence (Roig et al., 2010): vcg (virulence correlated gene: E, environmental type, C, clinical type, C/E, clinical and environmental type) (Rosche et al., 2005).

[c]Waiting for definitive accession.

Genome Sequence Assembly

Reads for each genome were done using SPAdes version 3.6.1 (Bankevich et al., 2012) with kmers from 21 to 127 and the careful option to reduce mismatches and short indels. Then, multiple sequence alignments were obtained by using Progressive Mauve software with default options (Darling et al., 2004). Locally Collinear Blocks (LCBs) with a size larger than 1 kb that were present in all the genomes were used to define the core genomes. The selected LCBs were concatenated to be used in subsequent analyses to build a core genome multiple alignment.

Core Genome

We selected seven *Vibrio* species that had at least one fully sequenced and annotated genome to define the core genome of the genus (CGG). **Table S1** summarizes the characteristics of the closed *Vibrio* genomes selected for the study.

The identification of each gene was performed using local BLAST searches (Wang et al., 2003). Three independent searches, one per chromosome and plasmid, were performed, and a database was generated for each genome. The resulting sequences were mapped onto reference genes using Geneious 6.1.6 (Biomatters) software. As the sequences used for the

generation of the databases were not closed genomes (draft genomes/contigs), only the genes present in all the genomes, with a minimum length of 80% of the total length of the homologous gene in the reference genome and with a DNA identity higher than 70%, were included in the CGS, the CGP or the CGG for further analysis.

In order to refine the preliminary alignment for the entire core, individual alignments for each gene were performed using the program MAFFT (Katoh and Standley, 2013). The elimination of unaligned ends and genes that did not match the above requirements was performed using an *in-house* Python script. Once all the genes were aligned, a concatenated sequence was generated for the CGS, the CGP and the CGG.

Functional Analysis of the CGS

Functional annotation was performed using GPRO (Futami et al., 2011). Gene descriptions were obtained by blasting the predicted protein sequences against the NCBI non-redundant proteins (NR) and the Clusters of Orthologous Groups (COGs) databases (Tatusov, 2000) as reference subjects. Protein accessions obtained from the annotation based on NR were used to add additional Gene Ontology (GO) (Gene Ontology Consortium, 2008) and Enzyme Commission (EC) (Bairoch, 2000) designations. Information about metabolic pathways was retrieved via the web from the KEGG database (Kotera et al., 2012) based on EC numbers relation.

Single Nucleotide Polymorphisms (SNPs) Identification

SNPs were identified in the genomes from the coding regions as described previously (Harris et al., 2010) with appropriate SNP cutoffs to minimize the number of false-positive/negative calls. SNPs were filtered to remove those at sites with a SNP quality score lower than 40 and/or with read coverage below 25X in this region. SNPs at sites with heterogeneous mappings were also filtered out if the SNP was present in fewer than 85% of reads for that position.

Virulence Genes to Define the Human Virulence-Related Core Genome (VCGS)

We identified human virulence-related genes in the CGS from previously published data to define the human virulence-related core genome (VCGS) (Chakrabarti et al., 1999; Chen et al., 2004; Horstman et al., 2004; Gulig et al., 2005; Bogard and Oliver, 2007; Alice et al., 2008; Oh et al., 2008; Brown and Gulig, 2009; Liu et al., 2009; Lee et al., 2010; Chen and Chung, 2011; Kim et al., 2011).

Phylogenetic Reconstruction and Congruence Analysis

All the phylogenetic trees for CGG, CGS, CGP, and VCGS were reconstructed using the maximum-likelihood (ML) method with PhyML software (Guindon et al., 2009). The best evolutionary model for each dataset was determined with jModelTest (Posada, 2008) and considering the Akaike information criterion (AIC) (Akaike, 1974). The selected models were the generalized time-reversible model (GTR) with Gamma (+G) distribution and invariant positions (+I) (Tavaré, 1986) for SNP analysis. The pVvBt2 phylogeny was evaluated using the Hasegawa-Kishino-Yano model (Hasegawa et al., 1985). Support for the nodes derived in these reconstructions was evaluated by bootstrapping using 1,000 replicates (Felsenstein, 1985).

The congruence among the different phylogenetic reconstructions was evaluated using Shimodaira–Hasegawa (SH) (Shimodaira and Hasegawa, 1999) and expected-likelihood weight (ELW) tests as implemented in TreePuzzle version 5.2 (Schmidt et al., 2002; Strimmer and Rambaut, 2002).

RESULTS

Core genome

We obtained almost complete genome sequences of 38 *V. vulnificus* strains (**Table 1**). The CGS was inferred from these genomes and additional genomes taken from the database (80 in total). The analysis included strains of the three Bts, isolated from a variety of hosts and habitats all over the world (**Table 1**). The main characteristics of the analyzed genomes and the CGS as well as the number of SNPs identified per chromosome are detailed in **Tables 2**, **3**, respectively. The average gene identities in the CGS were 91 and 89% for chromosome I and II, respectively, which supports previous observations pertaining to the greater variability of chromosome II in *Vibrio* spp. (Kirkup et al., 2010).

The genes of the CGS and the associated metabolic pathways are shown in **Tables S2**, **S3**, respectively. The genes present in all the strains that did not match the criteria used to define the CGS (spanning less than 80% of the length and showing less than 70% identity with respect to the homologous gene in the reference YJ016 genome) are shown in **Table S4**. The CGS includes practically all of the genes for glycolysis, TCA cycle and pentose phosphate pathway, aerobic and anaerobic respiration, nitrate respiration, as well as for biosynthesis of metabolic intermediates, cofactors, nucleotides, amino acids and cell building blocks (**Tables S2**, **S3**). The CGS also includes genes involved in survival in different environments (water, animal intestine, chitinous surfaces), such as genes for resistance to different stressors (e.g., cold, heat, toxic oxygen forms, antimicrobials, and tellurite), genes for chitin degradation (i.e., various chitinases) and genes for surface colonization [e.g., pilus MSHA (mannose sensitive hemagglutination) and flagellum biogenesis]. It is generally assumed that genes on the *Vibrio* chromosome II have specific environmental functions related to habitat adaptation (Xu et al., 2003).

TABLE 2 | Some general data of the genome and core genome of the species (CGS) *V. vulnificus*.

	Average Genome size in bp (number of genes)	Core			
		Size (bp)	Genes	SNP	%G+C
Chromosome I	3,286,174 (3,103)	1,396,961	1,304	132,027	46.9
Chromosome II	1,803,986 (1,571)	519,724	414	45,027	46.8
Chromosome I+II	5,090,160 (4,674)	1,916,685	1,718	177,054	46.9

TABLE 3 | Number of SNPs sites per lineages identified in the core genome of the species (CGS).

Lineage	Chromosome I			Chromosome II		
	Specific SNPs	% Specific SNPs of total SNPs	% Specific SNPs of Core Genome	Specific SNPs	% Specific SNPs of total SNPs	% Specific SNPs of Core Genome
L1	83,107	62.95	5.95	28,193	62.61	5.42
L2	63,308	47.95	4.53	21,870	48.57	4.21
L3	25,034	18.96	1.79	9,619	21.36	1.85
L4	15,989	12.11	1.14	6,090	13.53	1.17
L5	132,027	100	9.45	45,027	100	8.66
L1-L2	117,305	88.85	8.4	39,924	88.67	7.68
L1-L3	91,163	69.05	6.53	30,933	68.7	5.95
L1-L4	89,724	67.96	6.42	30,230	67.14	5.82
L1-L5	87,874	66.56	6.29	29,778	66.13	5.73
L2-L3	74,594	56.5	5.34	26,214	58.22	5.04
L2-L4	85,808	64.99	6.14	29,097	64.62	5.6
L2-L5	71,527	54.18	5.12	24,968	55.45	4.8
L3-L4	52,708	39.92	3.77	18,246	40.52	3.51
L3-L5	34,277	25.96	2.45	12,514	27.79	2.41
L4-L5	42,081	31.87	3.01	13,676	30.37	2.63

Regarding the metabolic pathways, all of the common genes for general metabolic pathways such as butanoate metabolism, glycerolipid metabolism, peptidoglycan biosynthesis and pyrimidine metabolism are located on chromosome I while the genes for biosynthesis of siderophores, which could be related to habitat adaptation, are in the chromosome II but together with other metabolic genes not clearly related to habitat adaptation (i.e., fatty acid biosynthesis/elongation, glycerophospolipid metabolism, glyoxylate/dicarboxylate metabolism, nicotinate/nicotinamide metabolism, porphyrin metabolism, and valine/leucine/isoleucine metabolism) (**Table S3**). Regarding to survival genes, genes for the flagellum and MSHA pilus are located in chromosome I while those for the Flp/Tad pilus are located in chromosomes I and II (**Table S2**). In conclusion, there is no clear relationship between chromosome localization and habitat adaptation for the CGS-genes.

Figure 1 presents the distribution of gene ontology terms. Most of the genes in the CGS encode proteins associated with cell membranes that are related with regulation of transcription, transport or metabolic/oxido-reduction process and present catalytic/hydrolase, transferase or transcription-factor activity (**Figure 1**). Again, many functions associated to CGS genes in chromosome I were not found in chromosome II such as kinase activity, metal ion transport and phospho-relay response regulator activity, etc. (**Figure 1**). In addition, no CGS genes related to proteolysis, phospho-relay signal transduction, carbohydrate transport, phosphorylation, ATP catabolic process, chemotaxis, biosynthesis, DNA recombination and repair, and phosphoenolpyruvate-dependent sugar phosphotransferase system were identified in chromosome II (**Figure 1**).

To determine the CGP, the plasmid was reconstructed from the sequenced genomes of the Bt2 strains indicated in **Table 1**. According to the phylogenetic tree (**Figure 3**), we found six variants of the plasmid, two previously described (types II and IV) (Lee et al., 2008; Roig and Amaro, 2009) and four new ones (**Table 4**). The list of genes in the CGP is shown in **Table S5**. The CGP includes three virulence genes, two host-specific, *vep07* and *ftbp* (fish transferrin binding protein), involved in resistance to and growth in eel serum, respectively (unpublished results, Pajuelo et al., 2015), and one host-nonspecific, *rtxA1*. This is a mosaic gene related with resistance to phagocytosis by murine and eel phagocytes (Lee et al., 2013a; Satchell, 2015) that presents at least seven different forms in *V. vulnificus* (Roig et al., 2011; Satchell, 2015).

Phylogenomic Analysis

To root the *V. vulnificus* phylogenetic tree within the genus, we first reconstructed a phylogenetic tree from the closed genomes of 15 strains belonging to seven *Vibrio* species (*V. parahaemolyticus, V. anguillarum, V. algynolyticus, V. campbellii, V. furnissii, V. cholerae,* and *V. splendidus-clade*) (main genome characteristics shown in **Table S1**) together with 27 selected *V. vulnificus* genomes (**Table 1**). Maximum-likelihood (ML) trees for chromosome I, II and I+II were reconstructed based on the common genes of these *Vibrio* spp. (**Figure S1**). The ML trees showed *V. vulnificus* as a very compact and independent group. Next, we inferred the phylogeny of the species from the 80 *V. vulnificus* genomes indicated in **Table 1** by using ML reconstruction obtained from the SNPs of coding regions.

The phylogenetic trees for *V. vulnificus* are shown in **Figure 2** and **Figure S2**. The rooted *V. vulnificus* phylogenetic reconstructions for chromosomes I and II were compared by congruence tests (**Table 5**). The results clearly indicate that chromosomes I and II were not congruent with one

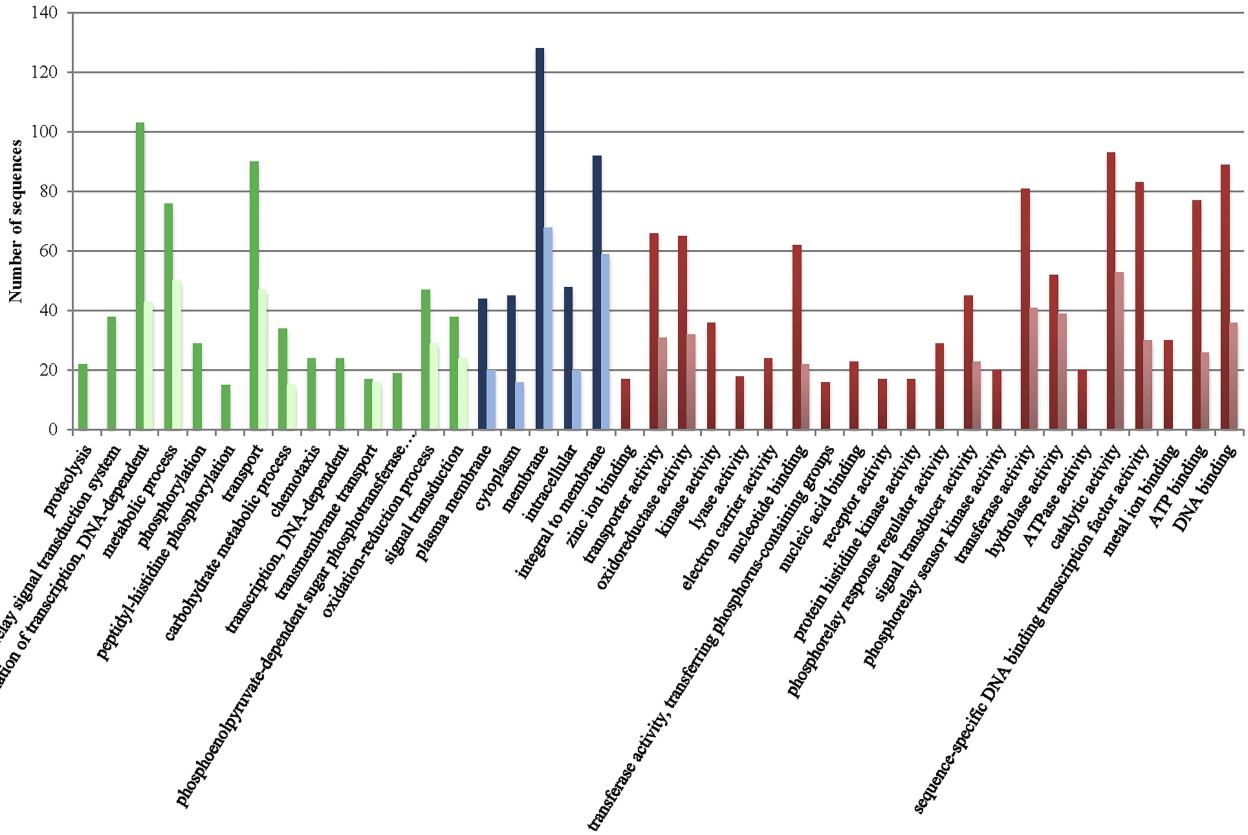


FIGURE 1 | Gene ontology terms distribution of the core genome of the species (CGS) in chromosomes I (dark color) and II (light color). Green, biological process (max. level 15); Blue, cellular component (max. level 15); Red, Molecular Function (max. level 15).

TABLE 4 | Virulence plasmid types and main characteristics.

pVvbt2 type	Strains*	% identity whole sequence	% identity core	Max size	Min size	Core size	GC
I	95-87, 95-86, A14	98.6	100	65,910	65,375	46,446	43.4
II	R02, CECT4999, CECT5763	80.2	100	68,446	66,603	46,446	43.3
III	94-8-112, 96-0426-1-4C	81.5	99.7	64,359	63,933	46,446	43.8
IV	CECT898, CECT4865, PD-2-51, Rae3, CECT4604, CECT4866, CIP8190	88.1	100.0	68,448	66,946	46,446	43.5
V	CECT7030, CECT5769	90.2	97.9	66,249	60,284	46,446	43.4
VI	95-8-161	100	100	63,933	63,933	46,446	43.3

**Serovar E strains: R02, CECT4999, CECT5763, 94-8-112, CECT898, CECT4865, PD-2-51, Rae3, CECT4604, CECT4866, CIP8190. Serovar A strains: CECT7030, CECT5769, A14. Serovar I, 95-87, 95-86, 95-8-161. Non typable, 96-0426-1-4C.*

another, which strongly suggests that both chromosomes have suffered inter and/or intra chromosomal rearrangement since the emergence of the *V. vulnificus* ancestor. In fact, several strains changed their relationships in the tree for each chromosome (**Figure 2**). Among them, we highlight the well-known human clinical isolates C7184, YJ016, and CMCP6 (**Figure 2**). Despite the global incongruence between the two chromosomal phylogenies, all the phylogenetic trees divided the species into five well-defined, highly supported lineages, two of them (L4 and 5) including only a few strains. No strain changed lineage between the two trees (**Figure 2**).

The pairwise identity for the strains of all the lineages was 82.5% (**Table 6**). Lineage 1 (L1) (pairwise identity 93.1%) includes clinical (50% of isolates) and environmental (50% of isolates) Bt1 isolates from the USA, South Korea, Taiwan, Israel, and Spain. The clinical L1 isolates were recovered from human infections in the USA, South Korea, and Taiwan, with 80% of these derived from blood (related to primary septicemia, when the etiology

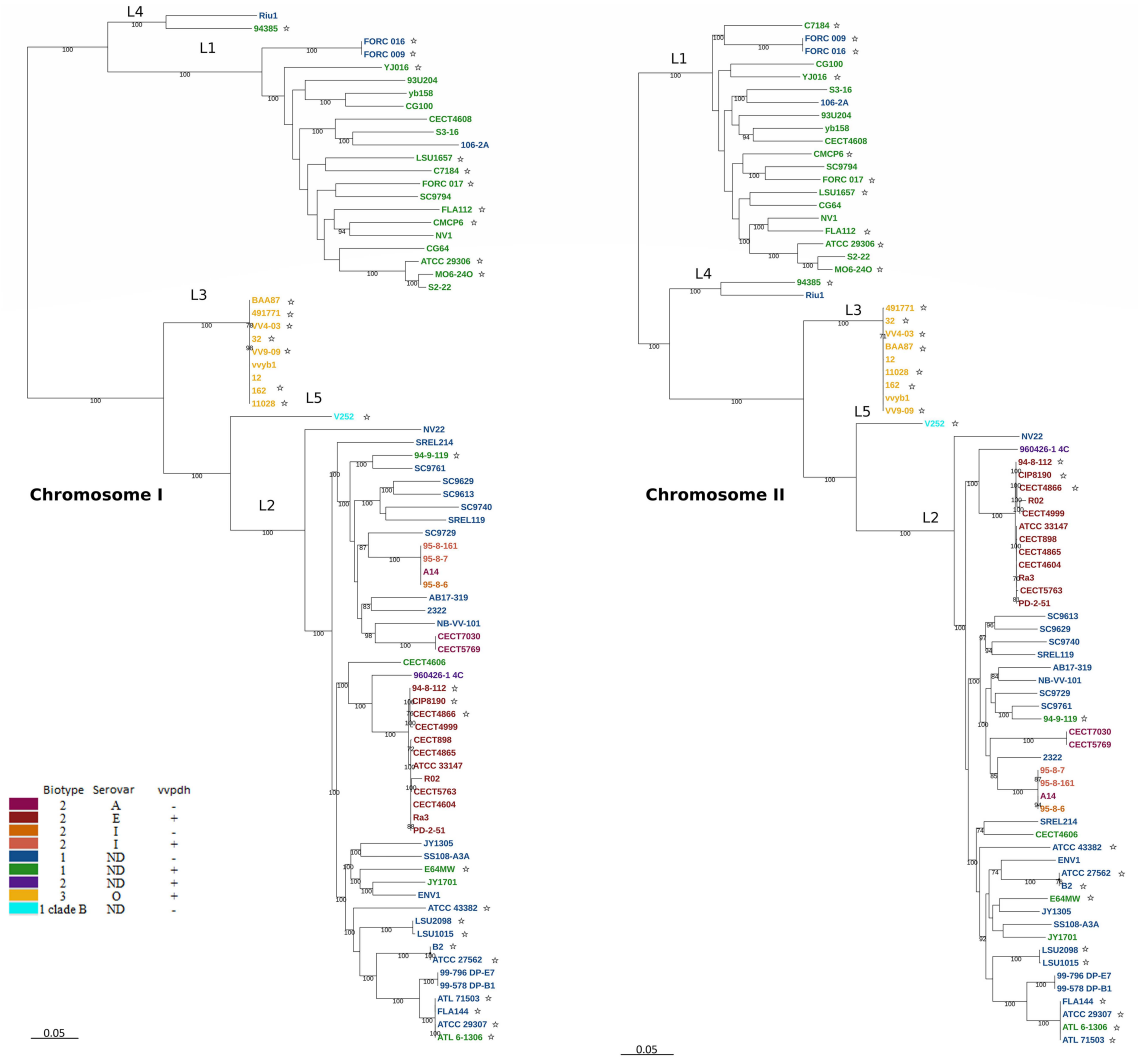


FIGURE 2 | *V. vulnificus* phylogeny reconstructed from single nucleotide polymorphisms (SNPs) of the coding regions in the CGS. *V. vulnificus* phylogeny based on single nucleotide polymorphisms (SNPs) of the coding regions in the core genome of the species (CGS). Maximum-likelihood tree derived using the generalized time-reversible model (GTR+G+I) model of evolution. Bootstrap support values higher than 70% are indicated in the corresponding nodes. *Human clinical isolate.

is known), 10% from stools (related to gastroenteritis) and 10% from wound samples. The environmental L1 isolates were recovered from oysters (40%), water (30%), and fish (30%) in Spain, Israel, Taiwan, South Korea, and the USA. Interestingly, the strain yv158, although environmental, belongs to a previously described highly clonal group (clade A), which includes both environmental and clinical isolates (wound samples), all of *vcg* C-type (clinical type according to the polymorphism in virulence correlated gene; Rosche et al., 2005), supporting the potential virulence of this specific strain (Broza et al., 2012).

L2 (pairwise identity 87.4%) includes Bt1 and Bt2 strains from environmental sources (39.6% of the isolates; 21% from water, 21% from fish, 53.5% from seafood and 4.5% from sediment), diseased humans [29.2% of the isolates; 57% from human blood (none of them with known etiology) and 43% from wounds] and diseased fish (31.2%) origins. All of the zoonotic strains (Bt2-SerE), regardless of their source (environment, diseased animal or diseased human), country (France, Australia, Denmark and Spain) or year of isolation (from 1980 to 2004), cluster in a highly homogeneous group (pairwise identity 97.7%). The non-typeable Bt2 strain (960426-1 4C) formed a subgroup together with the SerE clonal-complex while SerI and A strains formed another subgroup both within L2 (**Figure 2**).

L3 includes all the Bt3 strains, which form a highly homogeneous group (pairwise identity 97.8%). These Bt3 strains were isolated in Israel from outbreaks of human vibriosis associated with the handling of farmed-tilapia (year of isolation, 1996–2003) and from aquaculture fishponds of tilapia

TABLE 5 | Summary of Shimodaira-Hasegawa (SH) and Expected Likelihood Weight (ELW) tests.

Alignment	Topology	log-L	p-1sKH		p-SH		c-ELW		p-AU	
ChrI	ChrI	−2157881.655	1.000	+	1.000	+	1.000	+	1.000	+
	ChrII	−2179900.127	0.000	−	0.000	−	0.000	−	0.000	−
ChrII	ChrI	−761721.329	0.000	−	0.000	−	0.000	−	0.000	−
	ChrII	−757275.779	1.000	+	1.000	+	1.000	+	1.000	+
Virulence Plasmid	ChrI	−73205.493	0	−	0	−	0	−	0.0019	−
	ChrII	−73221.068	0	−	0	−	0	−	0.0009	−
Virulence Plasmid	Virulence Plasmid	−72555.771	1.000	+	1.000	+	1.000	+	0.9977	+
ChrI	ChrI	−340807.458	0.942	+	1.000	+	0.955	+	0.972	+
	ChrII	−340840.343	0.058	+	0.496	+	0.0454	−	0.028	−
	Virulence Plasmid	−405217.661	0	−	0	−	0	−	0.005	−
ChrII	ChrI	−149737.006	0.043	−	0.472	+	0.014	−	0.078	−
	ChrII	−149703.356	0.957	+	1.000	+	0.986	+	0.998	+
	Virulence Plasmid	−178030.177	0	−	0	−	0	−	0.0066	−

The columns show the results and p-values of the following tests: 1sKH—one sided KH test based on pairwise SH tests (Kishino and Hasegawa, 1989; Shimodaira and Hasegawa, 1999; Goldman et al., 2000); SH, Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999); ELW, Expected Likelihood Weight (Strimmer and Rambaut, 2002); p-AU—approximately unbiased test (Shimodaira, 2002). $p < 0.05$. +, positive congruence; –, negative congruence.

TABLE 6 | Global identity of groups and lineages.

Name	% Pairwise Identity
Whole	82.5
Biotype 3	97.8
Biotype 2	94.8
Lineage 1	93.1
Lineage 2	87.4
Lineage 3	97.8
Lineage 4	98.7
Lineage 5	100
Biotype 2 SerA	91.0
Biotype 2 SerE	97.7
Biotype 2 SerI	100.0

(2002–2004). In consequence, all of them are associated to the aquaculture industry.

Interestingly, all human clinical cases of known etiology caused by L2 and L3 isolates in Europe and Israel were related to farmed-fish handling regardless the Bt of the isolate.

L4 (similarity 98.7%) is formed by two Spanish isolates from the Ebro-Delta (a nature park by the Mediterranean Sea) area, one from seawater and the other from a human clinical (leg wound) case, both of Bt1. Finally, L5 is formed by a unique isolate of Bt1 from Israel and clinical origin that is considered as representative of a highly virulent clone designated as Clade B (Raz et al., 2014; Efimov et al., 2015).

The phylogenetic tree for pVvBt2 based on the CGP shows relationships among isolates that did not match the previous phylogenetic relationships (**Figures 2**, **3**). Further, the congruence analyses revealed that phylogenetic trees from the CGP and CGS were not congruent to each other, suggesting a different evolutionary history for the plasmid and the chromosomes (**Table 5**).

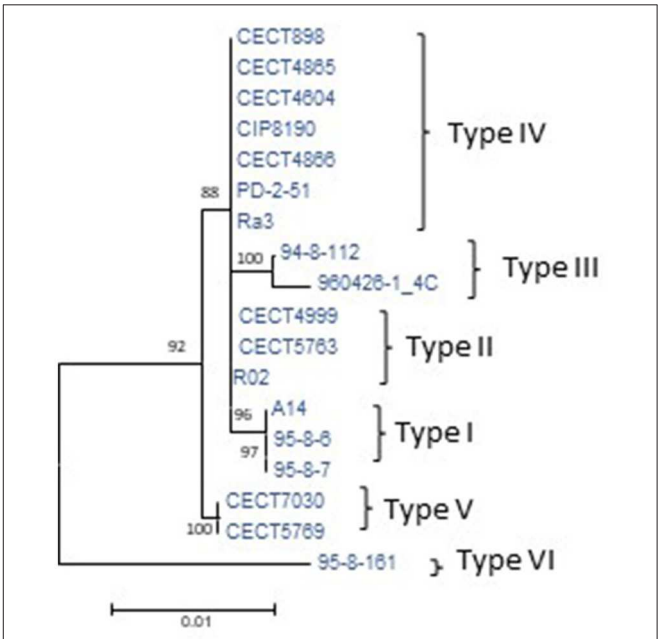


FIGURE 3 | Phylogeny of virulence plasmid pVvbt2 based on SNPs in the core genome of the plasmid (CGP). Maximum-likelihood tree using the Hasegawa-Kishino-Yano model of evolution with gamma distribution and invariant sites. Bootstrap support values higher than 70% are indicated in the corresponding nodes. The different types of plasmids detected and described in **Table 4** are marked to the right of the figure.

Virulence-Related Genes and VCGS

We searched for human virulence-related genes described in the literature for *V. vulnificus*, and we found that 75% of them were present in the CGS. This finding allowed us to define the human virulence-related core genome (VCGS) (**Table S6**). The VCGS includes genes for the flagellum, capsule, LPS-

and cell-wall biosynthesis, motility, hemolysin, and proteases (such as VvhA and a collagenase), resistance to human serum (*trkA;* Chen et al., 2004), heme uptake, biosynthesis and uptake of vulnibactin, several genes for transcriptional regulators such as Fur, and finally genes for MARTX (Multifunctional Autoprocessive Repeat in Toxin) transport and modification systems, genes that are duplicated in pVvbt2 (Chakrabarti et al., 1999; Horstman et al., 2004; Yu-Chung et al., 2004; Gulig et al., 2005; Bogard and Oliver, 2007; Alice et al., 2008; Lee et al., 2008; Oh et al., 2008; Brown and Gulig, 2009; Liu et al., 2009; Chen and Chung, 2011; Kim et al., 2011; Oliver, 2015; Satchell, 2015). Remarkably, genes for several collagenases and chitinases were found in both chromosomes while the MARTX operon and *vvhA* were located on chromosome II.

DISCUSSION

The core genome has been demonstrated to be an optimum data set for determining the phylogeny of a bacterial species since it primarily includes the essential genes of a species and excludes the non-essential ones, such as those present in MGE (mobile genetic element: genomic islands, prophages and plasmids) (Juhas et al., 2009; Segerman, 2012; Wolf et al., 2013). In the current study, we have used this approach to analyze the phylogeny of *V. vulnificus* and compare the phylogenetic groups with the current Bts of the species.

Our phylogenomic analysis suggests that *V. vulnificus* has diverged in five well-defined and separate lineages that do not correspond with the current Bts. L1 is formed by the most dangerous strains from a public health perspective. All of them correspond to Bt1 and were mostly isolated from human blood in North America and Asia, presumably from primary septicemia cases after ingestion of raw seafood. L2 and L3 comprise strains of the three Bts, mostly isolated from fish-farming related environments, including humans infected through handling of farm-fish in Europe and Israel (Bisharat et al., 1999; Haenen et al., 2013).

L2 includes Bt1 and Bt2 strains. Sanjuán et al. (2011) proposed that Bt2 is a polyphyletic group subdivided in Ser-related subgroups, one of which is a clonal complex (Bt2-SerE). Our phylogenomic study confirms that Bt2 is polyphyletic and that the SerE-subgroup is highly homogeneous (identity value of 97.7%). Bt2 was defined in 1982 based on the differential properties of the first fish-pathogenic strains, all of which belonged to SerE and were isolated from eel-farms in Japan (Muroga et al., 1976a,b; Tison et al., 1982). Later, Bt2-SerE isolates were recovered from human infections registered in the USA and Europe, some of them related to zoonotic cases and others of unknown etiology, as well as from different epizootic events of high mortality affecting different farmed-eels in Europe. Bt2-SerA and SerI emerged simultaneously in Spanish and Danish farms after the industry initiated the change from brackish- to fresh-waters in order to control the severity of vibriosis outbreaks due to Bt2-SerE as well as the probability of human infections (Fouz and Amaro, 2003). These new serovars are adapted to infect through and to survive in fresh-waters (Fouz et al., 2006).

All of the analyzed Bt2 strains contained the virulence plasmid pVvbt2, SerE strains present three variants of the plasmid, the two variants previously described (Lee et al., 2008) and a new one (**Table 4**). SerA and I strains showed three new variants (**Table 4**). It was previously hypothesized that Bt2 emerged in fish farms after acquisition of pVvbt2 by different clones of Bt1 strains (Sanjuán et al., 2011). To test this hypothesis, we compared the chromosomal phylogenetic trees reconstructed for Bt2 strains from the CGS with those from the CGP and found that they were not congruent. This result strongly supports the hypothesis of Sanjuán et al. (2011) and suggests that pVvbt2 has been acquired independently by different clones within L2. One of these plasmid-carrier clones successfully amplified in eel-farms and spread to other places and countries, probably in carrier fishes, giving rise to the worldwide expanded, current clonal complex. This clonal complex is supposed to be zoonotic because there are clinical Bt2-SerE isolates related to diseased fish handling and because all the fish and environmental isolates examined to date are virulent for both fish and mice (Sanjuán and Amaro, 2004).

L3 includes all Bt3 strains regardless of their origin (human infections related to fish-farms or environmental), which constitute a clonal group. Bt3 emerged in Israel in 1990 in farms of tilapia and is the only one that has produced outbreaks of human infections, all of them through severe wound infections or secondary septicemia cases (Bisharat et al., 1999). By using different genomic approaches, Raz et al. (2014) and Koton et al. (2014) have hypothesized that Bt3 emerged in the nutrient-enriched environment represented by the aquaculture industry from a Bt1 ancestor that acquired a rather small number of genes from different donors, leading to a change in biotype. The proposed ancestor was v252, a representative strain from a highly virulent clade designated as clade B that shares high similarity and appeared close to Bt3 (Raz et al., 2014; Efimov et al., 2015). Our analysis does not support this hypothesis. Instead, clade B shares the closest common ancestor with L2 and not with L3 in spite of having been isolated from the same "melting pot" where biotype 3 was evolved, i.e., aquaculture fish farms in Israel.

Comparisons of the core genome between clinical and environmental strains of the closely related species *V. cholerae* reveal that this species is divided into two linages, with most of the epidemic strains appearing closely related, regardless of their geographical origins (Eppinger et al., 2014). The only clonal *V. vulnificus* group with a worldwide distribution is that formed by Bt2-SerE strains, a group that combines the ability to infect fish with that of infecting humans and of surviving in the environment without nutrients for years in a viable but non-culturable state (Marco-Noales et al., 1999). Moreover, in *V. cholerae* the ability to cause cholera epidemics lies on mobile genetic elements, such as phages and pathogenicity islands, that carry the genes encoding the cholera toxin, TCP pilus, etc. (Ramamurthy and Bhattacharya, 2011; Das et al., 2016). In contrast, our phylogenomic analysis, as well as those based on MLSA (Cohen et al., 2007; Sanjuán et al., 2011) and microarray hybridization (Raz et al., 2014), show that environmental and clinical strains of *V. vulnificus* are distributed throughout the phylogenetic lineages, regardless the Bt, country of origin, or year

of isolation. This result is compatible with the hypothesis that essentially all *V. vulnificus* isolates, unlike *V. cholerae*, have the ability to infect humans. To confirm this, we investigated which virulence-related genes were present in the CGS and found that most of them belong to the core genome.

Summarizing, all of the phylogenetic reconstructions from the core genome of the species, the fish-virulence plasmid and the human-virulence genes strongly suggest that *V. vulnificus* emerged from an ancestor potentially virulent for humans that diverged in five lineages that do not correspond with the current Bts. Our results also highlight the importance of the aquaculture industry in the recent evolution and epidemic spread of the species and, finally, support the intra-specific classification in lineages instead of in Bts as well as the inclusion of a pathovar grouping all fish pathogenic isolates for which we propose the name "piscis."

AUTHOR CONTRIBUTIONS

CA, FG-C, EF, and FR designed the work, FR, ES, and FG-C performed the phylogenomic analysis, YD-P, BF, CG, CB-A, PG, and SM discussed the preliminary results. CA and FR wrote the paper. All the authors contributed to the discussion and improvement of the MS.

FUNDING

This work has been financed by grants AICO/2018/123, AGL2017-87723-P (both co-funded with FEDER funds), BFU2014-58656-R, Programa Consolider-Ingenio 2010 CSD2009-00006 from MICINN (Spain), and PROMETEO/2016/122 from Generalitat Valenciana.

ACKNOWLEDGMENTS

The authors also thank the SCSIE of the University of Valencia for technical support in determining the sequences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02613/full#supplementary-material

Figure S1 | *Vibrio* phylogeny based on the concatenated whole core genome of seven *Vibrio* species with closed genomes together with selected *V. vulnificus* strains for chromosome I, chromosome II and chromosome I+II. Maximum-likelihood tree derived from the aligned regions by using the GTR+G+I model of evolution. Bootstrap support values higher than 70% are indicated in the corresponding nodes. Color code: green, Bt1 vvpdh+; blue; Bt1 vvpdh−; yellow; Bt3 vvpdh+; red; Bt2 vvpdh+; magenta; Br2 vvpdh−.

Figure S2 | *V. vulnificus* phylogeny reconstructed from single nucleotide polymorphisms (SNPs) of the coding regions in the CGS for both chromosmes (ChrI+ChrII). *V. vulnificus* phylogeny based on single nucleotide polymorphisms (SNPs) of the coding regions in the core genome of the species (CGS). Maximum-likelihood tree derived using the generalized time-reversible model (GTR+G+I) model of evolution. Bootstrap support values higher than 70% are indicated in the corresponding nodes. *Human clinical isolate.

Table S1 | Characteristics of the Vibrio genomes used to define the core genome of the genus (CGG) according to the NCBI Databases.

Table S2 | *V. vulnificus* core genome.

Table S3 | Metabolic pathways associated to *V. vulnificus* core genome.

Table S4 | *V. vulnificus* common genes that were not considered to be part of the Core Genome because its identity and length with respect to the reference sequence were lower than 70 and 80%, respectively.

Table S5 | Core genes in the *V. vulnificus* plasmid pVvbt2 (CGP).

Table S6 | *V. vulnificus* virulence genes in the core genome.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Alice, A. F., Naka, H., and Crosa, J. H. (2008). Global gene expression as a function of the iron status of the bacterial cell: influence of differentially expressed genes in the virulence of the human pathogen *Vibrio vulnificus*. *Infect. Immun.* 76, 4019–4037. doi: 10.1128/IAI.00208-08

Amaro, C., Sanjuán, E., Fouz, B., Pajuelo, D., Lee, C.-T., Hor, L.-L., et al. (2015). The fish pathogen *Vibrio vulnificus* Biotype 2: epidemiology, phylogeny, and virulence factors involved in warm-water vibriosis. *Microbiol. Spectr.* 3:VE-0005-2014. doi: 10.1128/microbiolspec.VE-0005-2014

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Res.* 28, 304–305. doi: 10.1093/nar/28.1.304

Baker-Austin, C., Trinanes, J. A., Taylor, N. G. H., Hartnell, R., Siitonen, A., and Martinez-Urtaza, J. (2013). Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Nat. Clim. Chang.* 3, 73–77. doi: 10.1038/nclimate1628

Baker-Austin, C., Trinanes, J., Gonzalez-Escalona, N., and Martinez-Urtaza, J. (2017). Non-Cholera Vibrios: the microbial barometer of climate change. *Trends Microbiol.* 25, 76–84. doi: 10.1016/j.tim.2016.09.008

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

Biosca, E. G., Amaro, C., Larsen, J. L., and Pedersen, K. (1997). Phenotypic and genotypic characterization of *Vibrio vulnificus* - proposal for the substitution of the subspecific taxon biotype for serovar. *Appl. Environ. Microbiol.* 63, 1460–1466.

Bisharat, N., Agmon, V., Finkelstein, R., Raz, R., Ben-Dror, G., Lerner, L., et al. (1999). Clinical, epidemiological, and microbiological features of *Vibrio vulnificus* biogroup 3 causing outbreaks of wound infection and bacteraemia in Israel. *Lancet* 354, 1421–1424. doi: 10.1016/S0140-6736(99)02471-X

Bogard, R. W., and Oliver, J. D. (2007). Role of iron in human serum resistance of the clinical and environmental *Vibrio vulnificus* genotypes. *Appl. Environ. Microbiol.* 73, 7501–7505. doi: 10.1128/AEM.01551-07

Brown, R. N., and Gulig, P. A. (2009). Roles of RseB, sigmaE, and DegP in virulence and phase variation of colony morphotype of *Vibrio vulnificus*. *Infect. Immun.* 77, 3768–3781. doi: 10.1128/IAI.00205-09

Broza, Y. Y., Raz, N., Lerner, L., Danin-Poleg, Y., and Kashi, Y. (2012). Genetic diversity of the human pathogen *Vibrio vulnificus*: a new phylogroup. *Int. J. Food Microbiol.* 153, 436–443. doi: 10.1016/j.ijfoodmicro.2011.12.011

Chakrabarti, S., Sengupta, N., and Chowdhury, R. (1999). Role of DnaK in *in vitro* and *in vivo* expression of virulence factors of *Vibrio cholerae*. *Infect. Immun.* 67, 1025–1033.

Chen, C.-Y., Wu, K.-M., Chang, Y.-C., Chang, C.-H., Tsai, H.-C., Liao, T.-L., et al. (2003). Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res.* 13, 2577–2587. doi: 10.1101/gr.1295503

Chen, Y.-C., Chuang, Y.-C., Chang, C.-C., Jeang, C.-L., and Chang, M.-C. (2004). A K+ uptake protein, TrkA, is required for serum, protamine, and polymyxin B Resistance in *Vibrio vulnificus*. *Infect. Immun.* 72, 629–636. doi: 10.1128/IAI.72.2.629-636.2004

Chen, Y. C., and Chung, Y. T. (2011). A conserved GTPase YchF of *Vibrio vulnificus* is involved in macrophage cytotoxicity, iron acquisition, and mouse virulence. *Int. J. Med. Microbiol.* 301, 469–474. doi: 10.1016/j.ijmm.2011.02.002

Cohen, A. L. V., Oliver, J. D., DePaola, A., Feil, E. J., Boyd, E. F., Fidelma Boyd, E., et al. (2007). Emergence of a virulent clade of *Vibrio vulnificus* and correlation with the presence of a 33-kilobase genomic island. *Appl. Environ. Microbiol.* 73, 5553–5565. doi: 10.1128/AEM.00635-07

Danin-Poleg, Y., Raz, N., Roig, F. J., Amaro, C., Kashi, Y. (2015). Draft genome sequence of environmental bacterium *Vibrio vulnificus* CladeA-yb158. *Genome Announc*. 3:e00754-15. doi: 10.1128/genomeA.00754-15

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

Das, B., Pazhani, G. P., Sarkar, A., Mukhopadhyay, A. K., Nair, G. B., and Ramamurthy, T. (2016). Molecular evolution and functional divergence of *Vibrio cholerae*. *Curr. Opin. Infect. Dis.* 29, 520–527. doi: 10.1097/QCO.0000000000000306

Efimov, V., Danin-Poleg, Y., Green, S., Elgavish, S., and Kashi, Y. (2015). Draft genome sequence of the pathogenic bacterium *Vibrio vulnificus* V252 biotype 1, isolated in Israel. *Genome Announc.* 3:e01182-15. doi: 10.1128/genomeA.01182-15

Eppinger, M., Pearson, T., Koenig, S. S. K., Pearson, O., Hicks, N., Agrawal, S., et al. (2014). Genomic epidemiology of the Haitian cholera outbreak: A single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *MBio* 5:e01721-14. doi: 10.1128/mBio.01721-14

Farmer, J. J. III. (1979). *Vibrio ("Beneckea") vulnificus*, the bacterium associated with sepsis, septicaemia, and the sea. *Lancet* 2, 903.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.

Fouz, B., and Amaro, C. (2003). Isolation of a new serovar of *Vibrio vulnificus* pathogenic for eels cultured in freshwater farms. *Aquaculture* 217, 677–682. doi: 10.1016/S0044-8486(02)00200-4

Fouz, B., Larsen, J. L., and Amaro, C. (2006). *Vibrio vulnificus* serovar A: an emerging pathogen in European anguilliculture. *J. Fish Dis.* 29, 285–291. doi: 10.1111/j.1365-2761.2006.00719.x

Futami, R., Muñoz-Pomer, L., Viu, J. M., Domínguez-Escribá, L., Covelli, L., Bernet, G. P., et al. (2011). GPRO: the professional tool for management, functional analysis and annotation of omic sequences and databases. *Biotech. Bioinform.* 2011–SOFT3.

Gene Ontology Consortium (2008). The Gene Ontology project in 2008. *Nucleic Acids Res.* 36, D440–D444. doi: 10.1093/nar/gkm883

Goldman, N., Anderson, J. P., Rodrigo, A. G. (2000). Likelihood-based tests of topologies in phylogenetics. *System. Biol.* 49, 652–670. doi: 10.1080/106351500750049752

Guindon, S., Delsuc, F., Dufayard, J. F., and Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.* 537, 113–137. doi: 10.1007/978-1-59745-251-9_6

Gulig, P. A., Bourdage, K. L., and Starks, A. M. (2005). Molecular Pathogenesis of *Vibrio vulnificus*. *J. Microbiol.* 43, 118–131.

Haenen, O. L. M., Evans, J. J., and Berthe, F. (2013). Bacterial infections from aquatic species: potential for and prevention of contact zoonoses. *Rev. Off. Int. Epizoot.* 32, 497–507. doi: 10.20506/rst.32.2.2245

Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., et al. (2010). Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327, 469–474. doi: 10.1126/science.1182395

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174. doi: 10.1007/BF02101694

Hlady, W. G., and Klontz, K. C. (1996). The epidemiology of *Vibrio* infections in Florida, 1981-1993. *J. Infect. Dis.* 173, 1176–1183. doi: 10.1093/infdis/173.5.1176

Horstman, A. L., Bauman, S. J., and Kuehn, M. J. (2004). Lipopolysaccharide 3-Deoxy-D-manno-octulosonic Acid (Kdo) core determines bacterial association of secreted toxins. *J. Biol. Chem.* 279, 8070–8075. doi: 10.1074/jbc.M308633200

Jones, M. K., and Oliver, J. D. (2009). *Vibrio vulnificus*: disease and pathogenesis. *Infect. Immun.* 77, 1723–1733. doi: 10.1128/IAI.01046-08

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393. doi: 10.1111/j.1574-6976.2008.00136.x

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

Kim, I. H., Kim, B. S., Lee, K. S., Kim, I. J., Son, J. S., and Kim, K. S. (2011). Identification of virulence factors in *Vibrio vulnificus* by comparative transcriptomic analyses between clinical and environmental isolates using cDNA microarray. *J. Microbiol. Biotechnol.* 21, 1228–1235. doi: 10.4014/jmb.1111.11016

Kirkup, B. C., Chang, L., Chang, S., Gevers, D., and Polz, M. F. (2010). *Vibrio* chromosomes share common history. *BMC Microbiol.* 10:137. doi: 10.1186/1471-2180-10-137

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evolut.* 29, 170–179. doi: 10.1007/BF02100115

Kotera, M., Hirakawa, M., Tokimatsu, T., Goto, S., and Kanehisa, M. (2012). The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. *Methods Mol. Biol.* 802, 19–39. doi: 10.1007/978-1-61779-400-1_2

Koton, Y., Gordon, M., Chalifa-Caspi, V., and Bisharat, N. (2014). Comparative genomic analysis of clinical and environmental *Vibrio vulnificus* isolates revealed biotype 3 evolutionary relationships. *Front. Microbiol.* 5:803. doi: 10.3389/fmicb.2014.00803

Lee, C.-T., Amaro, C., Wu, K.-M., Valiente, E., Chang, Y.-F., Tsai, S.-F., et al. (2008). A common virulence plasmid in biotype 2 *Vibrio vulnificus* and its dissemination aided by a conjugal plasmid. *J. Bacteriol.* 190, 1638–1648. doi: 10.1128/JB.01484-07

Lee, C.-T., Pajuelo, D., Llorens, A., Chen, Y.-H., Leiro, J. M., Padrós, F., et al. (2013a). MARTX of *Vibrio vulnificus* biotype 2 is a virulence and survival factor. *Environ. Microbiol.* 15, 419–432. doi: 10.1111/j.1462-2920.2012.02854.x

Lee, K. J., Lee, N. Y., Han, Y. S., Kim, J., Lee, K. H., and Park, S. J. (2010). Functional characterization of the IlpA protein of *Vibrio vulnificus* as an adhesin and its role in bacterial pathogenesis. *Infect. Immun.* 78, 2408–2417. doi: 10.1128/IAI.01194-09

Lee, S. H., Chung, B. H., and Lee, W. C. (2013b). Retrospective analysis of epidemiological aspects of *Vibrio vulnificus* infections in Korea in 2001-2010. *Jpn. J. Infect. Dis.* 66, 331–333. doi: 10.7883/yoken.66.331

Liu, M., Naka, H., and Crosa, J. H. (2009). HlyU acts as an H-NS antirepressor in the regulation of the RTX toxin gene essential for the virulence of the human pathogen *Vibrio vulnificus* CMCP6. *Mol. Microbiol.* 72, 491–505. doi: 10.1111/j.1365-2958.2009.06664.x

Marco-Noales, E., Biosca, E. G., and Amaro, C. (1999). Effects of salinity and temperature on long-term survival of the eel pathogen *Vibrio vulnificus* biotype 2 (serovar E). *Appl. Environ. Microbiol.* 65, 1117–1126.

Marco-Noales, E., Milán, M., Fouz, B., Sanjuán, E., Amaro, C., Milan, M., et al. (2001). Transmission to eels, portals of entry, and putative reservoirs of *Vibrio vulnificus* Serovar E (Biotype 2). *Appl. Environ. Microbiol.* 67, 4717–4725. doi: 10.1128/AEM.67.10.4717-4725.2001

Muroga, K., Jo, Y., and Nishibuchi, M. (1976a). Pathogenic *Vibrio* isolated from cultured eels. I. Characteristics and taxonomic status. *Fish Pathol.* 11, 141–145. doi: 10.3147/jsfp.11.141

Muroga, K., Nishibuchi, M., and Jo, Y. (1976b). Pathogenic *Vibrio* isollated from cultured eels. II. Physiological characteristics and pathogenicity. *Fish Pathol.* 11, 147–151. doi: 10.3147/jsfp.11.147

Newton, A., Kendall, M., Vugia, D. J., Henao, O. L., and Mahon, B. E. (2012). Increasing rates of vibriosis in the United States, 1996–2010: review of surveillance data from 2 systems. *Clin. Infect. Dis.* 54, S391–S395. doi: 10.1093/cid/cis243

Oh, M. H., Jeong, H. G., and Choi, S. H. (2008). Proteomic identification and characterization of *Vibrio vulnificus* proteins induced upon exposure to INT-407 intestinal epithelial cells. *J. Microbiol. Biotechnol.* 18, 968–974.

Oliver, J. D. (2015). The biology of *Vibrio vulnificus*. *Microbiol. Spectr.* 3:VE-0001-2014. doi: 10.1128/microbiolspec.VE-0001-2014

Pajuelo, D., Lee, C.-T., Roig, F. J., Hor, L.-I., and Amaro, C. (2015). Novel host-specific iron acquisition system in the zoonotic pathogen *Vibrio vulnificus*. *Environ. Microbiol.* 17, 2076–2089. doi: 10.1111/1462-2920.12782

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Ramamurthy, T., and Bhattacharya, S. K. (2011). *Epidemiological and Molecular Aspects on Cholera*. New York, NY: Springer-Verlag.

Raz, N., Danin-Poleg, Y., Hayman, R. B., Bar-On, Y., Linetsky, A., Shmoish, M., et al. (2014). Genome-wide SNP-genotyping array to study the evolution of the human pathogen *Vibrio vulnificus* biotype 3. *PLoS ONE* 9:e114576. doi: 10.1371/journal.pone.0114576

Rippey, S. R. (1994). Infectious diseases associated with molluscan shellfish consumption. *Clin. Microbiol. Rev.* 7, 419–425. doi: 10.1128/CMR.7.4.419

Roig, F. J., and Amaro, C. (2009). Plasmid diversity in *Vibrio vulnificus* biotypes. *Microbiology* 155, 489–497. doi: 10.1099/mic.0.023424-0

Roig, F. J., González-Candelas, F., Amaro, C., Gonzalez-Candelas, F., and Amaro, C. (2011). Domain organization and evolution of multifunctional autoprocessing repeats-in-toxin (MARTX) toxin in *Vibrio vulnificus*. *Appl. Environ. Microbiol.* 77, 657–668. doi: 10.1128/AEM.01806-10

Roig, F. J., Sanjuán, E., Llorens, A., and Amaro, C. (2010). *pilF* polymorphism-based PCR to distinguish *Vibrio vulnificus* strains potentially dangerous to public health. *Appl. Environ. Microbiol.* 76, 1328–1333. doi: 10.1128/AEM.01042-09

Rosche, T. M., Yano, Y., and Oliver, J. D. (2005). A rapid and simple PCR analysis indicates there are two subgroups of *Vibrio vulnificus* which correlate with clinical or environmental isolation. *Microbiol. Immunol.* 49, 381–389. doi: 10.1111/j.1348-0421.2005.tb03731.x

Sanjuán, E., and Amaro, C. (2004). Protocol for specific isolation of virulent strains of *Vibrio vulnificus* serovar E (biotype 2) from environmental samples. *Appl. Environ. Microbiol.* 70, 7024–7032. doi: 10.1128/AEM.70.12.7024-7032.2004

Sanjuán, E., González-Candelas, F., and Amaro, C. (2011). Polyphyletic origin of *Vibrio vulnificus*biotype 2 as revealed by sequence-based analysis. *Appl. Environ. Microbiol.* 77, 688–695. doi: 10.1128/AEM.01263-10

Satchell, K. J. F. (2015). Multifunctional-autoprocessing repeats-in-toxin (MARTX) Toxins of *Vibrios*. *Microbiol. Spectr.* 3:VE-0002-2014. doi: 10.1128/microbiolspec.VE-0002-2014

Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504. doi: 10.1093/bioinformatics/18.3.502

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* 2:116. doi: 10.3389/fcimb.2012.00116

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *System. Biol.* 51, 492–508. doi: 10.1080/10635150290069913

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116. doi: 10.1093/oxfordjournals.molbev.a026201

Strimmer, K., and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. B Biol. Sci.* 269, 137–142. doi: 10.1098/rspb.2001.1862

Tatusov, R. L. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Am. Math. Soc. Lect. Math. Life Sci.* 17, 57–86.

Tison, D. L., Nishibuchi, M., Greenwood, J. D., and Seidler, R. J. (1982). *Vibrio vulnificus* biogroup 2: new biogroup pathogenic for eels. *Appl. Environ. Microbiol.* 44, 640–646.

Wang, H., Ooi, B. C., Tan, K. L., Ong, T. H., and Zhou, L. (2003). BLAST++: BLASTing queries in batches. *Bioinformatics* 19, 2323–2324. doi: 10.1093/bioinformatics/btg310

Wolf, Y. I., Snir, S., and Koonin, E. V. (2013). Stability along with extreme variability in core genome evolution. *Genome Biol. Evol.* 5, 1393–1402. doi: 10.1093/gbe/evt098

Xu, Q., Dziejman, M., and Mekalanos, J. J. (2003). Determination of the transcriptome of *Vibrio cholerae* during intraintestinal growth and midexponential phase *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1286–1291. doi: 10.1073/pnas.0337479100

Yu-Chung, C., Ming-Chung, C., Yin-Ching, C., and Chii-Ling, J. (2004). Characterization and virulence of hemolysin III from *Vibrio vulnificus*. *Curr. Microbiol.* 49, 175. doi: 10.1007/s00284-004-4288-5

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 Roig, González-Candelas, Sanjuán, Fouz, Feil, Llorens, Baker-Austin, Oliver, Danin-Poleg, Gibas, Kashi, Gulig, Morrison and Amaro. 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

CORRECTION
published: 21 August 2019
doi: 10.3389/fmicb.2019.01904

Corrigendum: Phylogeny of *Vibrio vulnificus* From the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy

*Francisco J. Roig[1,2,3], Fernando González-Candelas[4,5], Eva Sanjuán[1,2], Belén Fouz[1,2], Edward J. Feil[6], Carlos Llorens[3], Craig Baker-Austin[7], James D. Oliver[8,9], Yael Danin-Poleg[10], Cynthia J. Gibas[11], Yechezkel Kashi[10], Paul A. Gulig[12], Shatavia S. Morrison[11] and Carmen Amaro[1,2]**

[1] Estructura de Investigación Interdisciplinar en Biotecnología y Biomedicina BIOTECMED, University of Valencia, Valencia, Spain, [2] Departmento de Microbiología y Ecología, Universidad de Valencia, Valencia, Spain, [3] Biotechvana, Parc Cientific, Universitat de Valencia, Valencia, Spain, [4] Joint Research Unit on Infection and Public Health FISABIO-Salud Pública and Universitat de Valencia-I2SysBio, Valencia, Spain, [5] CIBEResp, National Network Center for Research on Epidemiology and Public Health, Instituto de Salud Carlos III, Valencia, Spain, [6] Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom, [7] Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom, [8] Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States, [9] Duke University Marine Lab, Beaufort, NC, United States, [10] Faculty of Biotechnology and Food Engineering, Technion–Israel Institute of Technology, Haifa, Israel, [11] Department of Bioinformatics and Genomics, the University of North Carolina at Charlotte, Charlotte, NC, United States, [12] Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Carmen Amaro
Carmen.amaro@uv.es

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

Received: 17 July 2019
Accepted: 02 August 2019
Published: 21 August 2019

Citation:
Roig FJ, González-Candelas F, Sanjuán E, Fouz B, Feil EJ, Llorens C, Baker-Austin C, Oliver JD, Danin-Poleg Y, Gibas CJ, Kashi Y, Gulig PA, Morrison SS and Amaro C (2019) Corrigendum: Phylogeny of Vibrio vulnificus From the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy. Front. Microbiol. 10:1904. doi: 10.3389/fmicb.2019.01904

Keywords: microbial evolution, pathogens, SNP, *Vibrio vulnificus*, core genome, virulence plasmid, pathovar, biotype

A Corrigendum on

Phylogeny of *Vibrio vulnificus* From the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy

by Roig, F. J., González-Candelas, F., Sanjuán, E., Fouz, B., Feil, E. J., Llorens, C., et al. (2018). Front Microbiol. 8:2613. doi: 10.3389/fmicb.2017.02613

There is an error in the Funding statement. The first funding number is incorrect and should be AICO/2018/123. The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Copyright © 2019 Roig, González-Candelas, Sanjuán, Fouz, Feil, Llorens, Baker-Austin, Oliver, Danin-Poleg, Gibas, Kashi, Gulig, Morrison and Amaro. 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: 08 December 2017
doi: 10.3389/fmicb.2017.02434

Strong Genomic and Phenotypic Heterogeneity in the *Aeromonas sobria* Species Complex

Jeff Gauthier[1], Antony T. Vincent[2,3], Steve J. Charette[2,3] and Nicolas Derome[1]*

[1] *Département de Biologie, Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec City, QC, Canada,*
[2] *Centre de Recherche de l'Institut Universitaire de Cardiologie et de Pneumologie de Québec, Quebec City, QC, Canada,*
[3] *Département de Biochimie, de Microbiologie et de Bio-informatique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec City, QC, Canada*

Aeromonas sobria is a mesophilic motile aeromonad currently depicted as an opportunistic pathogen, despite increasing evidence of mutualistic interactions in salmonid fish. However, the determinants of its host-microbe associations, either mutualistic or pathogenic, remain less understood than for other aeromonad species. On one side, there is an over-representation of pathogenic interactions in the *A. sobria* literature, of which only three articles to date report mutualistic interactions; on the other side, genomic characterization of this species is still fairly incomplete as only two draft genomes were published prior to the present work. Consequently, no study specifically investigated the biodiversity of *A. sobria*. In fact, the investigation of *A. sobria* as a species complex may have been clouded by: (i) confusion with *A. veronii* biovar *sobria* because of their similar biochemical profiles, and (ii) the intrinsic low resolution of previous studies based on 16S rRNA gene sequences and multilocus sequence typing. So far, the only high-resolution, phylogenomic studies of the genus *Aeromonas* included one *A. sobria* strain (CECT 4245 / Popoff 208), making it impossible to robustly conclude on the phylogenetic intra-species diversity and the positioning among other *Aeromonas* species. To further understand the biodiversity and the spectrum of host-microbe interactions in *A. sobria* as well as its potential genomic diversity, we assessed the genomic and phenotypic heterogeneity among five *A. sobria* strains: two clinical isolates recovered from infected fish (JF2635 and CECT 4245), one from an infected amphibian (08005) and two recently isolated brook charr probionts (TM12 and TM18) which inhibit *in vitro* growth of *A. salmonicida* subsp. *salmonicida* (a salmonid fish pathogen). A phylogenomic assessment including 2,154 softcore genes corresponding to 946,687 variable sites from 33 *Aeromonas* genomes confirms the status of *A. sobria* as a distinct species divided in two subclades, with 100% bootstrap support. The phylogenomic split of *A. sobria* in two subclades is corroborated by a deep dichotomy between all five *A. sobria* strains in terms of inhibitory effect against *A. salmonicida* subsp. *salmonicida,* gene contents and codon usage. Finally, the antagonistic effect of *A. sobria* strains TM12 and TM18 suggests novel control methods against *A. salmonicida* subsp. *salmonicida*.

Keywords: *Aeromonas sobria*, host–microbe interactions, bacterial genomics, microbial diversity, molecular systematics

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
M. Carmen Fuste, University of Barcelona, Spain
Brigitte Lamy, Centre Hospitalier Universitaire de Nice, France

****Correspondence:***
Jeff Gauthier
jeff.gauthier.1@ulaval.ca

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

Received: *26 September 2017*
Accepted: *23 November 2017*
Published: *08 December 2017*

Citation:
Gauthier J, Vincent AT, Charette SJ and Derome N (2017) Strong Genomic and Phenotypic Heterogeneity in the Aeromonas sobria Species Complex. Front. Microbiol. 8:2434. doi: 10.3389/fmicb.2017.02434

INTRODUCTION

Aeromonas spp. is a genus of Gammaproteobacteria with substantial heterogeneity among species and subspecies in terms of environmental distribution, host range and growth conditions (Cahill, 1990; Janda and Abbott, 2010; Vincent et al., 2016). Aeromonads are ubiquitous in aquatic environments worldwide (Hazen et al., 1978; Chowdhury et al., 1990), either as (i) free-living organisms (Acinas et al., 1999), (ii) sessile life biofilms on biotic and abiotic surfaces (Talagrand-Reboul et al., 2017a) or (iii) as part of the natural microbiota of amphibians, mammals, reptiles and fish (Popoff and Véron, 1976; Cahill, 1990).

For aeromonads associated with fish hosts, there is a broad spectrum of symbiotic interactions, from mutualism (Gibson et al., 1998; Gunasekara et al., 2011) to pathogenicity (Austin and Austin, 2012a,b). The type of interaction can shift in a given host–microbe system, depending on overall host health and several environmental factors such as temperature, fish population density and water quality (Austin and Austin, 2012a,b). For instance, water temperatures above 17°C may trigger acute episodes of furunculosis (*A. salmonicida* subsp. *salmonicida*) in salmonids, with >90% mortality rates in less than a week post-infection (Scott, 1968), whereas for temperatures less than 12°C, fish may either be chronically infected (presence of skin nodules without mortalities) or become asymptomatic carriers (Schachte, 2002). Water temperature is indeed a critical factor, as climate change is predicted to elevate mean temperatures of some North American lakes and rivers to an optimum for *A. salmonicida* subsp. *salmonicida* growth (Tam, 2009; Tam et al., 2011).

Interactions of aeromonads in a given host-microbe system are further controlled by other microbial strains that colonize host body surfaces composing the so called microbiota. Among members of the microbiota, some strains are documented to exert antagonistic effects against pathogenic strains (Boutin et al., 2012, 2013; Goulden et al., 2012; Schubiger et al., 2015) including aeromonads (Zhang et al., 2016; Gao et al., 2017; Gauthier et al., 2017). Whether a host-associated aeromonad exerts mutualistic or pathogenic interactions is highly dependent on which host(s) it infects, but also its genetic repertoire, which varies greatly between strains/species (Ghatak et al., 2016).

In *A. salmonicida*, for example, there are five officially recognized subspecies (Dallaire-Dufresne et al., 2014). Three of them (*achromogenes*, *masoucida* and *smithia*) infect a broad range of hosts including cod (*Gadus morhua*), black rockfish (*Sebastes schlegeli*) and turbot (*Scophtalmus maximus*) (Cornick et al., 1984; Larsen and Pedersen, 1996; Han et al., 2011). Subspecies *pectinolytica* is without any report of pathogenicity, and subspecies *salmonicida* almost exclusively infects salmonid fish (Austin and Austin, 2012a). The broad host range of *A. salmonicida*, as well as the presence of mesophilic strains in this mainly psychrophilic species (Vincent et al., 2016), are evidence of a great genomic heterogeneity and complexity. Indeed, the *A. salmonicida* pangenome (i.e., total non-redundant genes among 26 strains) is made of 8,164 genes, of which 59.2% are accessory genes (Vincent and Charette, 2017). The *A. salmonicida* pangenome is "open," suggesting a high prevalence of genetic material exchanges across other bacteria sharing the same environment (Rouli et al., 2015).

Similarly, motile aeromonads *A. hydrophila, A. veronii* and *A. caviae* also exhibit open pangenomes with a high species-wise proportion of accessory genes (61.7%, 53.4%, and 50.9% respectively), with strong variation in terms of antimicrobial resistance and virulence genes (Ghatak et al., 2016). *A. media,* noted for its remarkable genomic and phenotypic heterogeneities, shows the highest known species-wise proportion of accessory genes for an *Aeromonas* species (68.4%) (Talagrand-Reboul et al., 2017b).

Genome "openness," i.e., strong heterogeneity and high exchangeability of genes, seems to be a defining trait of this genus. This strong genomic heterogeneity, revealed by next-generation sequencing, has led to major reclassifications in the *Aeromonas* taxonomy (Beaz-Hidalgo et al., 2013, 2015). Indeed, of all 30 *Aeromonas* species with valid taxonomic status, half were described since the 2005 edition of the Bergey's Manual of Systematic Bacteriology (Boone et al., 2005). Novel species continue to be described (Figueras et al., 2017).

However, there are other relevant *Aeromonas* species complexes whose diversity and complexity has not been as thoroughly characterized. To this respect, one example of interest is *Aeromonas sobria* (*sensu* Popoff and Véron, 1976), a mesophilic motile aeromonad currently depicted as an opportunistic pathogen of freshwater fish, amphibians and reptiles (Wahli et al., 2005; Janda and Abbott, 2010; Austin and Austin, 2012b; Yang Q.-H. et al., 2017). In spite of increasing evidence of mutualistic interactions mediated by *A. sobria* strains (Brunt and Austin, 2005; Brunt et al., 2007; Pieters et al., 2008), the determinants of its host-microbe associations remain less understood than for other aeromonad species such as *A. salmonicida* subsp. *salmonicida* (Garduño and Kay, 1992; Garduño et al., 1993; Vanden Bergh and Frey, 2013). Indeed, the current literature on *A. sobria* is not only scarce with respect to other aeromonads, but is also strongly biased by an over-representation of pathogenic interactions.

From 1981 to date, PubMed referenced 142 articles dedicated to *A. sobria* while ISI Web of Science referenced 140 *A. sobria* articles from 1978 to date. Articles specifically discussing *A. veronii* biovar *sobria* are not included in this estimate. This constitutes about 10 times less literature than for *A. hydrophila,* a well-documented fish and human opportunistic pathogen (Cipriano et al., 1984; Janda and Abbott, 2010) and about five times less literature than for *A. salmonicida* subsp. *salmonicida,* a major pathogen of salmonids (Austin and Austin, 2012a; Dallaire-Dufresne et al., 2014). There are only three reports on mutualistic interactions by *A. sobria* (Brunt and Austin, 2005; Brunt et al., 2007; Pieters et al., 2008), all of which dealt with the same *A. sobria* strain (GC2).

This rough estimate of 140 *A. sobria* articles may be lower: prior to the description of *A. veronii* (Hickman-Brenner et al., 1987), several ornithine-decarboxylase negative *A. veronii* strains (now referred to as *A. veronii* biovar *sobria*) were incorrectly labeled as *A. sobria*. As a consequence, epidemiology prior to 1987 may be unreliable to this regard. Ironically, the majority of those articles, referenced in either PubMed or Web of Science,

are outbreak reports of immunocompromised human patients. Unlike *A. veronii,* few *A. sobria* (*sensu stricto*) strains have been isolated from sources other than fish and aquatic environments (Janda and Abbott, 2010).

Next-generation sequencing data on *A. sobria* is also scarce. Prior to this publication, only two genome assemblies (both drafts) were available on GenBank, compared to the 61 entries for *A. hydrophila* and 36 for *A. salmonicida*. The genome of the type strain CECT 4245 was sequenced through a large-scale study on the genus *Aeromonas* (Colston et al., 2014) while the one of 08005 was published as a Genome Announcement (Yang Q.-H. et al., 2017). Consequently, no study specifically investigated the genomic features and diversity of *A. sobria*.

To increase our knowledge about the spectrum of host–microbe interactions in *A. sobria* as well as its biodiversity, we report the comparative phenotypic and genomic analysis of five host-associated *A. sobria* strains including two mutualistic strains with strong *in vitro* antagonistic effect against *A. salmonicida* subsp. *salmonicida*. Genome sequencing and comparative analyses of these strains revealed unexpected heterogeneity between all five *A. sobria* strains in terms of phylogeny, codon usage and gene contents, which closely correlates with their phenotype regarding their inhibitory effect against *A. salmonicida* subsp. *salmonicida*.

MATERIALS AND METHODS

Bacterial Isolates and Growth Conditions

Aeromonas sobria strains TM12 and TM18 were both isolated in 2015 from the intestinal contents of an adult brook charr (*Salvelinus fontinalis*) from Lake Prime-Huron, Quebec, Canada. Strain JF2635 was isolated in 2001 from a European perch (*Perca fluviatilis*) in Switzerland (Wahli et al., 2005). Strain CECT 4245 (formerly known as Popoff 208) was isolated from an infected fish specimen (Popoff and Véron, 1976). All *A. sobria* isolates were grown on lysogeny broth (LB) agar or Tryptic Soy Agar plates (TSA, BD Diagnostics) plates at 18°C or 30°C, except recently sequenced strain 08005 (Yang Q.-H. et al., 2017) which was only included in the comparative genomic analyses. Indeed, this novel genome sequence (08005) was published after *in vitro* assays were completed. Given the scarcity of *A. sobria* genome sequences, we chose to include this strain in comparative genomic analyses, despite the absence of *in vitro* results, in order to maximize taxon sampling.

Phenotypic Characterization

Interspecific Antagonism Assays

Bacterial lawns of ten *A. salmonicida* subsp. *salmonicida* strains (Supplementary Table 1) were prepared by streaking a sterile swab dipped in liquid culture (OD_{600} = 0.7) on TSA plates. Wells were punched in the agar using sterile pipette tips with a diameter of 3.5 mm. For each *A. sobria* isolate, 10 μL of liquid culture (OD_{600} = 0.7) were dispensed in an assigned well. Plates were incubated at 18°C for 96 h. Inhibition surfaces around the wells were measured on 23.6 pixel/mm scans with software ImageJ version 1.48k (Schneider et al., 2012), with the well area subtracted from the whole inhibition area.

Antimicrobial Activity of Extracellular Products (ECP)

For each *A. sobria* isolate, ECPs were recovered by centrifuging overnight liquid cultures incubated at 18°C for 24 h in LB broth, all adjusted to OD_{600} = 0.7. Culture supernatants (CS), obtained by centrifugation at 4,000 × *g* for 20 min at 4°C, were then filtered with a 0.2 μm Filtropur S syringe disk filter (Sarstedt), arrayed on a Bioscreen C microplate (Growth Curves AB Ltd, Helsinki, Finland), and supplemented with an equal volume of *A. salmonicida* subsp. *salmonicida* 01-B526 liquid culture in LB broth adjusted at $OD_{600} \sim 0.05$. *A. salmonicida* subsp. *salmonicida* 01-B526 CS and fresh LB medium were used as neutral and negative controls for *A. sobria* CSs, respectively. Mixtures were incubated in a Bioscreen C plate reader at 18°C for 48 h, with OD_{600} measured at each hour. Statistical significance of growth differences between conditions was assessed with a one-way ANOVA for repeated measures over time (df_{time} = 47, $df_{conditions}$ = 4, $\alpha = 0.05$). Data used for testing were balanced (i.e., equal number of observations per condition), and respected the assumptions of normality (Shapiro–Wilk test: W = 0.98407, $p = 0.9899$) and homoscedasticity (Bartlett's test: $K^2 = 2.2485$, $df = 4$, $p = 0.6902$). *Post hoc* comparisons of means were performed using Tukey's HSD test only if a statistically significant difference was detected by ANOVA.

Biofilm Formation

The ability of *A. sobria* to produce biofilms in liquid broth was verified with the microtiter dish assay described by O'Toole (2011) with minor modifications. Briefly, overnight LB broth cultures adjusted at OD_{600} = 0.9 were diluted 1:100 in either LB-Miller or Tryptic Soy Broth (TSB, BD Diagnostics); 100 μL of each diluted culture were arrayed in triplicates on disposable PVC U-bottomed plates (VWR International). Plates were incubated at 30°C without shaking for 6 h. After incubation, OD_{600} was measured to estimate bacterial abundance in each culture. Biofilms were stained by adding 25 μL of 1% aqueous Crystal Violet solution in each well, and were let standing for 15 min at room temperature (RT). Wells were rinsed abundantly with distilled water. Wells were then washed twice with 200 μL 95% ethanol, which was kept and arrayed on a clean microtiter plate. Biofilms were quantified by reading the OD_{600} in each ethanol/Crystal Violet mixture. Statistical significance of growth differences between conditions was assessed with a two-way ANOVA (Factors: Growth media and *A. sobria* strain, df_{media} = 1, df_{strain} = 4, $\alpha = 0.05$). Data used for testing were balanced (i.e., had an equal number of observations per condition), and were $\log_{10}$-transformed to improve normality (Shapiro–Wilk test: $W = 0.90401$, $p = 0.01054$) and homoscedasticity (Levene's test: $df = 9$, $F = 1.0206$, $p = 0.4571$). No other data transformation improved normality and homoscedasticity as efficiently as the $\log_{10}$ transformation. *Post hoc* comparisons of means were performed using Tukey's HSD test only if a statistically significant difference was detected by ANOVA.

Growth Kinetics

For each *A. sobria* isolate, 400 μL of overnight culture in LB adjusted at $OD_{600} \sim 0.05$ were arrayed on a sterile transparent covered plate (CORNING). The plate was incubated at 30°C in an Infinite 200 PRO microplate incubator/reader equipped with a 595 nm absorbance filter (TECAN, Morrisville, NC, United States). The plate was shaken (200 RPM) for 48 h; OD_{595} was measured at each incubation cycle of 15 min.

Comparative Genomics Analyses

DNA Extraction and Genome Sequencing

The total genomic DNA of *A. sobria* strains JF2635, TM12, and TM18 was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Canada). The sequencing libraries were prepared using a KAPA Hyper Prep Kit and were sequenced by next-generation sequencing (NGS) on a MiSeq instrument (Illumina technology) by the Plateforme d'Analyse Génomique of the Institut de Biologie Intégrative et des Systèmes (IBIS, Université Laval). The resulting sequencing reads were *de novo* assembled into contiguous sequences using the A5-miseq pipeline version 20160825 (Tritt et al., 2012). Contigs were ordered using mauveAligner (Rissman et al., 2009) with the complete genome of *Aeromonas veronii* B565 (CP002607.1) as a reference. The complete draft genomes were annotated with RAST (Aziz et al., 2008) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and deposited in the public database GenBank (TM12: NQML00000000, TM18: NQMM00000000, and JF2635: LJZX00000000). The whole genome sequence of *A. sobria* CECT 4245 (GenBank: CDBW00000000.1) and 08005 (GenBank: NZ_MKFU00000000) were already available prior to this study.

Plasmid Assembly and Annotation

The high-copy plasmid sequences were recovered by downsampling the sequencing reads using seqtk[1] before re-performing *de novo* assemblies with the A5-miseq pipeline. The plasmid sequences were annotated with the RAST web server (Aziz et al., 2008) and were manually curated. The presence or absence of a type II toxin-antitoxin locus was assessed for each sequence using TAfinder (Shao et al., 2011). Plasmid sequences were deposited on GenBank (MF770238 and MF770239 for strain JF2635; MF770240, MF770241, and MF770242 for strain TM18).

Molecular Systematics

In addition to three *A. sobria* genome sequences produced by the present study, 30 genome sequences from representative strains of all *Aeromonas* species available in GenBank were downloaded, thus making a dataset of 33 genomes (Supplementary Table 2). To avoid annotation bias, all the sequences were locally annotated with Prokka version 1.12-beta (Seemann, 2014). Homology links between the coding sequences were detected with GET_HOMOLOGUES version 20170105 (Contreras-Moreira and Vinuesa, 2013) using two algorithms, COG (Kristensen et al., 2010) and OMCL (Li et al., 2003). Homologous sequences detected with both algorithms were kept for the subsequent analyzes. The 2,154 nucleotidic sequences corresponding to orthologous genes of the softcore (genes present in at least 95% of the genomes) and without paralogous ambiguity were codon aligned by muscle version 3.7 (Edgar, 2004) through TranslatorX (Abascal et al., 2010). Monomorphic sites were removed for each alignment with BMGE version 1.2 (Criscuolo and Gribaldo, 2010). All the sequences were concatenated and partitioned into a supermatrix by AMAS (Borowiec, 2016). The best-fit model was found for each partition using IQ-TREE version 1.5.3 (Nguyen et al., 2015). Finally, a maximum-likelihood tree was inferred also using IQ-TREE and the branch supports obtained with 10,000 ultrafast bootstraps (Minh et al., 2013). The average nucleotide identity (ANI) was computed for the 33 taxa using pyani (Pritchard et al., 2016) and NUCmer version 3.1 (Kurtz et al., 2004).

Other Analyses

The pangenome of all five *A. sobria* strains studied here was inferred using GET_HOMOLOGUES (Contreras-Moreira and Vinuesa, 2013), allowing to sort the genes in four categories based on orthologous gene cluster frequency distribution: core genes (present in all genomes), softcore (present in 95% of all genomes), cloud (present in 1 or 2 genomes only) and shell (all remaining genes). Relative Synonymous Codon Usage (RSCU) was computed using DAMBE6 (Xia, 2017). The principal components analysis (PCA) used to differentiate the isolates in two groups based on the RSCU values was performed by the R package ade4 (Dray and Dufour, 2007). Antibiotic resistance genes were found using the resistance gene identifier (RGI) from the CARD database (McArthur et al., 2013). Genes implicated in a secretion system were found by TXSScan (Abby et al., 2016). Prophages were detected with PHASTER (Arndt et al., 2016) using pseudo-finished genome assemblies prepared with CONTIGuator v2.7.1 (Galardini et al., 2011) using the *A. veronii* B565 complete genome (CP002607.1) as a reference chromosome for contig alignments.

RESULTS AND DISCUSSION

Strong Phenotypic Heterogeneity

Interspecific Antagonism

Two of the *A. sobria* strains (TM12 and TM18) analyzed in this study are gut symbionts recovered from healthy brook charr (*Salvelinus fontinalis*) (**Table 1**). Both were isolated in a research project aiming to study the interactions between resident brook charr bacteria and *A. salmonicida* subsp. *salmonicida*. Both strains had strong, yet qualitatively different *in vitro* inhibitory effects against fish pathogen *A. salmonicida* subsp. *salmonicida* (**Table 2**). This finding was interesting because two *A. sobria* strains were recovered from healthy specimens, yet also had an inhibitory effect against *A. salmonicida* subsp. *salmonicida* which is also part of the resident brook charr microbiota (Dallaire-Dufresne et al., 2014). This prompted the assessment of this antagonistic effect in other *A. sobria* strains (JF2635, CECT 4245), which are clinical isolates recovered from infected fish (**Table 1**). No data regarding their inhibitory effect against

[1]https://github.com/lh3/seqtk

TABLE 1 | *Aeromonas sobria* strains used in this study.

Strain	Source organism	Host disease status	Year of isolation	Reference
TM12	Brook charr (*Salvelinus fontinalis*)	Healthy	2015	This study
TM18	Brook charr (*Salvelinus fontinalis*)	Healthy	2015	This study
JF2635	European perch (*Perca fluviatilis)*	Moribund	2004	a
CECT 4245^T	Fish (unknown)	Unknown	ca. 1967–74	b
08005 *	American bullfrog (*Rana castebeiana)*	Moribund	2016	c

*a: Wahli et al. (2005). b: Popoff and Véron (1976). c: Yang Q.-H. et al. (2017). T, type strain (CECT, 1991); duplicated from Popoff 208 (1976). *Strain present only in genomic analyses.*

TABLE 2 | Diffusible inhibitory effect of *A. sobria* strains on TSA bacterial lawns of *A. salmonicida* subsp. *salmonicida*, after 96 h at 18°C.

A. salmonicida* subsp. *salmonicida		***A. sobria***			
Origin	**Strain**	**TM12**	**TM18**	**JF2635**	**CECT 4245^T**
Province of Quebec, Canada	01-B522	+	+++	–	–
	01-B526	+	++	–	–
	09-0167	++	+++	–	–
	M15879-11	+	+	–	–
	m23067-09	+	+++	–	–
New Brunswick, Canada	04-05MF26	++	+++	–	–
	09-144K3	++	+++	–	–
Norway	HER1085	++	+++	–	–
Switzerland	JF2267	+	+	–	–
France	A449^T	+	+++	–	–

+ 50–100 mm^2; ++ 100–200 mm^2; +++ more than 200 mm^2; – no visible inhibition plaque; T, type strain.

A. salmonicida subsp. *salmonicida* was available prior to this study.

While *A. sobria* isolates TM12 and TM18 had an antagonistic effect against 10 strains of *A. salmonicida* subsp. *salmonicida* from various hosts and geographical origins (**Table 2**), CECT 4245 and JF2635 did not show any disruptive effect on the growth of any *A. salmonicida* subsp. *salmonicida* strain, even after 96 h. Radial inhibition on *A. salmonicida* subsp. *salmonicida* bacterial lawns by TM12 and TM18 suggests involvement of diffusible inhibitory compounds, even though major differences were observed between these two strains. Strain TM12 produced inhibition halos overlapping the diffusion area of a blue pigment while TM18 did not exhibit any visible pigmentation, but a significantly stronger antagonistic effect (Supplementary Figure 1). Production of inhibition zones on *A. salmonicida* subsp. *salmonicida* lawns indicates that *A. sobria* TM12 and TM18 (but neither JF2635 nor CECT 4245) can produce diffusible antimicrobial compounds that inhibit *A. salmonicida* subsp. *salmonicida.* Interestingly, both strains that were isolated as causative infectious agents (JF2635 and CECT 4245) had no inhibitory effect against *A. salmonicida* subsp. *salmonicida,* whereas TM12 and TM18 (recovered from asymptomatic fish) had a strong antimicrobial effect on *A. salmonicida* subsp. *salmonicida.*

The production of antimicrobial compounds targeting *A. salmonicida* subsp. *salmonicida* has also been assessed by exposing *A. salmonicida* subsp. *salmonicida* 01-B526 to *A. sobria* extracellular products (ECPs) in culture supernatants (Supplementary Figure 2). Interestingly, results do not follow the trend observed in the agar assays described above. In fact, no significant change of *A. salmonicida* growth was detected after 48 h growth [$F_{(4,10)} = 0.482$, $p = 0.749$].

It is possible that the inhibitory compounds produced in agar are not produced when grown in liquid broth. Indeed, in solid medium assays, *A. sobria* strains were in conditions of high cell density without direct contact with *A. salmonicida* subsp. *salmonicida* cells, i.e., conditions resembling a bacterial biofilm (McBain, 2009). On the opposite, in liquid broth assays, *A. sobria* cells were in conditions more akin to planktonic life before recovery of their ECPs (i.e., in liquid broth with continuous shaking). Therefore, the growth characteristics of *A. sobria* isolates in liquid broth were investigated.

Growth Kinetics

All *A. sobria* strains undergo a similar growth pattern in LB broth (**Figure 1**). Stationary phase is reached after 10 h ($OD_{600} \sim [0.8; 0.9]$), followed by gradual decline. However, one striking difference is the high levels of background noise in the TM18 growth curve throughout the stationary and decline phases. This suggests that either cell aggregation occurs in liquid cultures, or that TM18 has the ability to form significant levels of biofilms in liquid broth. The latter possibility was subsequently assessed.

Biofilm Formation

The levels of biofilm production by *A. sobria* are significantly different in LB than in TSB media [$F_{(1,20)} = 62.80$, $p = 1.35 \times 10^{-7}$], and vary significantly between strains

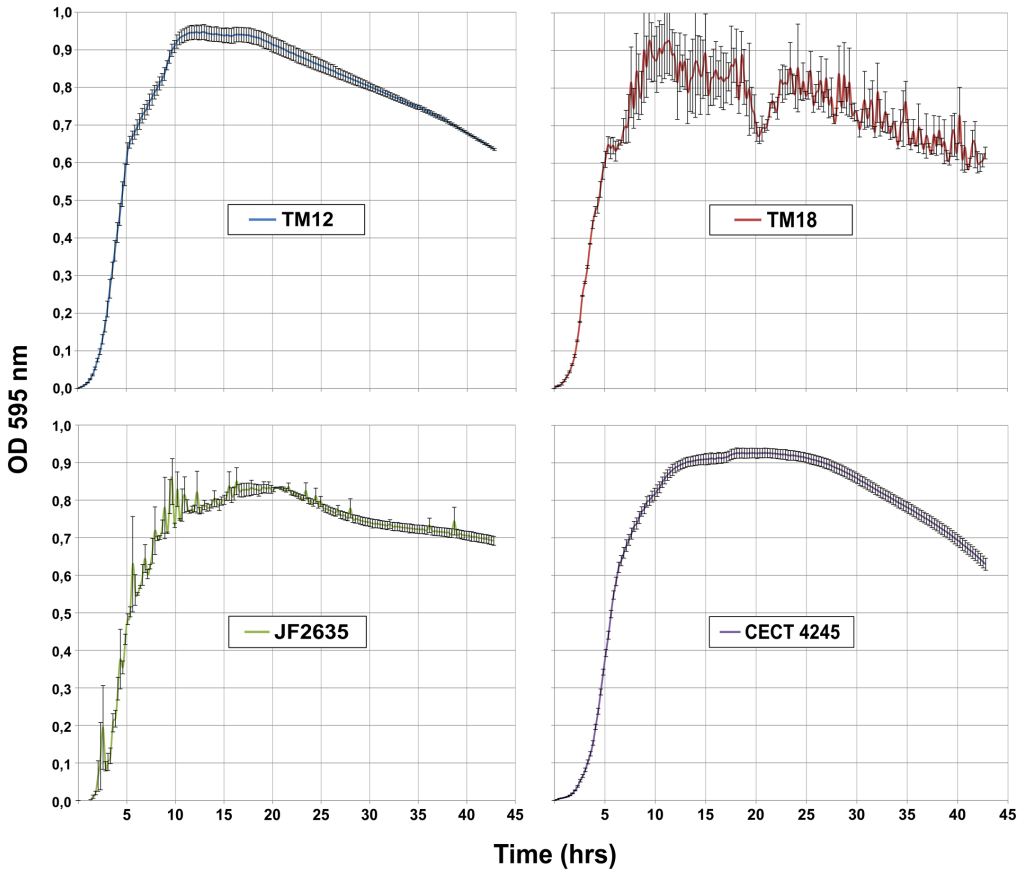


FIGURE 1 | Growth kinetics of *A. sobria* strains in LB broth at 30°C (in triplicates).

[$F_{(4,20)} = 17.87$, $p = 2.20 \times 10^{-6}$) (**Figure 2**). There is a strong interaction between the growth medium and the ability of *A. sobria* strains to produce biofilms [$F_{(4,20)} = 16.30$, $p = 4.39 \times 10^{-6}$).

Post hoc multiple comparisons revealed that *A. sobria* strains TM12, TM18 and CECT 4245 produced significant levels of biofilm compared to the no-cell control (Tukey's HSD test, $p \leq 0.021$), but with marginal between-strain differences (Tukey's HSD test, $p \geq 0.084$). Biofilm-producing strains do better in LB than in TSB (Tukey's HSD test, $p = 10^{-7}$). Strain JF2635 produced no detectable amount of biofilm over the no-cell control (Tukey's HSD test, $p = 0.65$).

RAST genome annotations of *A. sobria* strains revealed that strains TM18 and CECT 4245 lack the *pga* operon required for the biosynthesis of biofilm adhesin poly-β-1,6-N-acetyl-D-glucosamine (**Table 3**). This finding suggests that (i) TM18 and CECT 4245 could be more vulnerable to compounds inhibiting surface attachment (i.e., surfactants), and (ii) certain nutrients present in TSB broth but not in LB broth (i.e., enzymatic soymeal digest or glucose), could act as inhibitors of biofilm formation. Indeed, mono- and diglycerides are known surfactants (Prajapati et al., 2012) and inhibitors of biofilm formation in *Aeromonas* (Ham and Kim, 2016) that can be biologically synthesized in high abundance from enzymatic soymeal digests, due to its high phospholipid content (Küllenberg et al., 2012).

Interestingly, *A. sobria* strains possess several phospholipase genes that could produce surfactant metabolites (**Table 3**). Strains TM12, TM18 and CECT 4245 are likely able to produce (i) diacylglycerol (DAG) *via* phosphatidylcholine-specific phospholipase C, and (ii) free fatty acids *via* lysophospholipase L2. The latter are known to play a role as signal molecules involved in either biofilm formation or dispersion (Marques et al., 2015), and could lead to biofilm inhibition. Strain JF2635 (which produces no biofilm in either LB or TSB) possesses a phospholipase A1 gene but lacks lysophospholipase. In a liquid broth rich in glycerophospholipids such as TSB, this may result in a buildup of extracellular lysophospholipids which have strong detergent properties (Huang et al., 1998).

Heterogeneity in the *A. sobria* Species Pangenome

There is significant quantitative and qualitative heterogeneity among the four *A. sobria* strains of this study in terms of basic phenotypic traits such as (i) antagonism against another *Aeromonas* species; (ii) growth kinetics and iii) production of biofilms in different growth conditions. Knowing that those four strains were isolated from fairly different backgrounds

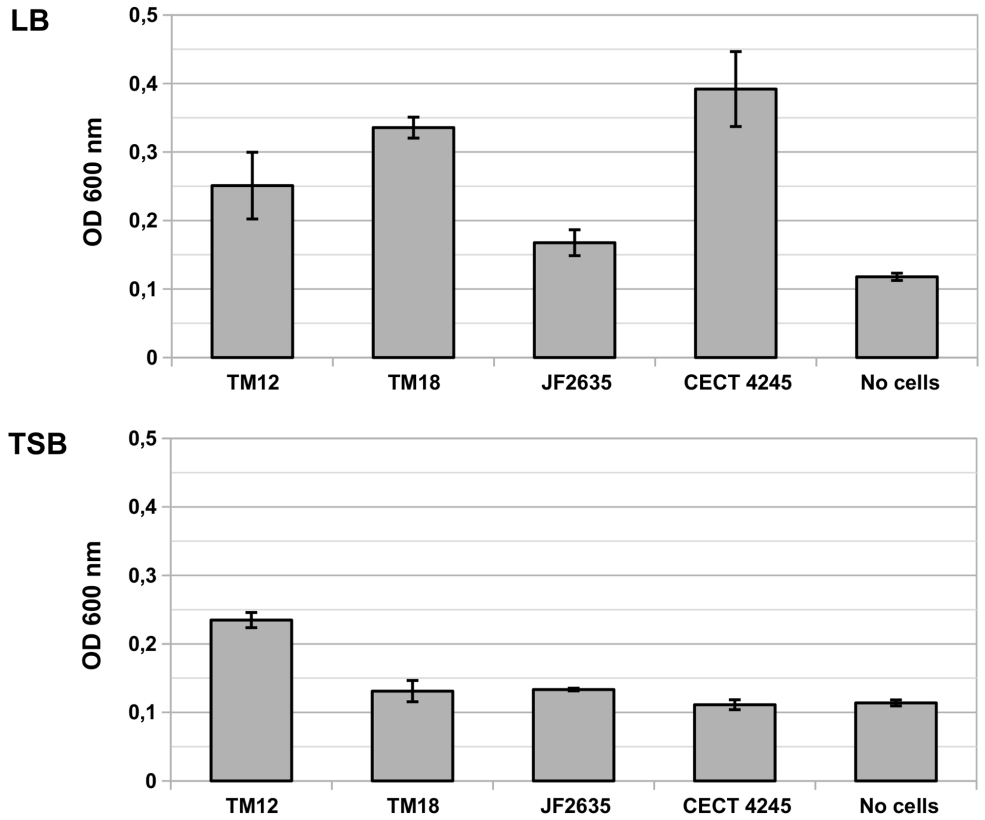


FIGURE 2 | Biofilm production in liquid cultures of *A. sobria* over 6 h static incubation at 30°C in LB and TSB broth. Bars indicate the OD_{600} of crystal violet that adsorbed in biofilms. Vertical segments indicate the standard error of the mean. This experiment was performed in triplicates.

(**Table 1**), this heterogeneity may be underlain by strong genomic divergence resulting from adaptation to different niches. It was therefore tempting to verify if the phylogenomic clustering among the *A. sobria* strains studied here also reflects this heterogeneity.

Core and Accessory Genomes

The pangenome of *A. sobria* strains studied here exhibits a species-wise proportion of accessory genes of 2,084/5,586 = 37.3% (**Table 4**). This is lower than reported values for other *Aeromonas* species complexes, where accessory genes represent 50–70% of the pangenome (see Introduction). This low proportion can be explained by the scarcity of sequence data for *A. sobria* (five genomes including those introduced in this publication). Indeed, a development plot of the core genome (i.e., a fit of core genome size *vs.* number of subsampled genomes) reveals that an asymptote has not yet been reached (i.e., the number of core genes will decrease by adding more genomes; Supplementary Figure 3A). Conversely, a development plot shows that the pangenome is "open," i.e., it would increase if more *A. sobria* genomes were included in the study (Supplementary Figure 3B) (Guimarães et al., 2015; Rouli et al., 2015). This suggests much of the phylogenomic diversity of *A. sobria* remains to be assessed, which will be solved by the addition of more whole genome data.

Molecular Phylogeny

In addition to the *A. sobria* isolates, a set of representative strains of each *Aeromonas* species with whole genome sequences available in GenBank (Supplementary Table 2) has been added to get a more accurate phylogenetic resolution of the *A. sobria* isolates. This includes the genome sequences of a fifth *A. sobria* strain, 08005. This additional strain was recovered from an infected amphibian. As expected, the five *A. sobria* isolates formed a monophyletic group (**Figure 3**).

As mentioned earlier, there is apparent confusion between *A. sobria sensu stricto* (Popoff and Véron, 1976) and *A. veronii* biovar *sobria* in the scientific literature, because of the similarity in their phenotypic profiles (Janda and Abbott, 2010; Austin and Austin, 2012b). However, the softcore genome phylogeny supports that *A. veronii* and *A. sobria* are distinct clades (**Figure 3**), as previously demonstrated by other clustering methods (Martino et al., 2011). To our knowledge, our phylogenetic assessment, which is based on 2,154 softcore gene sequences including 946,687 variable sites of 33 *Aeromonas* genomes, is the most robust and accurate phylogenetic positioning of *A. sobria* to date.

Interestingly, the *sobria* clade shares a near common ancestor with the *A. finlandiensis* species. Multilocus sequence analysis trees (7 and 15 housekeeping genes) from the paper having reported this species placed it near the species

TABLE 3 | Differential presence of genes involved in biofilm synthesis and glycerophospholipid catabolism, with special relevance to biofilm formation in *A. sobria*.

Category	RAST annotation	Major product/function	*A. sobria*			
			TM12	TM18	JF2635	CECT 4245
Biofilm adhesin biosynthesis	*pgaA* (Biofilm PGA outer membrane secretin)	Outer membrane PGA	P		P	
	pgaC [Biofilm PGA synthesis *N*-glycosyl transferase (EC 2.4.-.-)]	GlcNAc export to periplasm	P		P	
	pgaB (Biofilm PGA synthesis deacetylase (EC 3.-))	Promotes PGA export through the PgaA porin	P		P	
Glycero- phospholipid Catabolism	Phospholipase C 4 precursor (EC 3.1.4.3)	DAG* + phospholipid head group		P		P
	Putative phospholipase A1-like (EC 3.1.1.32)	Lysophospholipid* + free fatty acid	P	P	P	P
	Lysophospholipase L2 (EC 3.1.1.5)	Free fatty acid	P	P		P
Biofilm formation in LB			+	++	−	++
Biofilm formation in TSB			+	−	−	−

*PGA, Poly-beta-1,6-N-acetyl-D-glucosamine. DAG, diacylglycerol. GlcNAc, N-acetyl glucosamine. EC, Enzyme Commission number for enzymes. P, present. +, weak biofilm formation. ++, strong biofilm formation. −, no biofilm production. *known surfactants.*

A. allosaccharophila and *A. veronii,* while *A. sobria* was more basal (Beaz-Hidalgo et al., 2015). A recent study, based on two concatenated gene sequences reported *A. sobria* forming a clade along with *A. allosaccharophila* while *A. veronii* was predicted to share a recent common ancestor with the one of *A. finlandiensis* (Sanglas et al., 2017). In addition to the present study, two papers describing phylogenies of the *Aeromonas* genus based on core and softcore genomes have been published (Colston et al., 2014; Vincent et al., 2016). Unfortunately, these two publications did not include *A. finlandiensis* because those two studies were initiated prior to its description (Beaz-Hidalgo et al., 2015). We believe that more complete genomes of strains from the *A. finlandiensis* species is required to have a clearer taxonomic positioning relative to *A. sobria*.

Average Nucleotide Identity

The average nucleotide identity (ANI) is known to be a gold standard to determine the relatedness of bacterial species, where a value of ∼95–96% correlates with the ∼70–75% DNA:DNA hybridization threshold used as a gold standard to define prokaryotic species (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Colston et al., 2014; Federhen et al., 2016). The ANI values confirmed that the five *A. sobria* isolates are members of the same species (**Figure 3**). This analysis, in addition to the short phylogenetic branch lengths, showed that the strains 08005 and CECT 4245 (hereby referred to as "Clade 1") are evolutionarily close (Shared ANI: 99.9%). It is worth noting that strains JF2635, TM12 and TM18 (hereby referred to as "Clade 2"), which all grouped together in the softcore phylogeny, exhibited substantial nucleotide diversity (Shared ANI: 96.4 ± 0.4%). The split corresponding to both clades is strongly supported with a bootstrap score of 100.

TABLE 4 | Core and accessory gene counts across *A. sobria* strains of this study.

A. sobria	Cloud	Shell	Soft core	Core	Total
08005	461	169	3,502	3,339	4,132
CECT 4245	455	170	3,502	3,339	4,127
JF2635	667	161	3,425	3,339	4,330
TM12	412	93	3,455	3,339	3,960
TM18	505	142	3,471	3,339	4,118
Total *	2,084		3,502	3,339	5,586

*Cloud: present in 1 or 2 genomes only. Core: present in all genomes. Softcore: present in 95% of all genomes. Shell: all remaining genes. *Column totals are not necessarily sums of counts for each strain, since certain elements are shared between one or more strains.*

Codon Usage

The genomic dissimilarity underlying the split into two clades was striking as it resulted mostly from the number of tRNA genes encoded by these genomes (**Table 5**). Clade 1 isolates harbored 20–22% less tRNA genes than clade 2. Given this extensive dichotomy in tRNA genes, it was reasonable to hypothesize that some codons could be preferred, depending on the clade. The relative synonymous codon usage (RSCU) was found for each set of genes and the result was analyzed by a principal component analysis (PCA) in which the isolates were distributed as expected as in the phylogenetic tree (i.e., in two distinct groups), suggesting that a codon bias exists depending on the clade (**Figure 4**). Furthermore, the PCA also confirmed the more important heterogeneity within clade 2, previously evidenced in other analyses.

Antibiotic Resistance Genes

Bacterial genomes harbor various key genes to enhance their fitness, including drug resistance and virulence factors. In *A. sobria*, little is known about the pool of coding genes used for antibiotic resistance mechanisms and to colonize new environments. Thorough genomic sequence investigation

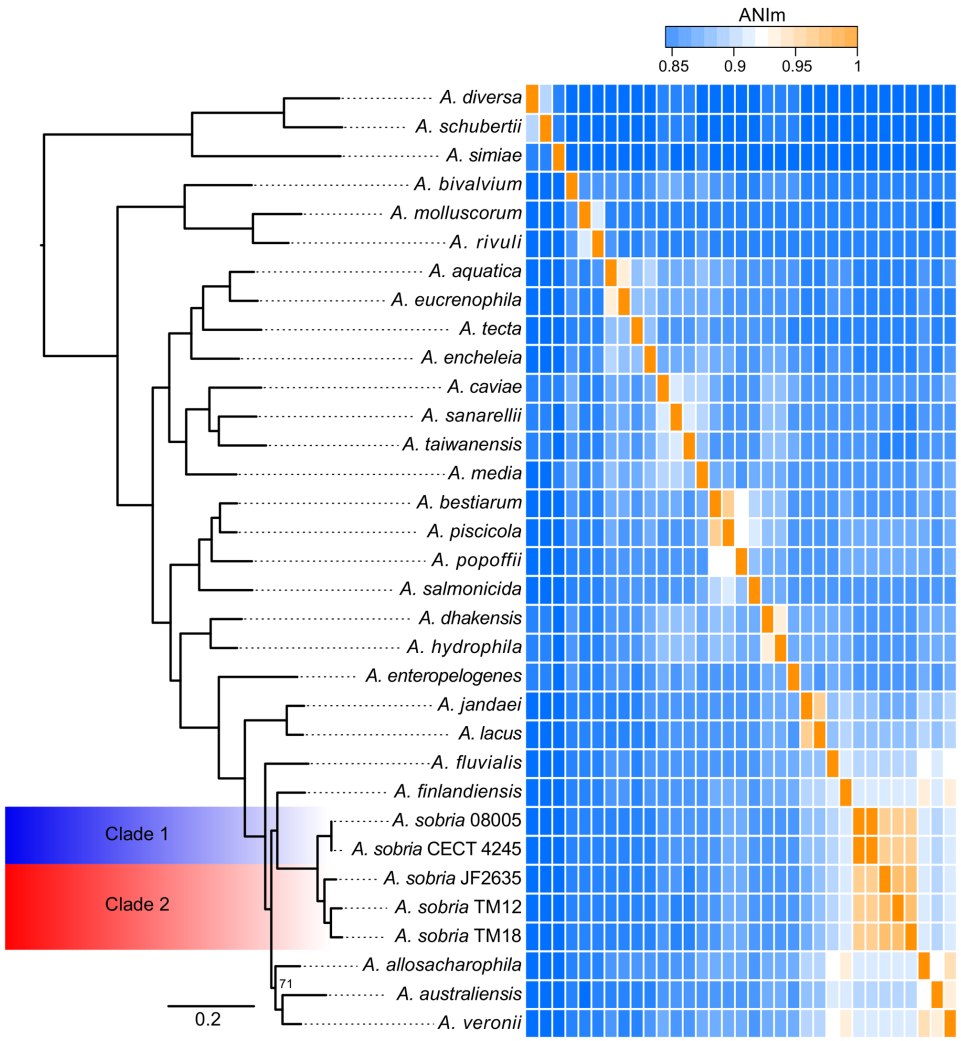


FIGURE 3 | Phylogenomic tree of the *Aeromonas* softcore genome (2,154 genes present in at least 95% of 33 *Aeromonas* genomes), coupled to an ANIm analysis of the *Aeromonas* genus with an emphasis on the *sobria* species. All nodes are supported by bootstrap values of 100, excepted the one of *allosaccharophila*, which is 71. The ANIm heatmap is a square matrix; rows and columns are ordered identically.

TABLE 5 | Genome sequences of *Aeromonas sobria* strains used in the present study.

Strain	Contigs	N50 (pb)	Coverage (x)	GC (%)	CDS	tRNA	GenBank	Ref.
TM12	98	136 123	112.26	57.81	4036	115	NQML00000000	This study
TM18	101	127 327	109.57	57.72	4256	116	NQMM00000000	This study
JF2635	121	128 869	75.04	57.84	4426	115	LJZX00000000	This study
CECT 4245	48	171 779	34	57.62	4223	91	NZ_CDBW00000000	a
08005	52	186 724	109.0	57.58	4219	90	NZ_MKFU00000000	b

a: Colston et al. (2014). b: Yang L. et al. (2017).

allowed to identify several genes conferring drug resistance (**Figure 5A**), many of which are coding either for efflux pumps or beta-lactam resistance proteins. This was not unsuspected knowing that aquatic environments are favorable for the spread of antibiotic resistance genes (Baquero et al., 2008), and that aeromonads are documented to be effective vectors for such genes (Henriques et al., 2006; Vincent et al., 2014; Piotrowska and Popowska, 2015; Trudel et al., 2016). Interestingly, there was a congruence regarding the phylogenetic signal between antibiotic resistance genes and the overall genomic sequence (i.e., supporting the same clade 1 and clade 2 dichotomy). The sole incongruence concerns the cluster 2 root: JF2635 roots the whole genome based phylogeny while TM18 roots the antibiotic resistance genes-based clustering (**Figure 5A**). Even if

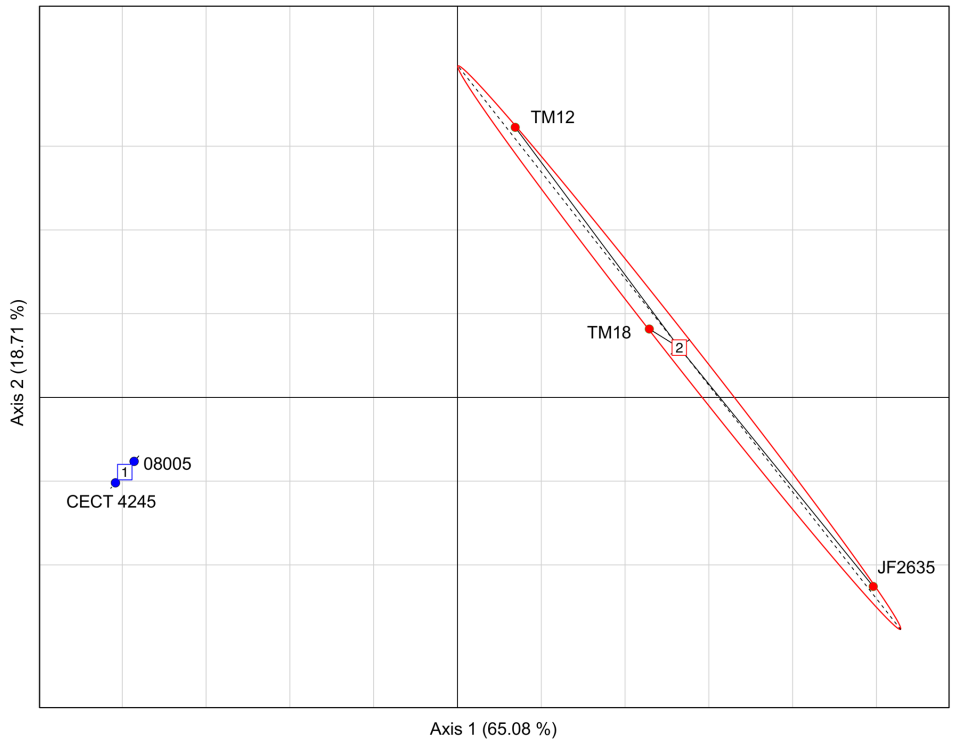


FIGURE 4 | Principal components analysis (PCA) based on the RSCU values showing a separation between isolates from clade 1 (blue) and those from clade 2 (red).

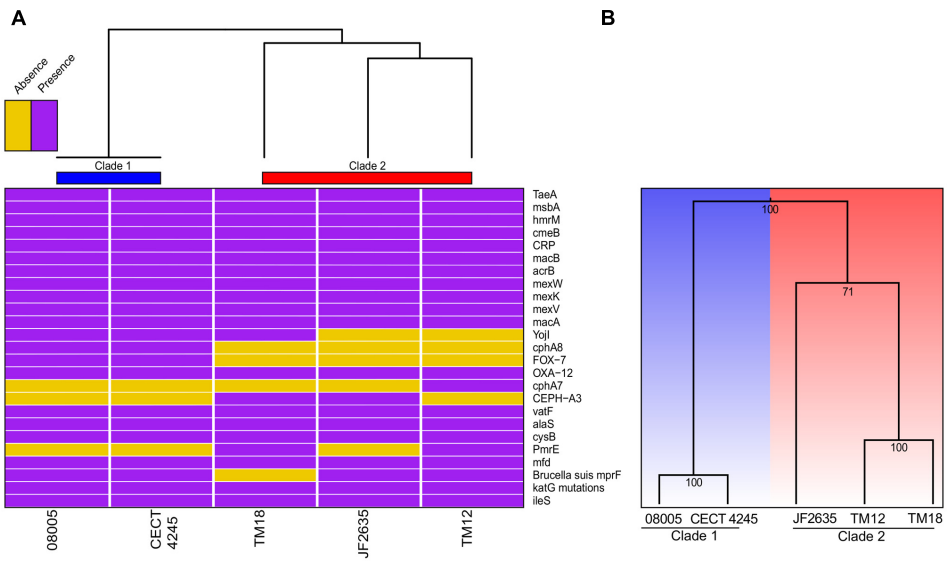


FIGURE 5 | (A) Heatmap based on the presence or absence of the antibiotic resistance genes. **(B)** Clustering based on the presence or absence of genes implicated in secretion systems.

it is perilous to draw conclusions about evolutionary history of resistance genes in *A. sobria* given the small number of markers comparatively to the molecular phylogeny, the fact that both topologies are similar lets us believe that resistance genes are not mobile and are stable.

Virulence Factors

Secretion systems are well characterized as sophisticated protein machineries widely distributed among bacteria and are, among other things, major determinants in virulence (Costa et al., 2015; Abby et al., 2016; Green and Mecsas, 2016). As for

antibiotic resistance genes, it was consequently relevant to verify the presence or absence of genes implicated in these systems (Supplementary Table 4). Here, the clustering analysis of virulence genes was even more congruent with the whole genome phylogeny than what was observed for resistance genes (**Figure 5B**). One of the most salient features was the absence of mandatory genes involved in the formation of a T6SSi, a secretion system exporting effectors to both bacterial and eukaryotic cells (Ho et al., 2014), for clade 2 isolates. A functional T6SS was previously reported in *A. hydrophila* (Suarez et al., 2008). Then, the five *A. sobria* genomes were predicted to harbor all mandatory genes for functional T1SS, T2SS, type IV pilus (T4P) and flagellum (Supplementary Table 5). Also the genomes of 08005, CECT 4245 and JF2635 were predicted to bear the single mandatory gene (coding for a protein having both the translocator and passenger domains) to have a functional T5aSS (Leo et al., 2012).

Plasmids

Of all five *A. sobria* strains included in this study, only two (TM18 and JF2635 from clade 2) harbored small high-copy-number plasmids (Supplementary Figure 4). TM18 has three plasmids ranging from 4,393 to 5,190 bp, whilst JF2635 harbors two plasmids (3,818 and 5,381 bp). Plasmids pJF2635-1 and pTM18-3 are ColE1-like replicons, as evidenced by the presence of genes encoding regulatory RNAs I and II involved in ColE1-type replication (Tomizawa, 1984). The other plasmids (pJF2635-2, pTM18-1 and pTM18-2) are ColE2-type replicons which have no RNA II gene but a RNA I gene complementary to the *repA* mRNA (Sugiyama and Itoh, 1993).

Aside from RNA I and RNA II genes, most genes found in the plasmid repertoire of *A. sobria* TM18 and JF2635 encode either hypothetical proteins or proteins involved in plasmid mobility and maintenance (Supplementary Figure 4; blue arrows). Two notable exceptions are:

(i) a putative *vapD* gene in pTM18-2, which may encode a virulence-associated protein with ssRNA endonuclease activity (Katz et al., 1992; Kwon et al., 2012). No other *vap* gene homologs were found in neither the plasmids nor the draft genome sequences.

(ii) a pseudogene in pTM18-1. A BLASTx search against the non-redundant protein database (NCBI) returned a hypothetical protein from an *Aeromonas* sp. as best hit. No putative conserved domain was detected.

Prophages

A total of nine predicted phage elements were found across the five *A. sobria* isolates (Supplementary Table 3). Seven of those prophages were found in JF2635, of which only two were presumably intact: a Phi018p-like element (Beilstein and Dreiseikelmann, 2008) and a SJ46-like element (Yang L. et al., 2017). Only one intact prophage, a Fels-2-like element is found in both CECT 4245 and 08005 strains (clade 1), whilst only one prophage, a 9.5 kb Phi018p-like element (presumably incomplete) was found in both TM12 and TM18 strains (clade 2). There is a clear dichotomy in terms of prophage contents between strains from clade 1 (CECT 4245 and 08005) and strains from clade 2 (TM12, TM18 and JF2635). Even among clade 2 strains, there is a split, with JF2635 having substantially more phage elements (Supplementary Figure 5). The presence of many degenerated elements in JF2635 suggests that this strain acquired prophages early in the evolutionary history of clade 2. The low number of phage elements in other *A. sobria* strains, as opposed to JF2635, requires further investigation.

CONCLUSION

Aeromonas sobria is a mesophilic motile aeromonad whose host–microbe associations, either mutualistic or pathogenic, are less understood than for other aeromonad species. We assessed the genomic and phenotypic heterogeneity among five *A. sobria* strains: two brook charr probionts (TM12 and TM18) which inhibit *in vitro* growth of *A. salmonicida* subsp. *salmonicida*, and three clinical isolates recovered from infected fish (JF2635 and CECT 4245) and an infected amphibian (08005). Comparative analysis supports a split of the *A. sobria* species complex in two clades.

- Clade 1 strains (08005 and CECT 4245) harbor no plasmids but a single intact Fels-2-like prophage. They also possess an identical antibiotic resistance gene profile. They have no inhibitory effect against *A. salmonicida* subsp. *salmonicida*.
- Clade 2 strains (TM12, TM18 and JF2635) possess 20–22% more tRNA genes than clade 1 strains, leading to major differences in relative synonymous codon usage. They harbor no Fels-2-like prophage, but minimally an incomplete Phi018p-like prophage. There are notable differences between clade 2 strains, however. Unlike TM12 and TM18, strain JF2635 has two intact and five incomplete prophages. Only TM18 and JF2635 harbor plasmids. Their antibioresistance gene and secretion gene profiles are more heterogeneous than within clade 1. Only clade 2 strains inhibit growth of *A. salmonicida* subsp. *salmonicida* (TM12 and TM18), with the exception of JF2635.

These findings illustrate how adaptation to a broad range of hosts and life strategies has shaped the evolution of the *A. sobria* species complex into two clades harboring significant within-clade and between-clade diversity. *A. sobria* has been treated as a monotypic bacterial species since its inception (Popoff and Véron, 1976; Martinez-Murcia et al., 1992; Abbott et al., 2003; Martino et al., 2011). However, the clear genomic and phenotypic division between clade 1 and clade 2 indicates that the *A. sobria* species complex may be composed of at least two candidates to subspecies status. Of course, the taxonomic assessment of *A. sobria* below the species level will be more accurate when more genome sequences will be available.

Finally, the antagonistic effect of clade 2 strains TM12 and TM18 against *A. salmonicida* subsp. *salmonicida* indicates that these strains (or their products) could lead to novel control and prevention methods to mitigate this opportunistic pathogen

of salmonid fish. This effect suggests a role of certain host-associated *A. sobria* strains in controlling the abundance of other opportunistic pathogens in their host microbiota (including other aeromonads) and deserves further investigation.

AUTHOR CONTRIBUTIONS

JG, AV, SC, and ND designed the experiments. JG and AV performed *in vitro* and *in silico* experiments. JG, AV, SC, and ND contributed to the manuscript.

FUNDING

The authors acknowledge funding from the Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (INNOVAMER Program), the Natural Sciences and Engineering Research Council of Canada (NSERC) and Ressources Aquatiques Québec (RAQ). JG received a Graduate Scholarship from the NSERC and AV received an Alexander Graham Bell Canada Graduate Scholarship from the NSERC. SC is a research scholar from the Fonds de Recherche du Québec en Santé.

ACKNOWLEDGMENTS

The authors thank Joachim Frey (University of Bern) for the JF2635 isolate, Typhaine Morvant for the isolation of strains TM12 and TM18 and Tom Van Acker for early characterization of those strains.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02434/full#supplementary-material

REFERENCES

Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38, W7–W13. doi: 10.1093/nar/gkq291

Abbott, S. L., Cheung, W. K. W., and Janda, J. M. (2003). The genus *Aeromonas*: biochemical characteristics, atypical reactions, and phenotypic identification schemes. *J. Clin. Microbiol.* 41, 2348–2357. doi: 10.1128/JCM.41.6.2348-2357.2003

Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E. P. C. (2016). Identification of protein secretion systems in bacterial genomes. *Sci. Rep.* 6:23080. doi: 10.1038/srep23080

Acinas, S. G., Antón, J., and Rodríguez-Valera, F. (1999). Diversity of free-living and attached bacteria in offshore Western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 65, 514–522.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Austin, B., and Austin, D. A. (eds). (2012a). "Aeromonadaceae representative (*Aeromonas salmonicida*)," in *Bacterial Fish Pathogens* (Dordrecht: Springer), 147–228. doi: 10.1007/978-94-007-4884-2_5

Austin, B., and Austin, D. A. (eds). (2012b). "Aeromonadaceae representatives (motile aeromonads)," in *Bacterial Fish Pathogens* (Dordrecht: Springer), 119–146. doi: 10.1007/978-94-007-4884-2_4

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

Baquero, F., Martínez, J.-L., and Cantón, R. (2008). Antibiotics and antibiotic resistance in water environments. *Curr. Opin. Biotechnol.* 19, 260–265. doi: 10.1016/j.copbio.2008.05.006

Beaz-Hidalgo, R., Latif-Eugenín, F., Hossain, M. J., Berg, K., Niemi, R. M., Rapala, J., et al. (2015). *Aeromonas aquatica* sp. nov. *Aeromonas finlandiensis* sp. nov. and *Aeromonas lacus* sp. nov. isolated from Finnish waters associated with cyanobacterial blooms. *Syst. Appl. Microbiol.* 38, 161–168. doi: 10.1016/j.syapm.2015.02.005

Beaz-Hidalgo, R., Martínez-Murcia, A., and Figueras, M. J. (2013). Reclassification of *Aeromonas hydrophila* subsp. *dhakensis* Huys et al. 2002 and *Aeromonas aquariorum* Martínez-Murcia et al. 2008 as *Aeromonas dhakensis* sp. nov. comb nov. and emendation of the species *Aeromonas hydrophila*. *Syst. Appl. Microbiol.* 36, 171–176. doi: 10.1016/j.syapm.2012.12.007

Beilstein, F., and Dreiseikelmann, B. (2008). Temperate bacteriophage PhiO18P from an *Aeromonas media* isolate: characterization and complete genome sequence. *Virology* 373, 25–29. doi: 10.1016/j.virol.2007.11.016

Boone, D. R., Castenholz, R. W., Garrity, G. M., Brenner, D. J., Krieg, N. R., and Staley, J. T. (2005). *Bergey's Manual® of Systematic Bacteriology*. Berlin: Springer.

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660. doi: 10.7717/peerj.1660

Boutin, S., Audet, C., and Derome, N. (2013). Probiotic treatment by indigenous bacteria decreases mortality without disturbing the natural microbiota of *Salvelinus fontinalis*. *Can. J. Microbiol.* 59, 662–670. doi: 10.1139/cjm-2013-0443

Boutin, S., Bernatchez, L., Audet, C., and Derôme, N. (2012). Antagonistic effect of indigenous skin bacteria of brook charr (*Salvelinus fontinalis*) against *Flavobacterium columnare* and *F. psychrophilum*. *Vet. Microbiol.* 155, 355–361. doi: 10.1016/j.vetmic.2011.09.002

Brunt, J., and Austin, B. (2005). Use of a probiotic to control lactococcosis and streptococcosis in rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J. Fish Dis.* 28, 693–701. doi: 10.1111/j.1365-2761.2005.00672.x

Brunt, J., Newaj-Fyzul, A., and Austin, B. (2007). The development of probiotics for the control of multiple bacterial diseases of rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J. Fish Dis.* 30, 573–579. doi: 10.1111/j.1365-2761.2007.00836.x

Cahill, M. M. (1990). Bacterial flora of fishes: a review. *Microb. Ecol.* 19, 21–41. doi: 10.1007/BF02015051

Chowdhury, M. A., Yamanaka, H., Miyoshi, S., and Shinoda, S. (1990). Ecology of mesophilic *Aeromonas* spp. in aquatic environments of a temperate region and relationship with some biotic and abiotic environmental parameters. *Zentralbl. Hyg. Umweltmed.* 190, 344–356.

Cipriano, R., Bullock, G., and Pyle, S. (1984). *Aeromonas hydrophila and Other Septicemias of Fish. US Fish & Wildlife Publications*. Available at: http://digitalcommons.unl.edu/usfwspubs/134

Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., and Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. *mBio* 5:e02136-14. doi: 10.1128/mBio.02136-14

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Cornick, J. W., Morrison, C. M., Zwicker, B., and Shum, G. (1984). Atypical *Aeromonas salmonicida* infection in Atlantic cod. *Gadus morhua* L. *J. Fish Dis.* 7, 495–499. doi: 10.1111/j.1365-2761.1984.tb01174.x

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., et al. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* 13, 343–359. doi: 10.1038/nrmicro3456

Criscuolo, A., and Gribaldo, S. (2010). BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from

multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210

Dallaire-Dufresne, S., Tanaka, K. H., Trudel, M. V., Lafaille, A., and Charette, S. J. (2014). Virulence, genomic features, and plasticity of *Aeromonas salmonicida* subsp. *salmonicida*, the causative agent of fish furunculosis. *Vet. Microbiol.* 169, 1–7. doi: 10.1016/j.vetmic.2013.06.025

Dray, S., and Dufour, A. B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20. doi: 10.18637/jss.v022.i04

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

Federhen, S., Rossello-Mora, R., Klenk, H.-P., Tindall, B. J., Konstantinidis, K. T., Whitman, W. B., et al. (2016). Meeting report: GenBank microbial genomic taxonomy workshop (12–13 May, 2015). *Stand. Genomic Sci.* 11:15. doi: 10.1186/s40793-016-0134-1

Figueras, M. J., Latif-Eugenín, F., Ballester, F., Pujol, I., Tena, D., Berg, K., et al. (2017). "*Aeromonas intestinalis*" and "*Aeromonas enterica*" isolated from human faeces, "*Aeromonas crassostreae*" from oyster and "*Aeromonas aquatilis*" isolated from lake water represent novel species. *New Microbes New Infect.* 15, 74–76. doi: 10.1016/j.nmni.2016.11.019

Galardini, M., Biondi, E. G., Bazzicalupo, M., and Mengoni, A. (2011). CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol. Med.* 6, 11. doi: 10.1186/1751-0473-6-11

Gao, X.-Y., Liu, Y., Miao, L.-L., Li, E.-W., Sun, G., Liu, Y., et al. (2017). Characterization and mechanism of anti-*Aeromonas salmonicida* activity of a marine probiotic strain, *Bacillus velezensis* V4. *Appl. Microbiol. Biotechnol.* 101, 3759–3768. doi: 10.1007/s00253-017-8095-x

Garduño, R. A., and Kay, W. W. (1992). Interaction of the fish pathogen *Aeromonas salmonicida* with rainbow trout macrophages. *Infect. Immun.* 60, 4612–4620.

Garduño, R. A., Thornton, J. C., and Kay, W. W. (1993). Fate of the fish pathogen *Aeromonas salmonicida* in the peritoneal cavity of rainbow trout. *Can. J. Microbiol.* 39, 1051–1058. doi: 10.1139/m93-159

Gauthier, J., Charette, S. J., and Derome, N. (2017). Draft genome sequence of *Pseudomonas fluorescens* ML11A, an endogenous strain from brook charr with antagonistic properties against *Aeromonas salmonicida* subsp. *salmonicida*. *Genome Announc.* 5:e01716-16. doi: 10.1128/genomeA.01716-16

Ghatak, S., Blom, J., Das, S., Sanjukta, R., Puro, K., Mawlong, M., et al. (2016). Pan-genome analysis of *Aeromonas hydrophila*, *Aeromonas veronii* and *Aeromonas caviae* indicates phylogenomic diversity and greater pathogenic potential for *Aeromonas hydrophila*. *Antonie Van Leeuwenhoek* 109, 945–956. doi: 10.1007/s10482-016-0693-6

Gibson, L. F., Woodworth, J., and George, A. M. (1998). Probiotic activity of *Aeromonas media* on the Pacific oyster, *Crassostrea gigas*, when challenged with *Vibrio tubiashii*. *Aquaculture* 169, 111–120. doi: 10.1016/S0044-8486(98)00369-X

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Goulden, E. F., Hall, M. R., Pereg, L. L., and Høj, L. (2012). Identification of an antagonistic probiotic combination protecting ornate spiny lobster (*Panulirus ornatus*) larvae against *Vibrio owensii* infection. *PLOS ONE* 7:e39667. doi: 10.1371/journal.pone.0039667

Green, E. R., and Mecsas, J. (2016). Bacterial secretion systems: an overview. *Microbiol. Spectr.* 4:VMBF-0012-2015. doi: 10.1128/microbiolspec.VMBF-0012-2015

Guimarães, L. C., Florczak-Wyspianska, J., de Jesus, L. B., Viana, M. V. C., Silva, A., Ramos, R. T. J., et al. (2015). Inside the pan-genome - methods and software overview. *Curr. Genomics* 16, 245–252. doi: 10.2174/1389202916666150423002311

Gunasekara, R. A., Cornillie, P., Casteleyn, C., De Spiegelaere, W., Sorgeloos, P., Simoens, P., et al. (2011). Stereology and computer assisted three-dimensional reconstruction as tools to study probiotic effects of *Aeromonas hydrophila* on the digestive tract of germ-free *Artemia franciscana* nauplii. *J. Appl. Microbiol.* 110, 98–105. doi: 10.1111/j.1365-2672.2010.04862.x

Ham, Y., and Kim, T.-J. (2016). Inhibitory activity of monoacylglycerols on biofilm formation in *Aeromonas hydrophila*, *Streptococcus mutans*, *Xanthomonas oryzae*, and *Yersinia enterocolitica*. *Springerplus* 5:1526. doi: 10.1186/s40064-016-3182-5

Han, H.-J., Kim, D.-Y., Kim, W.-S., Kim, C.-S., Jung, S.-J., Oh, M.-J., et al. (2011). Atypical *Aeromonas salmonicida* infection in the black rockfish, *Sebastes schlegeli* Hilgendorf, in Korea. *J. Fish Dis.* 34, 47–55. doi: 10.1111/j.1365-2761.2010.01217.x

Hazen, T. C., Fliermans, C. B., Hirsch, R. P., and Esch, G. W. (1978). Prevalence and distribution of *Aeromonas hydrophila* in the United States. *Appl. Environ. Microbiol.* 36, 731–738.

Henriques, I. S., Fonseca, F., Alves, A., Saavedra, M. J., and Correia, A. (2006). Occurrence and diversity of integrons and beta-lactamase genes among ampicillin-resistant isolates from estuarine waters. *Res. Microbiol.* 157, 938–947. doi: 10.1016/j.resmic.2006.09.003

Hickman-Brenner, F. W., MacDonald, K. L., Steigerwalt, A. G., Fanning, G. R., Brenner, D. J., and Farmer, J. J. (1987). *Aeromonas veronii*, a new ornithine decarboxylase-positive species that may cause diarrhea. *J. Clin. Microbiol.* 25, 900–906.

Ho, B. T., Dong, T. G., and Mekalanos, J. J. (2014). A view to a kill: the bacterial type VI secretion system. *Cell Host Microbe* 15, 9–21. doi: 10.1016/j.chom.2013.11.008

Huang, P., Liu, Q., and Scarborough, G. A. (1998). Lysophosphatidylglycerol: a novel effective detergent for solubilizing and purifying the cystic fibrosis transmembrane conductance regulator. *Anal. Biochem.* 259, 89–97. doi: 10.1006/abio.1998.2633

Janda, J. M., and Abbott, S. L. (2010). The genus *Aeromonas*: taxonomy. pathogenicity, and infection. *Clin. Microbiol. Rev.* 23, 35–73. doi: 10.1128/CMR.00039-09

Katz, M. E., Strugnell, R. A., and Rood, J. I. (1992). Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*. *Infect. Immun.* 60, 4586–4592.

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V., et al. (2010). A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* 26, 1481–1487. doi: 10.1093/bioinformatics/btq229

Küllenberg, D., Taylor, L. A., Schneider, M., and Massing, U. (2012). Health effects of dietary phospholipids. *Lipids Health Dis.* 11:3. doi: 10.1186/1476-511X-11-3

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Kwon, A.-R., Kim, J.-H., Park, S. J., Lee, K.-Y., Min, Y.-H., Im, H., et al. (2012). Structural and biochemical characterization of HP0315 from *Helicobacter pylori* as a VapD protein with an endoribonuclease activity. *Nucleic Acids Res.* 40, 4216–4228. doi: 10.1093/nar/gkr1305

Larsen, J. L., and Pedersen, K. (1996). Atypical *Aeromonas salmonicida* isolated from diseased turbot (*Scophtalmus maximus* L.). *Acta Vet. Scand.* 37, 139–146.

Leo, J. C., Grin, I., and Linke, D. (2012). Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1088–1101. doi: 10.1098/rstb.2011.0208

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Marques, C. N. H., Davies, D. G., and Sauer, K. (2015). Control of biofilms with the fatty acid signaling molecule cis-2-decenoic acid. *Pharmaceuticals* 8, 816–835. doi: 10.3390/ph8040816

Martinez-Murcia, A. J., Benlloch, S., and Collins, M. D. (1992). Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. *Int. J. Syst. Bacteriol.* 42, 412–421. doi: 10.1099/00207713-42-3-412

Martino, M. E., Fasolato, L., Montemurro, F., Rosteghin, M., Manfrin, A., Patarnello, T., et al. (2011). Determination of microbial diversity of *Aeromonas* strains on the basis of multilocus sequence typing, phenotype, and presence of putative virulence genes. *Appl. Environ. Microbiol.* 77, 4986–5000. doi: 10.1128/AEM.00708-11

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13

McBain, A. J. (eds). (2009). "Chapter 4 in vitro biofilm models: an overview," in *Advances in Applied Microbiology* (Cambridge, MA: Academic Press), 99–132. doi: 10.1016/S0065-2164(09)69004-3

Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

O'Toole, G. A. (2011). Microtiter dish biofilm formation assay. *J. Vis. Exp.* 47:2437. doi: 10.3791/2437

Pieters, N., Brunt, J., Austin, B., and Lyndon, A. R. (2008). Efficacy of in-feed probiotics against *Aeromonas bestiarum* and *Ichthyophthirius multifiliis* skin infections in rainbow trout (*Oncorhynchus mykiss*, Walbaum). *J. Appl. Microbiol.* 105, 723–732. doi: 10.1111/j.1365-2672.2008.03817.x

Piotrowska, M., and Popowska, M. (2015). Insight into the mobilome of *Aeromonas* strains. *Front. Microbiol.* 6:494. doi: 10.3389/fmicb.2015.00494

Popoff, M., and Véron, M. (1976). A taxonomic study of the *Aeromonas hydrophila-Aeromonas punctata* group. *Microbiology* 94, 11–22. doi: 10.1099/00221287-94-1-11

Prajapati, H. N., Dalrymple, D. M., and Serajuddin, A. T. M. (2012). A comparative evaluation of mono-, Di- and triglyceride of medium chain fatty acids by lipid/surfactant/water phase diagram, solubility determination and dispersion testing for application in pharmaceutical dosage form development. *Pharm. Res.* 29, 285–305. doi: 10.1007/s11095-011-0541-3

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., and Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 25, 2071–2073. doi: 10.1093/bioinformatics/btp356

Rouli, L., Merhej, V., Fournier, P.-E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7, 72–85. doi: 10.1016/j.nmni.2015.06.005

Sanglas, A., Albarral, V., Farfán, M., Lorén, J. G., and Fusté, M. C. (2017). Evolutionary roots and diversification of the genus *Aeromonas*. *Front. Microbiol.* 8:127. doi: 10.3389/fmicb.2017.00127

Schachte, J. H. (2002). *Furunculosis*. Long Island City, NY: New York Department of Environmental Conservatory.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Schubiger, C. B., Orfe, L. H., Sudheesh, P. S., Cain, K. D., Shah, D. H., and Call, D. R. (2015). Entericidin is required for a probiotic treatment (*Enterobacter* sp. strain C6-6) to protect trout from cold-water disease challenge. *Appl. Environ. Microbiol.* 81, 658–665. doi: 10.1128/AEM.02965-14

Scott, M. (1968). The pathogenicity of *Aeromonas salmonicida* (Griffin) in sea and brackish waters. *J. Gen. Microbiol.* 50, 321–327. doi: 10.1099/00221287-50-2-321

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shao, Y., Harrison, E. M., Bi, D., Tai, C., He, X., Ou, H.-Y., et al. (2011). TADB: a web-based resource for Type 2 toxin–antitoxin loci in bacteria and archaea. *Nucleic Acids Res.* 39, D606–D611. doi: 10.1093/nar/gkq908

Suarez, G., Sierra, J. C., Sha, J., Wang, S., Erova, T. E., Fadl, A. A., et al. (2008). Molecular characterization of a functional type VI secretion system from a clinical isolate of *Aeromonas hydrophila*. *Microb. Pathog.* 44, 344–361. doi: 10.1016/j.micpath.2007.10.005

Sugiyama, T., and Itoh, T. (1993). Control of ColE2 DNA replication: *in vitro* binding of the antisense RNA to the Rep mRNA. *Nucleic Acids Res.* 21, 5972–5977. doi: 10.1093/nar/21.25.5972

Talagrand-Reboul, E., Jumas-Bilak, E., and Lamy, B. (2017a). The social life of *Aeromonas* through biofilm and quorum sensing systems. *Front. Microbiol.* 8:37. doi: 10.3389/fmicb.2017.00037

Talagrand-Reboul, E., Roger, F., Kimper, J.-L., Colston, S. M., Graf, J., Latif-Eugenín, F., et al. (2017b). Delineation of taxonomic species within complex of species: *Aeromonas media* and related species as a test case. *Front. Microbiol.* 8:621. doi: 10.3389/fmicb.2017.00621

Tam, B. (2009). *A Climate Change Impact Assessment on the Spread of Furunculosis in the Ouje-Bougoumou Region*. Available at: https://tspace.library.utoronto.ca/handle/1807/17227 [accessed September 14, 2017].

Tam, B., Gough, W. A., and Tsuji, L. (2011). The impact of warming on the appearance of furunculosis in fish of the James Bay region, Quebec, Canada. *Reg. Environ. Change* 11, 123–132. doi: 10.1007/s10113-010-0122-8

Tomizawa, J. (1984). Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript. *Cell* 38, 861–870. doi: 10.1016/0092-8674(84)90281-2

Tritt, A., Eisen, J. A., Facciotti, M. T., and Darling, A. E. (2012). An integrated pipeline for *de novo* assembly of microbial genomes. *PLOS ONE* 7:e42304. doi: 10.1371/journal.pone.0042304

Trudel, M. V., Vincent, A. T., Attéré, S. A., Labbé, M., Derome, N., Culley, A. I., et al. (2016). Diversity of antibiotic-resistance genes in Canadian isolates of *Aeromonas salmonicida* subsp. *salmonicida*: dominance of pSN254b and discovery of pAsa8. *Sci. Rep.* 6:35617. doi: 10.1038/srep35617

Vanden Bergh, P., and Frey, J. (2013). *Aeromonas salmonicida* subsp. *salmonicida* in the light of its type-three secretion system. *Microb. Biotechnol.* 7, 381–400. doi: 10.1111/1751-7915.12091

Vincent, A. T., and Charette, S. J. (2017). Phylogenetic analysis of the fish pathogen *Aeromonas salmonicida* underlines the dichotomy between European and Canadian strains for the salmonicida subspecies. *J. Fish Dis.* 40, 1241–1247. doi: 10.1111/jfd.12595

Vincent, A. T., Trudel, M. V., Freschi, L., Nagar, V., Gagné-Thivierge, C., Levesque, R. C., et al. (2016). Increasing genomic diversity and evidence of constrained lifestyle evolution due to insertion sequences in *Aeromonas salmonicida*. *BMC Genomics* 17:44. doi: 10.1186/s12864-016-2381-3

Vincent, A. T., Trudel, M. V., Paquet, V. E., Boyle, B., Tanaka, K. H., Dallaire-Dufresne, S., et al. (2014). Detection of variants of the pRAS3, pAB5S9, and pSN254 plasmids in *Aeromonas salmonicida* subsp. *salmonicida*: multidrug resistance, Interspecies exchanges, and plasmid reshaping. *Antimicrob. Agents Chemother.* 58, 7367–7374. doi: 10.1128/AAC.03730-14

Wahli, T., Burr, S. E., Pugovkin, D., Mueller, O., and Frey, J. (2005). *Aeromonas sobria*, a causative agent of disease in farmed perch, *Perca fluviatilis* L. *J. Fish Dis.* 28, 141–150. doi: 10.1111/j.1365-2761.2005.00608.x

Xia, X. (2017). DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. *J. Hered.* 108, 431–437. doi: 10.1093/jhered/esx033

Yang, L., Li, W., Jiang, G.-Z., Zhang, W.-H., Ding, H.-Z., Liu, Y.-H., et al. (2017). Characterization of a P1-like bacteriophage carrying CTX-M-27 in *Salmonella* spp. resistant to third generation cephalosporins isolated from pork in China. *Sci. Rep.* 7:40710. doi: 10.1038/srep40710

Yang, Q.-H., Zhou, C., Lin, Q., Lu, Z., He, L.-B., and Guo, S.-L. (2017). Draft genome sequence of *Aeromonas sobria* strain 08005, isolated from sick *Rana catesbeiana*. *Genome Announc.* 5:e01352-16. doi: 10.1128/genomeA.01352-16

Zhang, D., Xu, D.-H., Qiu, J., Rasmussen-Ivey, C. R., Liles, M. R., and Beck, B. H. (2016). Draft genome sequence of *Pseudomonas mosselii* Gil3, isolated from catfish and antagonistic against hypervirulent *Aeromonas hydrophila*. *Genome Announc.* 4:e01305-16. doi: 10.1128/genomeA.01305-16

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 Gauthier, Vincent, Charette 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) 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 April 2018
doi: 10.3389/fmicb.2018.00805

A Polyphasic and Taxogenomic Evaluation Uncovers *Arcobacter cryaerophilus* as a Species Complex That Embraces Four Genomovars

Alba Pérez-Cataluña[1], Luis Collado[2], Oscar Salgado[2,3], Violeta Lefiñanco[2] and María J. Figueras[1]**

[1] Unit of Microbiology, Department of Basic Health Sciences, Faculty of Medicine and Health Sciences, IISPV, University Rovira i Virgili, Reus, Spain, [2] Faculty of Sciences, Institute of Biochemistry and Microbiology, Universidad Austral de Chile, Valdivia, Chile, [3] Laboratory of Microbial Ecology of Extreme Systems, Department of Molecular Genetics and Microbiology, Pontificia Universidad Católica de Chile, Santiago, Chile

OPEN ACCESS

Edited by:
Antonio Ventosa,
Universidad de Sevilla, Spain

Reviewed by:
Jason Sahl,
Northern Arizona University, United States
David John Studholme,
University of Exeter, United Kingdom

***Correspondence:**
Luis Collado
luiscollado@uach.cl
María J. Figueras
mariajose.figueras@urv.cat

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

Received: 28 January 2018
Accepted: 10 April 2018
Published: 27 April 2018

Citation:
Pérez-Cataluña A, Collado L, Salgado O, Lefiñanco V and Figueras MJ (2018) A Polyphasic and Taxogenomic Evaluation Uncovers Arcobacter cryaerophilus as a Species Complex That Embraces Four Genomovars. Front. Microbiol. 9:805. doi: 10.3389/fmicb.2018.00805

The species *Arcobacter cryaerophilus* is found in many food products of animal origin and is the dominating species in wastewater. In addition, it is associated with cases of farm animal and human infectious diseases,. The species embraces two subgroups i.e., 1A (LMG 24291^T = LMG 9904^T) and 1B (LMG 10829) that can be differentiated by their 16S rRNA-RFLP pattern. However, some authors, on the basis of the shared intermediate levels of DNA-DNA hybridization, have suggested abandoning the subgroup classification. This contradiction indicates that the taxonomy of this species is not yet resolved. The objective of the present study was to perform a taxonomic evaluation of the diversity of *A. cryaerophilus.* Genomic information was used along with a Multilocus Phylogenetic Analysis (MLPA) and phenotypic characterization on a group of 52 temporally and geographically dispersed strains, coming from different types of samples and hosts from nine countries. The MLPA analysis showed that those strains formed four clusters (I–IV). Values of Average Nucleotide Identity (ANI) and *in silico* DNA-DNA Hybridization (*is*DDH) obtained between 13 genomes representing strains of the four clusters were below the proposed cut-offs of 96 and 70%, respectively, confirming that each of the clusters represented a different genomic species. However, none of the evaluated phenotypic tests enabled their unequivocal differentiation into species. Therefore, the genomic delimited clusters should be considered genomovars of the species *A. cryaerophilus.* These genomovars could have different clinical importance, since only the cluster I included strains isolated from human specimens. The discovery of at least one stable distinctive phenotypic character would be needed to define each cluster or genomovar as a different species. Until then, we propose naming them "*A. cryaerophilus gv. pseudocryaerophilus*" (Cluster I = LMG 10229^T), "*A. cryaerophilus gv. crypticus*" (Cluster II = LMG 9065^T), "*A. cryaerophilus gv. cryaerophilus"* (Cluster III = LMG 24291^T) and "*A. cryaerophilus gv. occultus*" (Cluster IV = LMG 29976^T).

Keywords: *Arcobacter cryaerophilus*, *is*DDH, ANI, MLPA, genomovar

INTRODUCTION

The genus *Arcobacter,* within the family *Campylobacteraceae*, was proposed by Vandamme et al. (1991) to reclassify two species that were, at that time, assigned to the genus *Campylobacter* i.e., *Campylobacter nitrofigilis* (*Arcobacter nitrofigilis*, that was selected as the representative or the type species of the genus) and *Campylobacter cryaerophila* (now *Arcobacter cryaerophilus*). The phenotypic characteristics that differentiate *Campylobacter* and *Arcobacter* are the ability of the latter to grow in aerobic conditions and at lower temperatures (Vandamme et al., 1991; Collado and Figueras, 2011).

Using more than 4000 genomes, Waite et al. (2017) recently analyzed the 16S and 23S rRNA genes and 120 protein sequences and as a result they moved the *Epsilonproteobacteria* to the phylum level with the name Epsilonbacteraeota. In addition, they created a new family *Arcobacteraceae* that includes only the genus *Arcobacter*. Currently, the genus *Arcobacter* includes 27 species (Park et al., 2016; Whiteduck-Léveillée et al., 2016; Diéguez et al., 2017; Figueras et al., 2017; Tanaka et al., 2017; Pérez-Cataluña et al., 2018), four of which have been linked with human disease: *Arcobacter butzleri*, *A. cryaerophilus*, *A. thereius,* and *A. skirrowii* (Collado and Figueras, 2011; Figueras et al., 2014; Ferreira et al., 2015). The species *A. cryaerophilus* has been found in many food products of animal origin (like poultry, pork, lamb, and seafood and in dairy food processing facilities (Collado et al., 2008; Collado and Figueras, 2011).

On the basis of the different Restriction Fragment Length Polymorphism (RFLP) of the 16S and 23S rRNA genes, Kiehlbauch et al. (1991) and Vandamme et al. (1992) divided the species *A. cryaerophilus* into two subgroups, subgroup 1 or 1A and subgroup 2 or 1B (from here on we will call them subgroups 1A and 1B), represented by strains LMG 24291^T (=LMG 9904^T) and LMG 10829, respectively. Additionally, it was demonstrated that the two subgroups showed different whole-cell protein and fatty acid contents (Vandamme et al., 1992) and clustered apart by their Amplified Fragment Length Polymorphism (AFLP) patterns (On et al., 2003). A 16S rDNA-RFLP identification method established the separation of the subgroups on the basis of their restriction patterns (Figueras et al., 2008). Despite strains belonging to both subgroups having been found at the same time in animal and human clinical samples and in food products, 1B is generally much more frequently found than 1A (Collado and Figueras, 2011 and references therein). In 2010, Debruyne et al. (2010) reassessed the taxonomy of these two subgroups of *A. cryaerophilus* using 59 strains isolated mainly from aborted animals (74% of the strains) and human faces (19%). The clustering of the strains obtained by AFLP and by the phylogenetic analysis of the *cpn60* gene, together with the shared intermediate levels of DNA-DNA hybridization observed between the strains lead the authors to conclude that despite *A. cryaerophilus* having a complex taxonomy, the subgroup nomenclature should be abandoned (Debruyne et al., 2010). Furthermore, it was considered that the type strain (LMG 24291^T = LMG 9904^T) of *A. cryaerophilus* was not representative of the species because it corresponded with the less abundant 1A subgroup. They therefore proposed that it should be changed for the strain LMG 10829, representative of subgroup 1B (Debruyne et al., 2010). However, a recent metagenomic analysis of *Arcobacter* populations recovered from sewage samples of the wastewater treatment plant in the city of Reus (Spain) and from various cities of the United States gave evidence that both *A. cryaerophilus* subgroups (1A and 1B) were dominating in this environment (Fisher et al., 2014). In addition, a different prevalence of the two *A. cryaerophilus* subgroups was found depending on the wastewater temperature, 1B dominating in wastewater samples with temperatures above 20°C. Fisher et al. (2014) concluded that this finding is relevant because understanding the ecological factors that affect the fate of *Arcobacter* spp. in wastewater may help to better understand the risks associated with these emerging pathogens. The latter study showed that both subgroups of *A. cryaerophilus* were abundant and represented two different ecotypes. Therefore, based on those findings, a new polyphasic re-evaluation of the taxonomic diversity of this species is required. The aim of the present study was to investigate the taxonomy of *A. cryaerophilus,* evaluating strains from 9 different countries recovered from wastewater, different types of shellfish, human faces and various types of animal samples (feces, various viscera from fetuses, uterus, and milk). To our knowledge, this is the most diverse collection of strains of this species studied so far. The polyphasic study involved a phylogenetic analysis of the sequences of the 16S and 23S rRNA genes and of several housekeeping genes, an analysis of 13 genomes (7 of which were obtained in this study) from a representative strains and a phenotypic characterization.

Abbreviations: LMG, Laboratorium voor Microbiologie, Universiteit Gent, Belgium Culture Collection; MLPA, Multilocus Phylogenetic Analysis; ANI, Average Nucleotide Identity; *is*DDH, *in silico* DNA-DNA hybridization.

MATERIALS AND METHODS

Strains Used in This Study

The study included a total of 52 strains that were widely distributed, both geographically and by the type of sample from which they were isolated that, included different host species (humans, pigs, cow, deer, clams, etc.) and environments (water, milk, reclaimed water etc.) as show in **Table 1**. Six strains possessed their genomes available at the GenBank database, 36 were field isolates from different sources and countries collected over a broad time frame (1985–2013) and 10 strains were from the BCCM/LMG Bacteria Culture Collection (**Table 1**). Among the latter was the type strain of *A. cryaerophilus* LMG 24291^T that corresponds to subgroup 1A and the reference strain LMG 10829 of the subgroup 1B (**Table 1**). The 46 strains were re-evaluated or ascribed to subgroups 1A or 1B using the 16S rDNA-RFLP method described by Figueras et al. (2008, 2012). The method consists of the digestion of an amplified fragment (1026 bp) of the 16S rRNA gene with the enzyme *Mse*I, which produces a pattern with different band sizes for subgroup 1A (395, 216, 143, 138 bp) and for subgroup 1B (365, 216, 143, and 138 bp). The RFLP patterns of the six genomes from the GenBank database (genomes L397 to L401 and L406) were obtained by an *in silico* simulation of the enzymatic digestion using GeneQuest

TABLE 1 | Strains used ($n = 52$) in this study included field isolates, the type and reference collection strains of the species *A. cryaerophilus* and genomes from the NCBI database[a] and 7 obtained in this study[b] (accession numbers in **Table 2**).

Country	Strain	Source	Isolation year	16S-RFLP Pattern	Cluster
Brazil	F196	Aborted porcine fetus	1997	1B	I
Brazil	UF1T	Uterus, sow	1997	1B	I
Brazil	UF2T	Uterus, sow	1997	1B	I
Brazil	UPER3	Uterus, sow	1997	1B	I
Canada	LMG 10229[b]	Kidney, aborted porcine fetus	1990	1B	I
Canada	LMG 10241	Kidney, aborted porcine fetus	1990	1B	I
Canada	LMG 10210[b]	Aborted bovine fetus	1990	1B	IV
Canada	L397[a]	Wastewater	2008	1B	I
Canada	L398[a]	Water	2008	1B	I
Canada	L399[a]	Wastewater	2008	1B	I
Canada	L400[a]	Wastewater	2008	1B	I
Canada	L401[a]	Goose feces	2009	1B	I
Canada	L406[a]	Water	2008	1B	I
Chile	AB3A	Abomasum, aborted bovine fetus	2011	1B	I
Chile	AB74A	Abomasum, aborted bovine fetus	2013	1B	I
Chile	AO2A	Lungs, aborted ovine fetus	2011	1B	I
Chile	AL 20-1	Clam	2011	1B	II
Chile	CV-152	Feces, deer	2013	1A	III
Chile	CV-2101	Feces, deer	2013	1A	III
Chile	EMU-3	Feces, emu	2013	1A	III
Chile	FE7	Feces, chicken	2005	Abutz	II
Chile	HHS 118A	Feces, asymtomatic human	2013	1B	I
Chile	HHS 133A	Feces, asymtomatic human	2013	1B	I
Chile	HHS 188A	Feces, asymtomatic human	2013	1B	I
Chile	HHS 191A	Feces, asymtomatic human	2013	1B	I
Chile	HHS 205A	Feces, asymtomatic human	2013	1B	I
Chile	MC 2-2	Surf clam	2011	NP	I
Chile	MCV 42-1	Feces, cow	2011	1B	I
Chile	ME 15-4	Mussel	2011	Abutz	II
Chile	NAV 15-1	Razor clam	2011	1A	IV
Chile	NAV12-2	Razor clam	2011	NP	I
Chile	NB14A	Jejunum, calf	2011	1B	I
Costa Rica	14 PHA	Viscera, chicken	2011	1B	I
Costa Rica	20 PHF	Viscera, chicken	2011	1B	I
Ireland	LMG 24291[Tb]	Brain, aborted bovine fetus	1985	1A	III
Ireland	LMG 9065[b]	Placenta, aborted ovine fetus	1989	1A	II
Ireland	LMG 9861[b]	Peritoneum, aborted bovine fetus	1990	1B	I
Ireland	LMG 9863[b]	Placenta, aborted ovine fetus	1990	Abutz	II
Ireland	LMG 29976[b]	Eye, aborted porcine fetus	1990	1A	IV
Ireland	LMG 9871[b]	Kidney, aborted bovine fetus	1990	Abutz	II
Italy	284/1	Cow milk	2012	1B	I
Italy	BUF3	Buffalo milk	2012	1B	I
Italy	FEBU4	Feces, buffalo	2012	1B	I
New Zealand	8749401	Diarrhoeic feces, human	2008	1B	I
New Zealand	8756347	Diarrhoeic feces, human	2008	1B	I
Spain	8122333	Diarrhoeic feces, human	2012	1B	I
Spain	RW15-1	Reclaimed water	2013	1A	IV
Spain	RW17-4	Reclaimed water	2013	1A	IV
Spain	RW25-5	Reclaimed water	2013	1A	I
Spain	RW33-8	Reclaimed water	2013	1A	I
Spain	RW45-3	Reclaimed water	2013	1A	IV
USA	LMG 10829	Human blood	1990	1B	I

software (DNASTAR, USA). When a different pattern from that expected for *A. cryaerophilus* was obtained, it was compared with those patterns described for the type strains of all the *Arcobacter* species by Figueras et al. (2008, 2012). In addition the identity of the strains were confirmed by sequencing the *rpoB* gene using primers and conditions described in other studies (Collado et al., 2009; Levican et al., 2015).

Phylogenetic Analysis

A Multilocus Phylogenetic Analysis (MLPA) was carried out by amplifying and sequencing 4 housekeeping genes (*gyrB, rpoB, atpA*, and *cpn60*) following protocols described by Levican Asenjo (2013). In addition, these genes and the 16S and 23S rRNA genes were extracted from the 7 obtained genomes and from the 6 downloaded from the GenBank database. Accession number or locus tag of each gene and strain are show in Supplementary Table S1. Genes were aligned (Supplementary Figure S4) using CLUSTALW (Larkin et al., 2007) implemented in MEGA 6 software (Tamura et al., 2013). The same software was used for the phylogenetic analysis using Neighbor-Joining (NJ) algorithm (Kimura, 1980; Saitou and Nei, 1987) and the bootstrap support for individual nodes was calculated with 1,000 replicates.

Whole Genome Sequencing and Analysis

The genome sequence of the type strain of *A. cryaerophilus* (LMG 24291^T) and of six additional strains (LMG 10229^T, LMG 9861, LMG 9065^T, LMG 9871, LMG 29976^T, and LMG 10210) representative of the different MLPA clusters were obtained in the present study using Illumina MiSeq platform (San Diego, CA, USA). The genomic DNA was extracted from pure cultures using the Easy-DNA^{TM} gDNA Purification kit (Invitrogen, Madrid, Spain). Genomic libraries were prepared with the Nextera® XT DNA Sample Preparation Kit (Illumina) following manufacturer's instructions. Genome assembly was carried out with the SPAdes 3.9 (Nurk et al., 2013) and the CGE assemblers (Larsen et al., 2012) and the best results were selected for further analysis. Assembled genomes were annotated using Prokka v1.11 software (Seemann, 2014). Additionally, the protein-encoding sequences (CDS) were annotated using the Rapid Annotation Subsystem Technology (RAST) (Aziz et al., 2008) and the PATRIC server v3.5.2. (Wattam et al., 2017). The general characteristics derived from the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) and described for the 13 genomes (6 from the GenBank database and 7 from this study) were: genome size (Mb), number of contigs, N50 (bp), G+C content (%) and the number of predicted CDS. Furthermore, the genomes were compared by the Average Nucleotide Identity (ANI) and the *in silico* DNA-DNA hybridization (*is*DDH) indices using OrthoANI (Lee et al., 2015) and Genome-to-Genome Distance Calculator software (Meier-Kolthoff et al., 2013), respectively.

Additionally, a phylogenetic analysis of the 13 genomes (LMG 24291^T, LMG 10229^T, LMG 9861, L397-L401, L406, LMG 9065^T, LMG 9871, LMG 29976^T, and LMG 10210) was carried out using the Maximum Likelihood estimation using RAxML (Stamatakis, 2014) with the pipeline implemented in the PATRIC server (Wattam et al., 2017). The genome of *A. trophiarum* LMG 25534^{T}was used as outgroup. As a first step, the phylogeny was constructed using a set of homologous proteins identified with BLASTp (Boratyn et al., 2013) and clustered with the Markov Cluster Algorithm (MCL) (Dongen, 2000). The second step was an alignment of the protein set using MUSCLE (Edgar, 2004) and the Hidden Markov Models (HMM) were constructed with HMMER tools (Eddy, 1998).

Virulence and Antibiotic Resistance Genes

Virulence genes were searched by BLASTn analysis with default parameters using the Virulence Factors of Pathogenic Bacteria Database (VFDB) (Chen et al., 2005), Victors Database (University of Michigan, USA) and PATRIC_VF (Wattam et al., 2017). Antibiotic resistance genes were searched using the Antibiotic Resistance Database (ARDB) (Liu and Pop, 2009) and the Comprehensive Antibiotic Resistance Database (CARD) (Jia et al., 2017). The five mentioned databases are included at the Specialty Genes tool available at the PATRIC server (Wattam et al., 2017). Furthermore, the Antibiotic Resistance Gene-Annotation database (ARG-ANNOT) (Gupta et al., 2014) was also used to search antibiotic resistance genes by BLASTp analysis using default parameters and the database ARG-ANNOT AA V3 (March 2017). Virulence and resistance mechanisms were also searched for with RAST (Aziz et al., 2008) and PATRIC servers (Wattam et al., 2017). Additionally, genes related with the virulence of *Arcobacter* (Collado and Figueras, 2011; Douidah et al., 2012; Levican et al., 2013a) were searched for with BLASTn using sequences obtained from GenBank and from the annotated *Arcobacter* genomes of *A. butzleri* RM4018, *A. nitrofigilis* DSM 7299 and *Arcobacter* sp. L. The genes studied were *cadF* and *cj1349*, which encode two fibronectin binding proteins; *ciaB* encodes the invasion protein CiaB, *mviN* gene related to peptidoglycan synthesis; *pldA* gene encodes a phospholipase; *tlyA* gene codifies for a hemolysine; *hecB* related to hemolysis activation; *hecA* gene that encodes an adhesion protein and finally the gene *irgA* that codifies an iron-regulated outer membrane protein (Collado and Figueras, 2011; Douidah et al., 2012; Levican et al., 2013a). The accession number or locus tag of those genes are show in Supplementary Table S2. A phylogenetic analysis was conducted using the three virulence genes (*cj1349*, *mviN*, and *pldA*) present in all the studied genomes to evaluate their genetic relatedness and evolution.

Comparison of the Genome Derived Metabolic and Phenotypic Information

The genomes of the seven representative strains from each cluster (LMG 10229^T, LMG 9861, LMG 9065^T, LMG 9871, LMG 24291^T, LMG 29976^T, and LMG 10210) were compared using the Functional Comparison Tool implemented in the Seed Viewer (Overbeek et al., 2014). This software uses the protein sequences of each compared genome annotated with RAST (Aziz et al., 2008) and reconstructs the metabolic pathways. On the other hand, the phenotypic traits derived from each genome were obtained with Traitar software (Weimann et al., 2016) using the protein annotations obtained with Prokka v1.2 (Seemann, 2014). This software infers phenotypic traits using

data from the Global Infectious Disease and Epidemiology Online Network (GIDEON) and from the Bergey's Systematic Bacteriology (Goodfellow et al., 2012). The software works with a total of 67 traits that embrace different microbiological or biochemical characteristics involved in enzyme activity, growth, oxygen requirements, morphology, and hydrogen sulfide production (Weimann et al., 2016).

Phenotypic Characterization

Phenotypic characterization of the 46 strains included 9 tests recommended in the guidelines for defining new species of the family *Campylobacteraceae* (Ursing et al., 1994; On et al., 2017) and 7 additional tests used in the description of other *Arcobacter* spp. (Donachie et al., 2005; Houf et al., 2005). Most of these tests were chosen using as

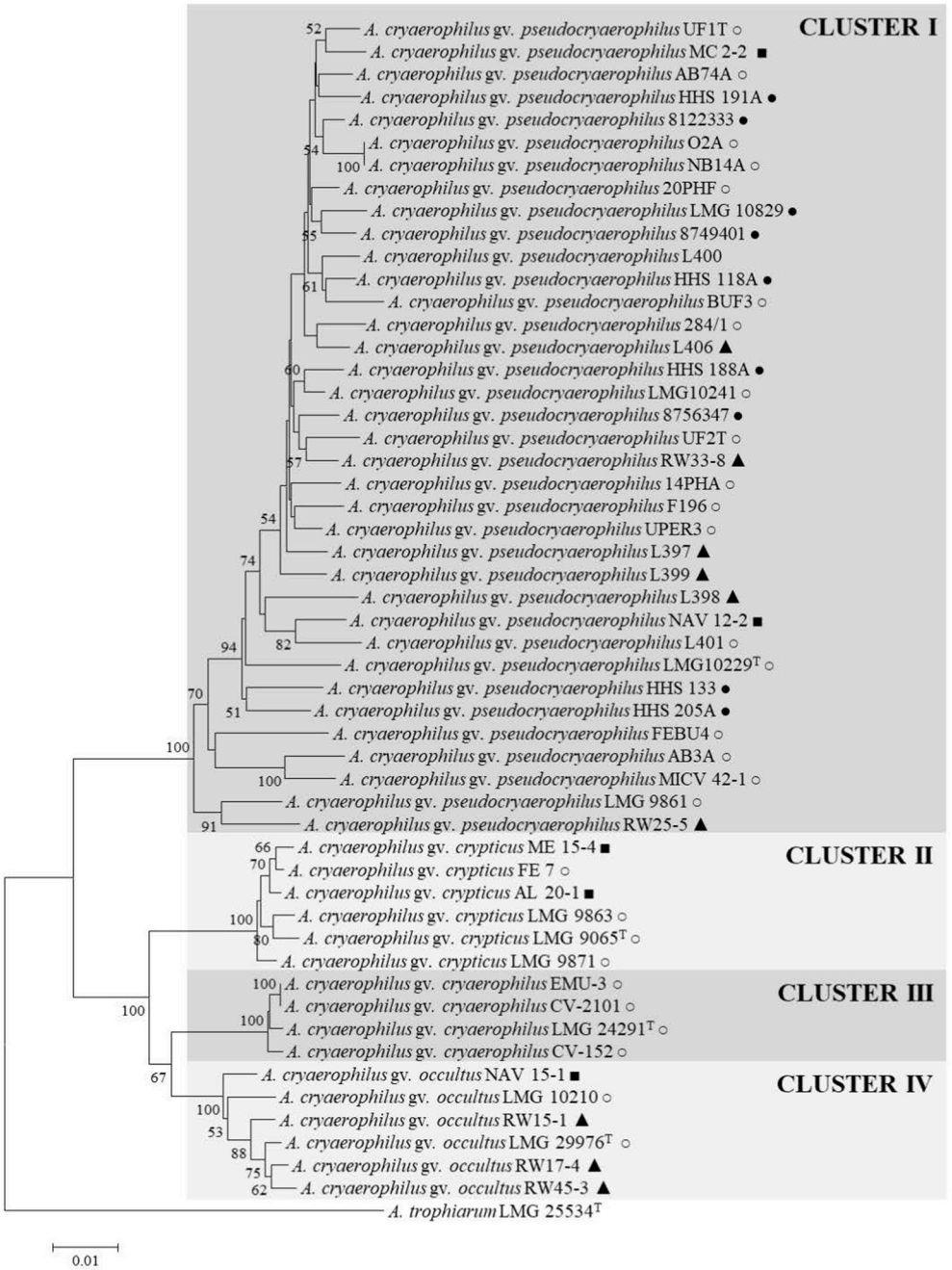


FIGURE 1 | Neighbor joining tree based on the concatenated sequences of *gyrB*, *rpoB*, *atpA,* and *hsp60* (2,408 bp) genes showing the distribution of the 52 strains in four clusters. Bootstrap values (>50%) based on 1,000 replications are shown at the nodes of the tree. Bar indicates 1 substitutions per 100 bp. Isolation source: ●, Human; ○, Animal; ■, Shellfish; ▲, Water.

a criterion the biochemical tests that gave variable results for both *A. cryaerophilus* subgroups in the previous study by On (1996), in which a total of 67 phenotypic tests were analyzed from 9 and 10 strains of subgroups 1A and 1B, respectively. Growth conditions on blood agar were tested (BD Difco, NJ, USA) at 37° and 42°C at three different atmospheres: aerobic, microaerobic ,and anaerobic conditions. The biochemical properties were tested at 30°C in aerobic conditions for the 46 strains using positive and negative controls in parallel for each specific test. To evaluate inter laboratory reproducibility, the strains LMG 9065, LMG 9861, LMG 9871, LMG 10229 and LMG 24291^{T}were tested in parallel in two different laboratories in different countries (Chile and Spain).

RESULTS AND DISCUSSION

Molecular Identification and Phylogeny

Table 1 shows that 46 of the 52 strains gave RFLP patterns defined by Figueras et al. (2008) for *A. cryaerophilus* and 4 showed the one for *A. butzleri* (FE7, ME15-4, LMG 9863, and LMG 9871). However, strains NAV12-2 and MC2-2 produced a new RFLP pattern different to the described ones (Figueras et al., 2008, 2012). From the 46 strains that gave the pattern of *A. cryaerophilus*, 34 gave the pattern of the subgroup 1B (including the in *silico* simulated patterns obtained from the 16S rRNA genes of the 6 GenBank genomes L397- L401 and L406) and 12 the one of the subgroup 1A. This demonstrated once more that subgroup 1B is more abundant than 1A, in agreement

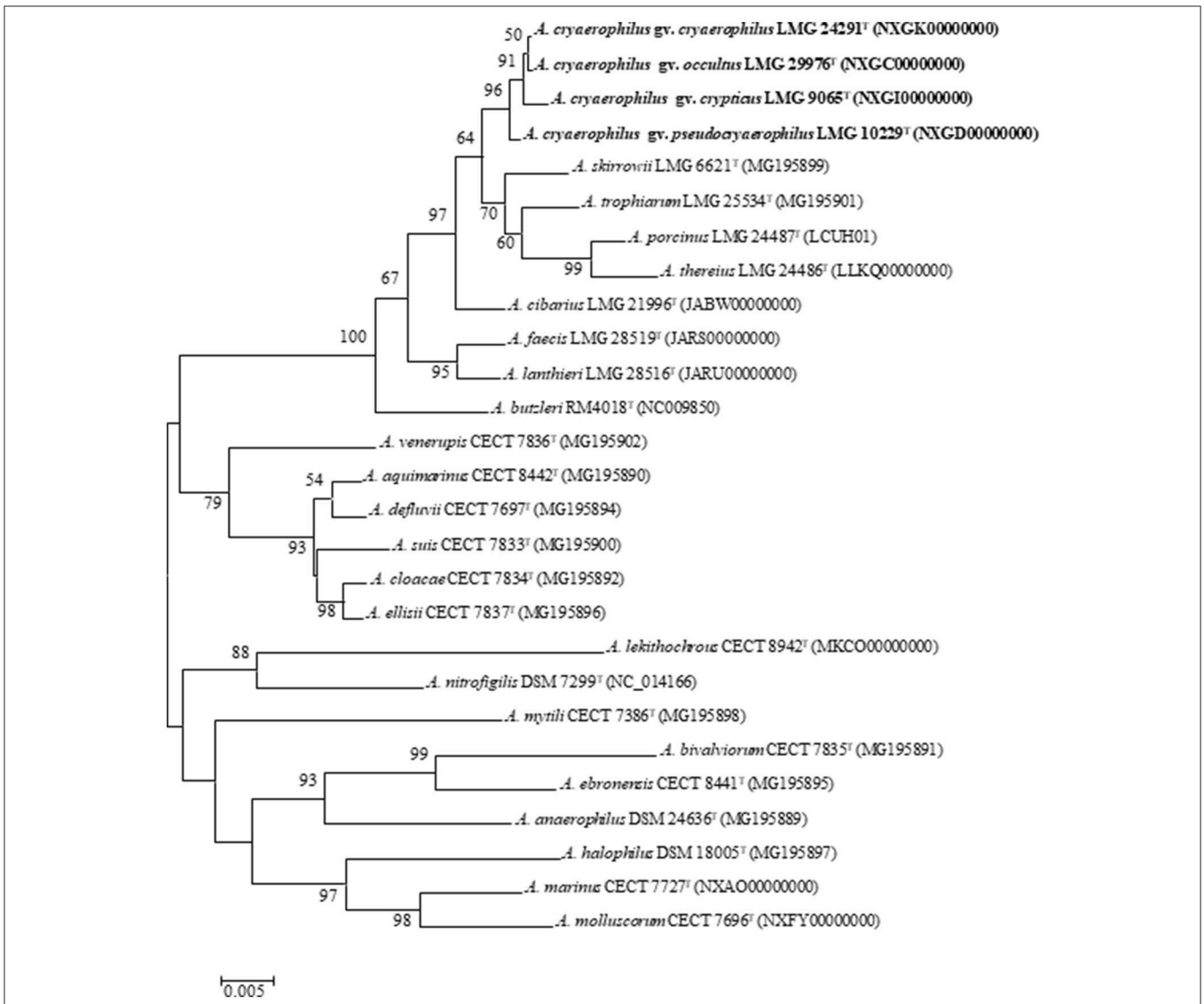


FIGURE 2 | Neighbor joining tree based on 16S rRNA (1,496 bp) sequences showing the phylogenetic position of the representative strains of the four clusters of *A. cryaerophilus* within the genus *Arcobacter*. Bootstrap values (>50%) based on 1,000 replications are shown at the nodes of the tree. Bar, 5 substitutions per 1,000 bp.

with results of previous studies (Debruyne et al., 2010; Collado and Figueras, 2011; Fisher et al., 2014). As Figueras et al. (2012) explained when describing the 16S rDNA-RFLP identification method, different RFLP patterns from those expected for the *Arcobacter* spp. can obtained for new species or might be due to the existence of a mutation on the targeted site of the endonucleases in a known species. The former occurred for instance in *A. mytili* (Collado et al., 2009) and *A. molluscorum* (Figueras et al., 2011a) among other species (Figueras et al., 2011b; Levican et al., 2012, 2013b, 2015). Mutations at the binding site of the endonuclease *MseI* were described in the strains LMG 9863 and LMG 9871 (used in this study, **Table 1**), but in this case instead of resulting in a new pattern they were responsible for generating the pattern for *A. butzleri* instead of *A. cryaerophilus* (Figueras et al., 2012).

The MLPA with the concatenated sequences (2,408 bp) of the four housekeeping genes (*gyrB*, *rpoB*, *atpA,* and *cpn60*) of the 52 strains showed that they grouped into four main clusters (**Figure 1**). Cluster I had 36 strains, most of them (88.8%) from the subgroup 1B, and included the reference strain for the 1B subgroup LMG 10829. The other four strain of this cluster presented the pattern of subgroup 1A ($n = 2$) and a different pattern to those described ($n = 2$). Cluster II ($n = 6$) corresponded to the four strains that showed a 16S rDNA-RFLP pattern similar to the one described for *A. butzleri* (Figueras et al., 2008) and two other strains with the patterns for the subgroups 1A and 1B. Cluster III, included the type strain of *A. cryaerophilus* LMG 24291^T and three field isolates from Chilean animals all belonging to the subgroup 1A, and Cluster IV comprised six strains, mostly from subgroup 1A ($n = 5$). Interestingly, strains recovered from human specimens belonged exclusively to Cluster I, suggesting potential host specificity because strains associated with farm animal abortions were present in the four clusters (**Figure 1**).

A representative type strain was selected from each cluster (I–IV) for further analysis and for constructing a 16S rRNA gene phylogenetic tree (**Figure 2**). The tree showed that the four strains formed separated branches, strains LMG 24291^T and LMG 29976^T being the nearest ones. The percentage of similarity of the 16S rRNA gene between the type strains ranged from 99.5% between strains LMG 10229^T (Cluster I) and LMG 9065^T (Cluster II) to 99.9% between the original type strain of *A. cryaerophilus* LMG 24291^T (Cluster III) and the representative strain of Cluster IV (LMG 29976^T). These results agree with what occurs between other species of *Arcobacter,* such as *A. ellisii* and *A. cloacae* (Figueras et al., 2011b; Levican et al., 2013b), where the 16S rRNA gene does not have enough resolution to differentiate the species. The phylogeny of the 23S rRNA gene (Supplementary Figure S1) and the one carried out with the concatenated sequences of the two rRNA genes (Supplementary Figure S2) presented the same topology shown with the 16S rRNA gene (**Figure 2**) and confirmed that the strains of Cluster III are more closely related to Cluster IV than to the other clusters.

Genome Analysis

The characteristics of the 13 compared genomes (8 representatives of Cluster I, two of clusters II and IV and

TABLE 2 | Characteristics of the 13 genomes from representative strains from each of the clusters.

Features	CLUSTERS												
	I							II		III	IV		
	LMG 10229^T (NXGD01[a])	LMG 9861 (NXGJ01[a])	L397 (LRUQ01[a])	L398 (LRUR01[a])	L399 (LRUS01[a])	L400 (LRUT01[a])	L401 (LRUU01[a])	L406 (LRUV01[a])	LMG9065^T (NXGI01[a])	LMG9871 (NXGH01[a])	LMG 24291^T (NXGK01[a])	LMG 29976^T (NXGK01[a])	LMG 10210 (NXGE01[a])
Deep Coverage	25X	37X	180X	129X	77X	106X	133X	190X	74X	19X	196X	187X	160X
Size (Mb)	2.06	2.02	2.31	2.03	2.10	2.20	2.17	2.02	2.05	2.08	2.05	2.19	2.26
Contigs	27	32	96	71	92	92	85	65	56	180	91	322	70
N50 (Kb)	199	109	56	64	54	54	58	64	138	38	54	241	355
G+C%	27.3	27.6	27.0	27.2	27.4	27.3	27.1	27.4	27.3	28.2	27.2	30.0	27.6
Genes (Total)	2,134	2,074	2,373	2,100	2,214	2,258	2,205	2,102	2,139	2,237	2,141	2,475	2,346
CDS (Coding)	2,071	2,000	2,246	2,002	2,100	2,138	2,117	2,020	2,070	2,136	2,081	2,288	2,255
Genes (RNA)	46	52	33	31	38	37	30	34	54	50	49	59	59
tRNAs	36	40	27	25	30	29	25	29	38	40	40	46	43
ncRNAs	2	3	3	2	3	2	2	2	2	2	3	3	3

[a]*Genome accession number; ncRNAs, non-coding RNAs.*

one of Cluster III) are shown in **Table 2**. The quality of the genome sequences was in general in agreement with the minimal standards established for the use of genome data for taxonomical purposes, that embraces characteristics of the sequencing and assembly of the genomes like the depth of coverage, the value of N50 and the number of contigs (Chun et al., 2018). The

TABLE 3 | Results of Average Nucleotide Identity (ANI) and *in silico* DNA-DNA hibridization (*is*DDH) between representative genomes of the four clusters.

	CLUSTER												
	I								II		III	IV	
	LMG 10229^T	LMG 9861	L397	L398	L399	L400	L401	L406	LMG 9065^T	LMG 9871	LMG 24291^T	LMG 10210	LMG 29976^T
Cluster I													
LMG 10229^T		74.7	74.3	72.3	73.5	73.5	71.9	75.7	50.3	50.4	50.1	63.7	62.0
LMG 9861	**97.2**		69.6	69.6	68.7	70.3	69.7	72.3	50.2	49.6	49.7	63.0	62.1
L397	**97.1**	**96.6**		73.5	75.3	78.4	71.9	82.1	49.3	49.0	49.2	60.1	59.9
L398	**96.9**	**96.6**	**97.1**		78.4	74.9	87.7	75.3	49.7	50.3	49.3	60.1	59.7
L399	**97.1**	**96.6**	**97.3**	**97.7**		75.3	79.0	75.8	49.8	49.5	49.5	60.2	59.2
L400	**97.1**	**96.7**	**97.7**	**97.2**	**97.3**		74.1	83.4	49.1	49.3	49.1	60.3	59.7
L401	**96.9**	**96.6**	**97.1**	**98.6**	**97.7**	**97.1**		74.6	49.9	49.9	50.5	60.7	59.5
L406	**97.3**	**96.9**	**98.1**	**97.3**	**97.4**	**98.2**	**97.2**		20.1	50.0	49.7	61.7	60.5
Cluster II													
LMG 9065^T	**93.1**	**93.1**	**92.9**	**93.1**	**93.1**	**92.9**	**93.1**	**93.0**		81.4	56.1	51.1	51.7
LMG 9871	**93.1**	**93.1**	**92.9**	**93.1**	**93.0**	**92.9**	**93.1**	**93.0**	**98.1**		49.7	51.0	52.2
Cluster III													
LMG 24291^T	**92.9**	**92.8**	**92.6**	**92.6**	**92.7**	**92.5**	**92.9**	**92.9**	**94.3**	**94.4**		52.3	54.4
Cluster IV													
LMG 10210	**95.7**	**95.8**	**95.2**	**95.2**	**95.3**	**95.3**	**95.2**	**95.5**	**93.5**	**93.4**	**93.4**		73.4
LMG 29976^T	**95.2**	**95.5**	**95.1**	**94.9**	**94.9**	**95.1**	**94.9**	**95.1**	**93.5**	**93.5**	**93.9**	**97.1**	

Values in bold in the lower triangle corresponds to ANI and in the upper triangle to isDDH.

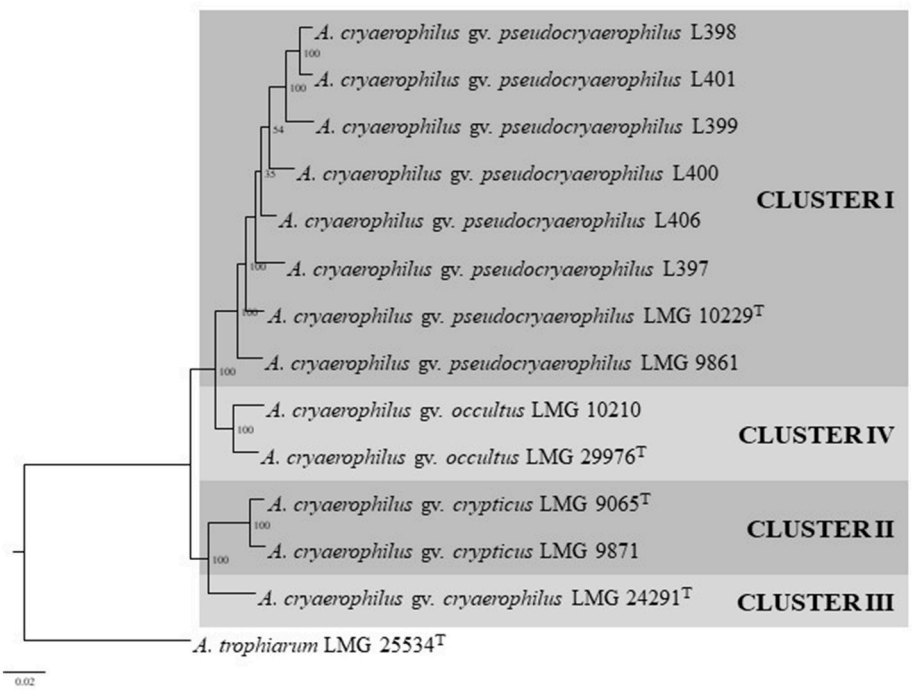


FIGURE 3 | Maximum likelihood tree based on the concatenated sequences of 893 core protein sequences showing the distribution of the 13 genomes used in the same four clusters shown in **Figure 1**. Boostrap values based on 1,000 replications are shown at the nodes of the tree. Bar indicates 2 substitution per 100 aa.

exceptions were the genome sequence data of strains LMG 10229^T, LMG 9871, and LMG 9861 that presented a depth of coverage lower than 50X proposed in the standards (**Table 2**). Globally, the genomic characteristics of the 13 compared genomes shown in **Table 2** were very similar, with sizes that did not differ in more than 0.29 Mb, with a %mol G+C content ranging between 27.0 and 30.0% and with a number of coding sequences or CDS of around 2000 (**Table 2**). The G+C values were in agreement with those (24.6–31%) described in the recent emended description of the genus *Arcobacter* (Sasi Jyothsna et al., 2013). **Table 3** shows the results from the calculated overall genome related taxonomical indices i.e., ANI and *is*DDH. For species delineation the generally accepted ANI and isDDH boundary values are 95–96 and 70%, respectively (Goris et al., 2007; Richter and Rossello-Mora, 2009; Meier-Kolthoff et al., 2013; Chun et al., 2018). However, for the genus *Arcobacter,* ANI values above 96% were the ones that better correlated with *is*DDH results above 70% in previous studies (Figueras et al., 2017; Pérez-Cataluña et al., 2018) in agreement with what happens in other genera (Beaz-Hidalgo et al., 2015; Figueras et al., 2017; Liu et al., 2017). The ANI values of the representative strains from each of the four different clusters were below the 96% cut-off indicating that the compared genomes belonged to different species, while the intra-cluster ANI values ranged from 96.6 to 98.6%. The *is*DDH results of <70% found between strains of the four clusters confirmed as the ANI results did that each cluster represented an independent species. The core genome phylogenetic tree inferred from 893 protein sequences of the

TABLE 4 | Antibiotic resistant genes and virulence factors.

	CLUSTERS												
	I								II		III	IV	
	1	2	3	4	5	6	7	8	9	10	11	12	13
ANTIBIOTIC RESISTANCE													
Multidrug efflux pumps													
CmeABC system[a]	+	+	+	+	+	+	+	+	+	+	+	+	+
MFS Superfamily[a]	−	+	−	+	+	+	+	+	+	−	−	−	−
Macrolids													
MacAB−TolC[a]	+	+	+	+	+	+	+	+	+	+	+	+	+
Quinolones													
gyrA mutation	−	−	−	−	−	−	−	−	−	−	−	−	−
23S rRNA mutations	−	−	−	−	−	−	−	−	−	−	−	−	−
OqxB[b]	+	+	+	+	+	+	+	+	+	+	+	+	+
β-lactamics													
β-lactamase[a]	−	−	−	−	−	−	−	+[d]	−	−	−	−	−
Colistin													
Mcr−1[b]	+	−	+	+	+	+	+	+	+	−	+	−	+
Mcr−2[b]	+	−	+	+	+	+	+	+	+	−	+	−	+
Acriflavin resistance[a]	+	+	+	+	+	+	+	+	+	+	+	+	+
Streptomycin/Spectinomycin[a]	−	−	−	−	−	−	−	−	−	−	−	+	−
VIRULENCE FACTORS													
Invasion													
ciaB[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
mviN[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
Adhesion													
cj1349[c]	+	+	+	+	+	+	+	+	+	+	+	+	+
cadF[c]	−	−	−	−	−	−	−	−	−	−	−	−	−
Filamentous hemmaglutinin													
hecA[c]	−	−	−	−	−	−	−	−	−	−	−	−	−
Hemolysis													
hecB[c]	−	−	−	−	−	−	−	−	−	−	−	−	−
tlyA[c]	−	−	−	−	−	−	−	−	+	−	+	−	−
Outer membarne protein													
irgA[c]	−	−	−	+	−	−	+	−	−	−	−	−	−
Phospholipase													
pldA[c]	+	+	+	+	+	+	+	+	+	+	+	+[e]	+

1, A. cryaerophilus gv. pseudocryaerophilus LMG 10229^T; 2, A. cryaerophilus gv. pseudocryaerophilus LMG 9861; 3, A. cryaerophilus gv. pseudocryaerophilus L397; 4, A. cryaerophilus gv. pseudocryaerophilus L398; 5, A. cryaerophilus gv. pseudocryaerophilus L399; 6, A. cryaerophilus gv. pseudocryaerophilus L400; 7, A. cryaerophilus gv. pseudocryaerophilus L401; 8, A. cryaerophilus gv. pseudocryaerophilus L406; 9, A. cryaerophilus gv. crypticus LMG 9065^T; 10, A. cryaerophilus gv. crypticus LMG 9871; 11, A. cryaerophilus gv. cryaerophilus LMG 24291^T; 12, A. cryaerophilus gv. occultus LMG 29976^T; 13, A. cryaerophilus gv. occultus LMG 10210. [a]RAST/PATRIC results, [b]ARG–ANNOT results, [c]BLASTn of virulence genes results (See Supplementary Table S2), [d]β-lactamase class D, [e]Phospholipase A and C.

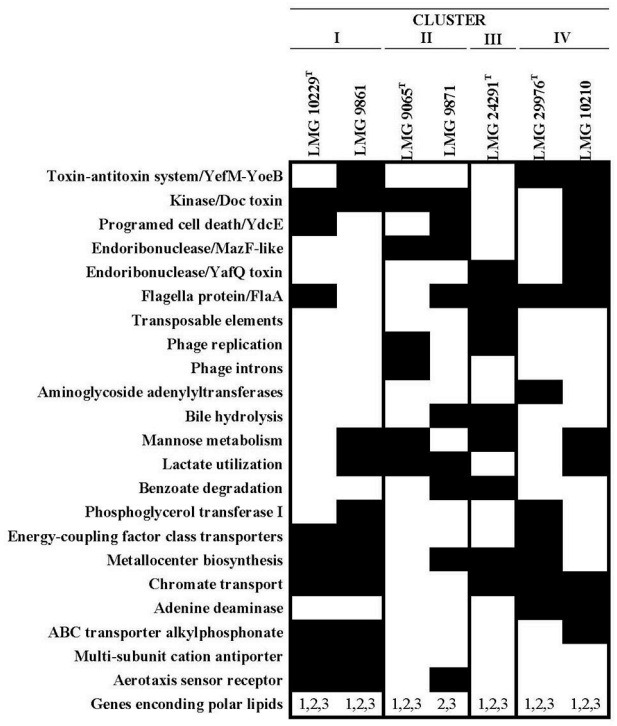


FIGURE 4 | Function based comparison between the representative genomes of each cluster using RAST annotation results. Black squares represent de presence and white squares the absence of each subsystem/protein. 1-3 Polar lipids genes: 1. Phosphatidilglycerol phosphatase A (*pspA*); 2. phosphatidate cytidylyltransferase (*cdsA*); 3. phosphatidyl serine decarboxylase (*psd*). Genes *pspA* and *cdsA* are involved in the synthesis of phosphatidilglycerol (PG) and *psd* gene in the synthesis of phosphatidylethanolamine (PE).

13 genomes obtained with PATRIC showed that the genomes also grouped into four different well-supported clusters with bootstraps of 100% (**Figure 3**). Interestingly, clusters IV and I formed a separate branch from clusters II and III. This indicates that the proteins of the genomes of clusters I and IV are more similar than the nucleotide sequences, and this was also in agreement with the higher values observed with ANI and *is*DDH for these two clusters.

Virulence and Antibiotic Resistance Genes

Of the different methods and databases used for recognizing virulence factors (Victors, VFDB and PATRIC_VF) none of them were useful for recognizing virulence genes. There were only a few exceptions. The phospholipase C identified with the databases PATRIC_VF and Victors in the genome of the strain LMG 29976^T (Cluster IV). The enzyme UDP-N-acetylglucosamine 4-6 dehidratase involved in flagelline glycosylation and identified with the VFDB database in the genomes L397 and L399 (Cluster I). Finally, the Pspa protein (EC 2.3.1.41), essential for gluconeogenesis, identified using PATRIC_VF in the genome LMG 9871 of Cluster II (**Table 4**). However the BLASTn carried out for the detection of virulence genes showed the presence of different genes related with adhesion (*cj1349*), invasion (*ciaB* and *mviN*) and phospholipase activity (*pldA*) (**Table 4**). None of the genomes showed the *cadF*, *hecA* and *hecB* genes that encode a fibronectin binding protein, an adhesion protein and a factor for hemolysis activation, respectively. These results agree with those obtained for the genome of *A. thereius* LMG 24486^T (Rovetto et al., 2017). The *irgA* and *tlyA* genes that encode an iron-regulated outer membrane protein and a hemolysine, respectively, were the only ones found i.e. the gene *irgA* in the genomes L398 and L401 (Cluster I); the gene *tlyA* in LMG 24291^T (Cluster III) and LMG 9065^T (Cluster II). The phylogenetic analysis of the concatenated sequences of the four virulence genes present in all the genomes (*cj1349*, *mviN*, *pldA* and *ciaB*) formed the same four clusters (Supplementary Figure S3). However, the distribution of the clusters was similar to the one obtained with the core genome tree (**Figure 3**), where clusters I and IV formed a separated branch from clusters II and III.

Regarding the presence of antibiotic resistant mechanisms, all the genomes showed the cmeABC multidrug efflux pump, the MacAB-TolC system for macrolide resistance, the *oxqB* gene related with quinolone resistance and genes related with the resistance to acriflavine. Resistance to colistin by the genes *mcr-1* and *mcr-2* were present in 85% of the genomes. The genome L406 was the only one that possessed a β-lactamase gene of class D. Resistance to β-lactamic compounds have been reported in other studies (Atabay and Aydin, 2001; Fera et al., 2003) and the same β-lactamase gene is present in the genome of *A. butleri* RM4018. However, this gene is absent in the genome of *A. thereius* LMG 24486^T (Rovetto et al., 2017). The genome LMG 29976 was the only one that presented genes for the resistance to streptomycin/spectomycin. The susceptibility of *A. cryaerophilus* to streptomycin has been previously demonstrated (Kabeya et al., 2004; Rahimi, 2014). However, this is the first report that show the presence of resistance genes to this antimicrobial compound. Mutations on the 23S rRNA (Ren et al., 2011) and the *gyrA* gene

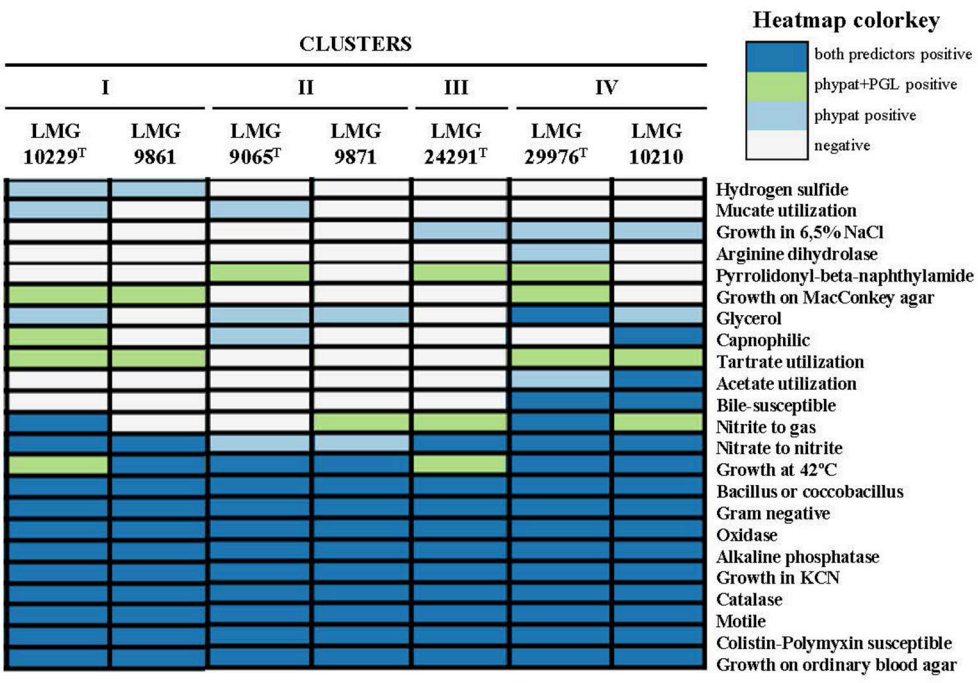


FIGURE 5 | Phenotypic inference using Traitar software for the representative genomes of each cluster. The software uses two prediction models, the phypat model (predicts the presence/absence of proteins found in the phenotype of 234 bacterial species) and a combination of phypat+PGL models (uses the information of phypat combined with the information of the acquisition or loss of protein families and phenotypes through the evolution), to determine the phenotypic characteristics.

(Carattoli et al., 2002) for erythromycin and quinolone resistance were not detected, despite *gyrA* mutations have been found in some quinolone-resistant *A. cryaerophilus* strains (Abdelbaqi et al., 2007; Van den Abeele et al., 2016).

Functional and Phenotypic Inference

Several subsystems where found to be characteristic of each Cluster on the basis of the functional-based comparison between the representative genomes (**Figure 4**). Cluster I genomes (LMG 10229^T and LMG 9861) carry specifically multi-subunit cation antiporters [Na(+) H(+) cation antiporter ABCDEFG] whose function includes sodium tolerance and pH homeostasis in an alkaline environment (Ito et al., 2017). Cluster II genomes (LMG 9065^T and LMG 9871) were the only ones that did not show the chromate transport protein ChrA, which confers resistance to chromate compounds present in the other studied genomes. Cluster III (LMG 24291^T) was the only one that presented transposable elements as the putative transposase TniA and the Nucleotide Triphosphate binding protein TniB. Finally, the enzyme Adenosine deaminase (EC 3.5.4.4) involved in purine metabolism was only detected in Cluster IV genomes (LMG 29976^T and LMG 10210).

From the 67 phenotypic inferred traits analyzed with Traitar 11(16.4%) were found in all the analyzed genomes while 12 were only found in some of them (**Figure 5**). The genomes of Cluster I (LMG 10229^T and LMG 9861) were predicted to produce hydrogen sulfide while those of Cluster IV (LMG 29976^T and LMG 10210) showed acetate utilization and bile susceptibility. However, none of these characteristics have been observed when they have been tested in the laboratory on those strains. This might be due to the inability to reproduce the necessary conditions in the laboratory for the expression of these features. None of the other nine traits recognized in some genomes enabled us to differentiate between the IV Clusters.

Phenotypical Characterization

Table 5 shows the phenotypical results obtained from the strains of each of the four clusters. In agreement with what was found in previous studies where phenotypic test did not differentiate between subgroups 1A and 1B (Neill et al., 1985; Vandamme et al., 1992; On, 1996), none of the performed phenotypic tests enabled to clearly distinguish strains from each of the four phylogenetic clusters. Most of the tests gave variable results except for Cluster IV. However, this might be due to the small number of strains ($n = 2$) analyzed in this group. Considering these results, each of the three genetically recognized new species (clusters I, II, and IV) should be considered a different genomovar (gv.) of the species *A. cryaerophilus*. A genomovar is a well-delimited group of strains that correspond to a new species by genomic information but that cannot be phenotypically differentiated (Ursing et al., 1995). Cluster III represents the original species *A. cryaerophilus* because it embraces the type strain of the species. The value of the phenotypic characterization has already been questioned considering the lack of reproducibility of results between laboratories and some authors have suggested it is now time to base the description of new taxa on the genome sequence analysis (Moore et al., 2010; Sutcliffe, 2015). According to Sutcliffe (2015), phenotypic characterization is harder to evaluate nowadays than the genotype. Considering that genomic characterization is objective and reproducible, we agree with Sutcliffe (2015) that we should be able to define species on the basis of genetic characters like the ones evaluated in this study. This will favor the faster discover of the large number of taxa waiting to be described (Sutcliffe, 2015). However, this will require a modification of the Bacteriological Code, which we hope will happen in the near future.

TABLE 5 | Phenotypic characteristics of the four clusters.

Characteristics	1	2	3	4
Growth in/on				
Air at 37°C	+	V(+)	+	−
Microaerobiosis at 37°C	+	V(+)	V(+)	−
Anaerobiosis at 37°C	V(+)	V(−)	V(−)	−
Air at 42°C	V(−)	−	−	−
2% (w/v) NaCl	V(+)	+	V(+)	+
3% NaCl	V(−)	−	−	−
4% (w/v) NaCl	−	−	−	−
1% bilis	+	V(+)	V(+)	+
1.5% bilis	+	+	V(+)	+
2% bilis	+	+	V(+)	+
1% (w/v) glycine	V(−)	−	−	−
0.1% sodium deoxycholate	V(+)	V(+)	V(+)	+
MacConkey	V(+)	+	V(+)	+
$CdCl_2$	V(−)	V(−)	V(−)	V
Resistance to				
Cefoperazone (64 mg/L)	V(+)	+	+	+
Enzyme activity				
Catalase	+	+	+	+

Taxa: 1, A. cryaerophilus gv. pseudocryaerophilus (n = 27) [Cluster I]; 2, A. cryaerophilus gv. crypticus (n = 5) [Cluster II]; 3, A. cryaerophilus gv. cryaerophilus(n = 3) [Cluster III]; 4, A. cryaerophilus gv. occultus (n = 2) [Cluster IV]. The specific responses for type strains were coincidental or expressed in brackets. Unless otherwise indicated: +, ≥95% strains positive; −, ≤11% strains positive; V, variable; (), main result of the strains; CO_2 indicates microaerobic conditions.

CONCLUSION

The phylogenetic and genomic analysis showed that the strains of the species *A. cryaerophilus* represent four separated species. In addition, phenotypical and functional traits were in evidence for the genomes selected as representative of each cluster. Despite all the results, phenotypic characterization carried out at the laboratory showed a high inter- and intra-cluster variability that did not allow us to determine specific phenotypic characteristics or therefore to define the three uncovered clusters as three new species. Following current bacterial taxonomic rules, we will not be able to define these species until we find phenotypical characteristics that allow us to discriminate the three new species from each other

and from the species *A. cryaerophilus*. Therefore, we describe them as four genomovars with the names "*A. cryaerophilus gv. pseudocryaerophilus*" (pseu.do.cry.a.e.ro'phi.lus. Gr. adj. *pseudês* false, N.L. masc. adj. *cryaerophilus* specific epithet of an *Arcobacter* species; N.L. masc. adj. *pseudocryaerophilus* false *cryaerophilus;* Cluster I = LMG 10229^T), "*A. cryaerophilus gv. crypticus*" (cryp'ti.cus. L. masc. adj. *Crypticus* hidden; Cluster II = LMG 9065^T), *A. cryaerophilus gv. cryaerophilus* (Cluster III = LMG 24291^T) and "*A. cryaerophilus gv. occultus*" (oc.cul′tus. L. adj. *occultus* occulted, hidden; Cluster IV = LMG 29976^T). Unfortunately, the phenotype derived from the genome could not be reproduced in the laboratory, either. This might be due to the inability to mimic *in vitro* the conditions for the expression of these pathways or characteristics. The phenotypic characterization limits a proper description and it might be considered an important shortcoming in the genomic era in which all the molecular and genomic data leave no doubts about the existence of four different species among the investigated *A. cryaerophilus* strains.

AUTHORS CONTRIBUTIONS

LC and MF: designed the work; LC and OS: carried out the phylogenetic analysis; VL and AP-C: carried out the phenotypic characterization of the strains; AP-C: carried out the genome sequencing and analysis; LC, MF, and AP-C: wrote the paper.

ACKNOWLEDGMENTS

The authors thank Dr. Maria Laura Arias (University of Costa Rica), Dr. Mary Nulsen (Massey University), Dr. Andrea Serraino (University of Bologna) and Dr. Sergio Oliveira (ULBRA University of Brazil) for kindly providing *Arcobacter* strains. We thank Prof. Aharon Oren from the Hebrew University of Jerusalem for supervising and correcting the species name etymology. This work was supported in part by the project DID-UACh S-2013-06 from the Universidad Austral de Chile and by the projects JPIW2013-69 095-C03-03 of MINECO (Spain) and AQUAVALENS of the Seventh Framework Program (FP7/2007-2013) grant agreement 311846 from the European Union. AP-C thanks *Institut d'Investigació Sanitária Pere Virgili* (IISPV) for her PhD fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00805/full#supplementary-material

REFERENCES

Abdelbaqi, K., Ménard, A., Prouzet-Mauleon, V., Bringaud, F., Lehours, P., and Mégraud, F. (2007). Nucleotide sequence of the *gyrA* gene of *Arcobacter* species and characterization of human ciprofloxacin-resistant clinical isolates. *FEMS Immunol. Med. Microbiol.* 49, 337–345. doi: 10.1111/j.1574-695X.2006.00208.x

Atabay, H. I., and Aydin, F. (2001). Susceptibility of *Arcobacter butzleri* isolates to 23 antimicrobial agents. *Lett. Appl. Microbiol.* 33, 430–433. doi: 10.1046/j.1472-765X.2001.01025.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

Beaz-Hidalgo, R., Hossain, M. J., Liles, M. R., and Figueras, M.-J. (2015). Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for *Aeromonas* genomes in the GenBank database. *PLoS ONE* 10:e0115813. doi: 10.1371/journal.pone.0115813

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., et al. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 41, W29–W33. doi: 10.1093/nar/gkt282

Carattoli, A., Dionisi, A., and Luzzi, I. (2002). Use of a LightCycler *gyrA* mutation assay for identification of ciprofloxacin-resistant *Campylobacter coli*. *FEMS Microbiol. Lett.* 214, 87–93. doi: 10.1128/JCM.39.4.1443-1448.2001

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., et al. (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. doi: 10.1093/nar/gki008

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Collado, L., Cleenwerck, I., Van Trappen, S., De Vos, P., and Figueras, M. J. (2009). *Arcobacter mytili* sp. nov., an indoxyl acetate-hydrolysis-negative bacterium isolated from mussels. *Int. J. Syst. Evol. Microbiol.* 59, 1391–1396. doi: 10.1099/ijs.0.003749-0

Collado, L., and Figueras, M. J. (2011). Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. *Clin. Microbiol. Rev.* 24, 174–192. doi: 10.1128/CMR.00034-10

Collado, L., Inza, I., Guarro, J., and Figueras, M. J. (2008). Presence of *Arcobacter* spp. in environmental waters correlates with high levels of fecal pollution. *Environ. Microbiol.* 10, 1635–1640. doi: 10.1111/j.1462-2920.2007.01555.x

Debruyne, L., Houf, K., Douidah, L., De Smet, S., and Vandamme, P. (2010). Reassessment of the taxonomy of *Arcobacter cryaerophilus*. *Syst. Appl. Microbiol.* 33, 7–14. doi: 10.1016/j.syapm.2009.10.001

Diéguez, A. L., Balboa, S., Magnesen, T., and Romalde, J. L. (2017). *Arcobacter lekithochrous* sp. nov., isolated from a molluscan hatchery. *Int. J. Syst. Evol. Microbiol.* 67, 1327–1332. doi: 10.1099/ijsem.0.001809

Donachie, S. P., Bowman, J. P., On, S. L. W., and Alam, M. (2005). *Arcobacter halophilus* sp. nov., the first obligate halophile in the genus Arcobacter. *Int. J. Syst. Evol. Microbiol.* 55, 1271–1277. doi: 10.1099/ijs.0.63581-0

Dongen, S. M. van. (2000). *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrech, Utrech.

Douidah, L., De Zutter, L., Bar,é, J., De Vos, P., Vandamme, P., Vandenberg, O., et al. (2012). Occurrence of putative virulence genes in *Arcobacter* species isolated from humans and animals. *J. Clin. Microbiol.* 50, 735–741. doi: 10.1128/JCM.05872-11

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755–763. doi: 10.1093/bioinformatics/14.9.755

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113

Fera, M., Carbone, M., Maugeri, T. L., Giannone, M., Gugliandolo, C., La Camera, E., et al. (2003). *In vitro* susceptibility of *Arcobacter butzleri* and *Arcobacter cryaerophilus* to different antimicrobial agents. *Int. J. Antimicrob. Agents* 21, 488–491. doi: 10.1016/S0924-8579(03)00004-9

Ferreira, S., Queiroz, J. A., Oleastro, M., and Domingues, F. C. (2015). Insights in the pathogenesis and resistance of *Arcobacter* : a review. *Crit. Rev. Microbiol.* 42, 364–383. doi: 10.3109/1040841X.2014.954523

Figueras, M. J., Collado, L., and Guarro, J. (2008). A new 16S rDNA-RFLP method for the discrimination of the accepted species of *Arcobacter*. *Diagn. Microbiol. Infect. Dis.* 62, 11–15. doi: 10.1016/j.diagmicrobio.2007.09.019

Figueras, M. J., Collado, L., Levican, A., Perez, J., Solsona, M. J., and Yustes, C. (2011a). *Arcobacter molluscorum* sp. nov., a new species isolated from shellfish. Syst. *Appl. Microbiol.* 34, 105–109. doi: 10.1016/j.syapm.2010.10.001

Figueras, M. J., Levican, A., and Collado, L. (2012). Updated 16S rRNA-RFLP method for the identification of all currently characterised *Arcobacter* spp. *BMC Microbiol.* 12:292. doi: 10.1186/1471-2180-12-292

Figueras, M. J., Levican, A., Collado, L., Inza, M. I., and Yustes, C. (2011b). *Arcobacter ellisii* sp. nov., isolated from mussels. *Syst. Appl. Microbiol.* 34, 414–418. doi: 10.1016/j.syapm.2011.04.004

Figueras, M. J., Levican, A., Pujol, I., Ballester, F., Rabada Quilez, M. J., and Gomez-Bertomeu, F. (2014). A severe case of persistent diarrhoea associated with *Arcobacter cryaerophilus* but attributed to *Campylobacter* sp. and a review of the clinical incidence of Arcobacter spp. *New microbes new Infect.* 2, 31–37. doi: 10.1002/2052-2975.35

Figueras, M. J., Pérez-Cataluña, A., Salas-Mass,ó, N., Levican, A., and Collado, L. (2017). "*Arcobacter porcinus*" sp. nov., a novel Arcobacter species uncovered by Arcobacter thereius. *New Microbes New Infect.* 15, 104–106. doi: 10.1016/j.nmni.2016.11.014

Fisher, J. C., Levican, A., Figueras, M. J., and McLellan, S. L. (2014). Population dynamics and ecology of *Arcobacter* in sewage. *Front. Microbiol.* 5:525. doi: 10.3389/fmicb.2014.00525

Goodfellow, M., Kämpfer, P., Busse, H. J., Trujillo, M. E., Suzuki, K. I., Ludwig, W., et al. (2012). *Bergey's Manual of Systematic Bacteriology*. New York, NY: Springer.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., et al. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* 58, 212–220. doi: 10.1128/AAC.01310-13

Houf, K., On, S. L. W., Coenye, T., Mast, J., Van Hoof, J., and Vandamme, P. (2005). *Arcobacter cibarius* sp. nov., isolated from broiler carcasses. *Int. J. Syst. Evol. Microbiol.* 55, 713–717. doi: 10.1099/ijs.0.63103-0

Ito, M., Morino, M., and Krulwich, T. A. (2017). Mrp antiporters have important roles in diverse bacteria and archaea. *Front. Microbiol.* 8:2325. doi: 10.3389/FMICB.2017.02325

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Kabeya, H., Maruyama, S., Morita, Y., Ohsuga, T., Ozawa, S., Kobayashi, Y., et al. (2004). Prevalence of *Arcobacter* species in retail meats and antimicrobial susceptibility of the isolates in Japan. *Int. J. Food Microbiol.* 90, 303–308. doi: 10.1016/S0168-1605(03)00322-2

Kiehlbauch, J. A., Plikaytis, B. D., Swaminathan, B., Cameron, D. N., and Wachsmuth, I. K. (1991). Restriction fragment length polymorphisms in the ribosomal genes for species identification and subtyping of aerotolerant *Campylobacter* species. *J. Clin. Microbiol.* 29, 1670–1676.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi: 10.1128/JCM.06094-11

Lee, I., Kim, Y. O., Park, S.-C., and Chun, J. (2015). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103. doi: 10.1099/ijsem.0.000760

Levican, A., Alkeskas, A., Günter, C., Forsythe, S. J., and Figueras, M. J. (2013a). Adherence to and invasion of human intestinal cells by *Arcobacter* species and their virulence genotypes. *Appl. Environ. Microbiol.* 79, 4951–4957. doi: 10.1128/AEM.01073-13

Levican, A., Collado, L., Aguilar, C., Yustes, C., Diéguez, A. L., Romalde, J. L., et al. (2012). *Arcobacter bivalviorum* sp. nov. and *Arcobacter venerupis* sp. nov., new species isolated from shellfish. *Syst. Appl. Microbiol.* 35, 133–138. doi: 10.1016/j.syapm.2012.01.002

Levican, A., Collado, L., and Figueras, M. J. (2013b). *Arcobacter cloacae* sp. nov. and *Arcobacter suis* sp. nov., two new species isolated from food and sewage. *Syst. Appl. Microbiol.* 36, 22–27. doi: 10.1016/j.syapm.2012.11.003

Levican, A., Rubio-Arcos, S., Martinez-Murcia, A., Collado, L., and Figueras, M. J. (2015). *Arcobacter ebronensis* sp. nov. and *Arcobacter aquimarinus* sp. nov., two new species isolated from marine environment. *Syst. Appl. Microbiol.* 38, 30–35. doi: 10.1016/j.syapm.2014.10.011

Levican Asenjo, A. (2013). *Sanitary Importance of Arcobacter*. PhD thesis, Universitat Rovira I Virgili, Reus.

Liu, B., and Pop, M. (2009). ARDB–Antibiotic Resistance Genes Database. *Nucleic Acids Res.* 37, D443–D447. doi: 10.1093/nar/gkn656

Liu, Y., Lai, Q., Du, J., and Shao, Z. (2017). Genetic diversity and population structure of the *Bacillus cereus* group bacteria from diverse marine environments. *Sci. Rep.* 7:689. doi: 10.1038/s41598-017-00817-1

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Moore, E. R. B., Mihaylova, S. A., Vandamme, P., Krichevsky, M. I., and Dijkshoorn, L. (2010). Microbial systematics and taxonomy: relevance for a microbial commons. *Res. Microbiol.* 161, 430–438. doi: 10.1016/j.resmic.2010.05.007

Neill, S. D., Campbell, J. N., O'Brien, J. J., Weatherup, S. T. C., and Ellis, W. A. (1985). Taxonomic position of *Campylobacter cryaerophila* sp. *nov. Int. J. Syst. Bacteriol.* 35, 342–356. doi: 10.1099/00207713-35-3-342

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20, 714–737. doi: 10.1089/cmb.2013.0084

On, S. L. (1996). Identification methods for campylobacters, helicobacters, and related organisms. *Clin. Microbiol. Rev.* 9, 405–422.

On, S. L. W., Harrington, C. S., and Atabay, H. I. (2003). Differentiation of *Arcobacter* species by numerical analysis of AFLP profiles and description of a novel *Arcobacter* from pig abortions and turkey faeces. *J. Appl. Microbiol.* 95, 1096–1105. doi: 10.1046/j.1365-2672.2003.02100.x

On, S. L. W., Miller, W. G., Houf, K., Fox, J. G., and Vandamme, P. (2017). Minimal standards for describing new species belonging to the families *Campylobacteraceae* and *Helicobacteraceae*: campylobacter, Arcobacter, Helicobacter and Wolinella spp. *Int. J. Syst. Evol. Microbiol.* 67, 5296–5311. doi: 10.1099/ijsem.0.002255

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, 206–214. doi: 10.1093/nar/gkt1226

Park, S., Jung, Y. T., Kim, S., and Yoon, J. H. (2016). *Arcobacter acticola* sp. nov isolated from seawater on the East Sea in South Korea. *J. Microbiol.* 54, 655–659. doi: 10.1007/s12275-016-6268-4

Pérez-Cataluña, A., Salas-Mass,ó, N., and Figueras, M. J. (2018). *Arcobacter canalis* sp. nov., isolated from a water canal contaminated with urban sewage. *Int. J. Syst. Evol. Microbiol.* 68, 1258–1264 doi: 10.1099/ijsem.0.002662

Rahimi, E. (2014). Prevalence and antimicrobial resistance of *Arcobacter* species isolated from poultry meat in Iran. *Br. Poult. Sci.* 55, 174–180. doi: 10.1080/00071668.2013.878783

Ren, G. W. N., Wang, Y., Shen, Z., Chen, X., Shen, J., and Wu, C. (2011). Rapid detection of point mutations in domain V of the 23S rRNA gene in erythromycin-resistant *Campylobacter* isolates by pyrosequencing. *Foodborne Pathog. Dis.* 8, 375–379. doi: 10.1089/fpd.2010.0676

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rovetto, F., Carlier, A., Van den Abeele, A.-M., Illeghems, K., Van Nieuwerburgh, F., Cocolin, L., et al. (2017). Characterization of the emerging zoonotic pathogen *Arcobacter thereius* by whole genome sequencing and comparative genomics. *PLoS ONE* 12:0180493. doi: 10.1371/journal.pone.0180493

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Sasi Jyothsna, T. S., Rahul, K., Ramaprasad, E. V. V., Sasikala, C., and Ramana, C. V. (2013). *Arcobacter anaerophilus* sp. nov., isolated from an estuarine sediment and emended description of the genus Arcobacter. *Int. J. Syst. Evol. Microbiol.* 63, 4619–4625. doi: 10.1099/ijs.0.054155-0

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

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

Sutcliffe, I. C. (2015). Challenging the anthropocentric emphasis on phenotypic testing in prokaryotic species descriptions: rip it up and start again. *Front. Genet.* 6:218. doi: 10.3389/fgene.2015.00218

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

Tanaka, R., Cleenwerck, I., Mizutani, Y., Iehata, S., Bossier, P., and Vandamme, P. (2017). *Arcobacter haliotis* sp. nov., isolated from abalone species *Haliotis gigantea*. *Int. J. Syst. Evol. Microbiol.* 67, 3050–3056. doi: 10.1099/ijsem.0.002080

Ursing, J. B., Lior, H., and Owen, R. J. (1994). Proposal of minimal standards for describing new species of the family *Campylobacteraceae*. *Int. J. Syst. Bacteriol.* 44, 842–845. doi: 10.1099/00207713-44-4-842

Ursing, J. B., Rosselló-Mora, R. A., García-Valdes, E., and Lalucat, J. (1995). Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups. *Int. J. Syst. Bacteriol.* 45:604. doi: 10.1099/00207713-45-3-604

Vandamme, P., Falsen, E., Rossau, R., Hoste, B., Segers, P., Tytgat, R., et al. (1991). Revision of Campylobacter, Helicobacter, and Wolinella taxonomy: emendation of generic descriptions and proposal of *Arcobacter* gen. *nov. Int. J. Syst. Bacteriol.* 41, 88–103. doi: 10.1099/00207713-41-1-88

Vandamme, P., Vancanneyt, M., Pot, B., Mels, L., Hoste, B., Dewettinck, D., et al. (1992). Polyphasic taxonomic study of the emended genus *Arcobacter* with *Arcobacter butzleri* comb. nov. and Arcobacter skirrowii sp. nov. an aerotolerant bacterium isolated from veterinary specimens. *Int. J. Syst. Bacteriol.* 42, 344–356. doi: 10.1099/00207713-42-3-344

Van den Abeele, A. M., Vogelaers, D., Vanlaere, E., and Houf, K. (2016). Antimicrobial susceptibility testing of *Arcobacter butzleri* and *Arcobacter cryaerophilus* strains isolated from Belgian patients. *J. Antimicrob. Chemother.* 71, 1241–1244. doi: 10.1093/jac/dkv483

Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., et al. (2017). Comparative genomic analysis of the class *Epsilonproteobacteria* and proposed reclassification to Epsilonbacteraeota (phyl. nov.). *Front. Microbiol.* 8:682. doi: 10.3389/fmicb.2017.00682

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* 45, D535–D542. doi: 10.1093/nar/gkw1017

Weimann, A., Mooren, K., Frank, J., Pope, P. B., Bremges, A., and McHardy, A. C. (2016). From genomes to phenotypes: traitar, the microbial trait analyzer. *mSystems* 1, e00101–e00116. doi: 10.1128/mSystems.00101-16

Whiteduck-Léveillée, K., Whiteduck-Léveillée, J., Cloutier, M., Tambong, J. T., Xu, R., Topp, E., et al. (2016). Identification, characterization and description of *Arcobacter faecis* sp. nov. isolated from a human waste septic tank. *Syst. Appl. Microbiol.* 39, 93–99. doi: 10.1016/j.syapm.2015.12.002

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 Pérez-Cataluña, Collado, Salgado, Lefiñanco and Figueras. 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: 02 November 2017
doi: 10.3389/fmicb.2017.02155

Culture Strategies for Isolation of Fastidious *Leptospira* Serovar Hardjo and Molecular Differentiation of Genotypes Hardjobovis and Hardjoprajitno

*Roberta T. Chideroli[1], Daniela D. Gonçalves[2], Suelen A. Suphoronski[1], Alice F. Alfieri[3], Amauri A. Alfieri[3], Admilton G. de Oliveira[4], Julio C. de Freitas[1] and Ulisses de Padua Pereira[1]**

[1] Department of Veterinary Preventive Medicine, State University of Londrina, Londrina, Brazil, [2] Department of Veterinary Preventive Medicine and Public Health, Paranaense University, Umuarama, Brazil, [3] National Institute of Science and Technology for Dairy Production Chain (INCT - LEITE), Universidade Estadual de Londrina, Londrina, Brazil, [4] Laboratory of Microbial Biotechnology, Department of Microbiology, State University of Londrina, Londrina, Brazil

OPEN ACCESS

Edited by:
Jesus L. Romalde, Universidade de Santiago de Compostela, Spain

Reviewed by:
Suresh Kumar, Universiti Putra Malaysia, Malaysia
Paula Ruybal, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

***Correspondence:**
Ulisses de Padua Pereira
upaduapereira@uel.br

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

Received: *23 August 2017*
Accepted: *20 October 2017*
Published: *02 November 2017*

Citation:
Chideroli RT, Gonçalves DD, Suphoronski SA, Alfieri AF, Alfieri AA, de Oliveira AG, de Freitas JC and Pereira UP (2017) Culture Strategies for Isolation of Fastidious Leptospira Serovar Hardjo and Molecular Differentiation of Genotypes Hardjobovis and Hardjoprajitno. Front. Microbiol. 8:2155. doi: 10.3389/fmicb.2017.02155

The *Leptospira* serovar Hedjo belongs to the serogroup sejroe and this serovar is the most prevalent in bovine herds worldwide. The sejroe serogroup is the most frequently detected by serology in Brazilian cattle herds suggesting that it is due serovar Hardjo. In the molecular classification, this serovar has two genotypes: Hardjobovis and Hardjoprajitno. This serovar is as considered as fastidious pathogens, and their isolation is one of the bottlenecks in leptospirosis laboratories. In addition, its molecular characterization using genomic approaches is oftentimes not simple and time-consuming. This study describes a method for isolating the two genotypes of serovar Hardjo using culture medium formulations and suggests a get-at-able molecular characterization. Ten cows naturally infected which were seropositive were selected from small dairy farms, and their urine was collected for bacterial isolation. We evaluated three modifications of liquid *Leptospira* medium culture supplemented with sodium pyruvate, superoxide dismutase enzyme and fetal bovine serum, and the isolates were characterized by molecular techniques. After isolation and adaptation in standard culture medium, the strains were subcultured for 1 week in the three modified culture media for morphologic evaluation using electronic microscopy. Strains were molecularly identified by multilocus variable-number tandem-repeat analysis (MLVA), partial sequencing and phylogenic analyses of gene *sec Y*. Combining the liquid culture medium formulations allowed growth of the *Leptospira* serovar Hardjo in three tubes. Two isolates were identified as genotype Hardjobovis, and the other as genotype Hardjoprajitno. Morphologically, compared with control media, cells in the medium supplemented with the superoxide dismutase enzyme were more elongated and showed many cells in division. The cells in the medium supplemented with fetal bovine serum were fewer and lost their spirochete morphology. This indicated that the additional supplementation with fetal bovine serum assisted in the initial growth and maintenance of the viable leptospires and the superoxide dismutase enzyme allowed

them to adapt to the medium. These culture strategies allowed for the isolation and convenient molecular characterization of two genotypes of serovar Hardjo, creating new insight into the seroepidemiology of leptospirosis and its specific genotypes. It also provides new information for the immunoprophylaxis of bovine leptospirosis.

Keywords: leptospirosis, culture medium, pyruvate sodium, superoxide dismutase, fetal bovine serum, DNA fingerprint

INTRODUCTION

The genus *Leptospira* comprises a heterogeneous group of pathogenic and saprophytic species belonging to the order *Spirochaetales* (Adler and de la Peña Moctezuma, 2010). Leptospiral serovar diversity results from structural heterogeneity in the carbohydrate component of the lipopolysaccharides (de la Peña-Moctezuma et al., 1999). Many serovars are adapted for specific mammalian hosts, which harbor these microorganisms in the renal tubules and intermittently eliminate them through the urine contaminating the surrounding environment (Adler and de la Peña Moctezuma, 2010).

Serovar Hardjo is one of serovars of sejroe serogroup. In bovine herds naturally infected, the serovar Hardjo is the most prevalent (Ellis, 2015). In Brazilian cattle herds, antibodies against the sejroe serogroup are the most frequently detected by the microscopic agglutination test (MAT) (Favero et al., 2001; Figueiredo et al., 2009; Hashimoto et al., 2012; Silva et al., 2012), suggesting that it is due to the serovar Hardjo. By molecular classification, this serovar has two genotypes (Hardjobovis and Hardjoprajitno). The genotype Hardjobovis belongs to the species *Leptospira borgpetersenii* and genotype Hardjoprajitno to the species *Leptospira interrogans*.

Serological methods are limited in that they can only distinguish the serovars at the serogroup level but cannot differentiate the genotypes of the Hardjo serovar (Picardeau, 2013), which are relevant for the epidemiology of these genotypes. Serovar determination is a very laborious methodology and the use of monoclonal antibody panels for the cross-agglutinin absorption test (CAAT), has a high cost for the implementation of this methodology and may only be performed at the Royal Tropical Institute reference laboratory in the Netherlands (Faine et al., 1999). Isolating leptospiral strains is useful for molecular characterization and genotyping; however, it is time-consuming and uncertain, particularly for the more fastidious serovars such as Hardjo (Pailhoriès et al., 2015). These microorganisms are slow-growing and require a rich medium at a neutral pH, which makes it difficult to cultivate leptospires from natural sources (Johnson and Gary, 1962; Staneck et al., 1973; Bey and Johnson, 1978; Adler et al., 1986; González et al., 2006; Zacarias et al., 2008).

Serovar Hardjo is difficult to culture and with low rates of success in the attempts of isolation due to its extreme nutritional requirements (Robertson et al., 1964; Flint and Liardet, 1980; Ellis and Thiermann, 1986; Leonard et al., 1992). While ordinary culture media are adequate to recover the less fastidious leptospires, they are ineffective for isolating serovar Hardjo.

Many studies are working to improve the culture media used for isolating *Leptospira* spp. by adding components that help the bacteria grow such as sodium pyruvate, different concentrations of Tween 80, bovine serum albumin (Johnson et al., 1973; Rodríguez et al., 2002; González et al., 2006; Wuthiekanun et al., 2014), and different combinations of antibiotics that inhibit contaminants (Johnson and Rogers, 1964; Myers, 1975; Zacarias et al., 2008; Miraglia et al., 2009; Chakraborty et al., 2011). However, there are few studies regarding the effect of these supplements on the initial isolation and maintenance of the strains as well as the supplement's influence on the viability, motility, and leptospiral cell morphology.

In the last three decades, molecular methods such as pulse field gel electrophoresis (PFGE), restriction fragment length polymorphism (RFLP), Multiple Locus Sequence Typing (MLST), and multilocus variable-number tandem-repeat analysis (MLVA) were introduced to diagnosis, identification and characterization of leptospires (Thiermann et al., 1985; Herrmann et al., 1992; Perolat et al., 1993; Ralph et al., 1993; Majed et al., 2005; Ahmed et al., 2006). In particular, the MLVA method is useful, low time-consuming and has accessible costs for detecting and identifying *Leptospira* serovars (Pourcel et al., 2003), including the genotype Hardjobovis of serovar Hardjo (Chideroli et al., 2016).

This work describes different culture media compositions for isolating two genotypes of serovar Hardjo, their molecular characterization by MLVA, and their phylogeny using a partial sequence of the *secY* gene.

MATERIALS AND METHODS

Selection of Animals

Cows naturally infected from three small dairy farms in the northern Parana state were monitored serologically for leptospirosis due to their history of reproductive failure. Of these animals, 10 cows that tested positive on the MAT test (titers between 100 and 1,600 for serogroup sejroe) were selected, and their urine was collected for bacterial isolation and DNA detection by PCR.

Animal Ethics and Usage

The study was carried out in accordance with the recommendations of National Council for Control of Animal Experimentation (CONCEA). The protocol was approved by The Ethics Committee on Animal Use (CEUA) from State University of Londrina number CEEA - 58/06.

Preparation of the New Culture Medium

Three formulations of base liquid culture media were produced to isolate and maintain the leptospires. All media contained basic ingredients and additional supplements (**Table 1**).

Culture medium A contained *Leptospira* Medium Base culture Ellinghausen-McCullough-Johnson-Harris (EMJH) (2.56 g/L; Difco™, InterLab, BR), *Leptospira* Enrichment EMJH (100 mL/L; Difco™, InterLab, BR), and sodium pyruvate (0.1 g/L; Sigma®, USA). Medium B contained the same components as medium A with the addition of the enzyme, superoxide dismutase (25,000 U/L; Sigma®, USA). Medium C was similar to B; however, the base Enrichment EMJH (rabbit serum supplement) was changed to fetal bovine serum (100 mL/L; Gibco®, USA).

When necessary, the following antibiotics were added to the three culture media formulations: 5-fluorouracil (400 mg/L, Sigma®, USA), chloramphenicol (5 mg/L, Sigma®, USA), nalidixic acid (50 mg/L, Sigma®, USA), neomycin (10 mg/L, Sigma®, USA), and vancomycin (10 mg/L, Acros®, USA) (Zacarias et al., 2008).

Urine Collection and Culture

A urine sample from each animal was obtained by perineal massage and immediately seeded in tubes containing either culture medium A, B, and C with the five antibiotics. After incubation at 28°C for 24 h, the cultures were seeded in duplicate using the same three different culture media without antibiotics. The initial tubes with antibiotics were discarded after subculturing for 24 h, and the subcultures tubes were evaluated weekly for 6 months with a dark field microscope (Olympus BX40 Model).

Extraction and Amplification of DNA for MLVA and *secY*

For genetic characterization, DNA from the leptospire cultures was extracted using the PureLink Genomic DNA Mini Kit (Invitrogen Life Technologies, Eugene, OR, USA). DNA from the leptospiral strain isolates was amplified using the Platinum PCR SuperMix Kit (Invitrogen Life Technologies, Eugene, OR, USA) according to the following conditions: 45 μL of each reaction containing SuperMix, 1 μL of each primer (10 nM), and 3 μL of DNA template (∼50 ng). All products were analyzed by electrophoresis in a 2% agarose gel with ethidium bromide (0.5 g/mL) in 0.5X TBE buffer (89 mM Tris, 89 mM boric acid, and 2 mM EDTA), pH 8.4, and visualized under ultraviolet light.

TABLE 1 | Resume of the combinations of the EMJH culture media.

Culture medium*	Formulation
A	EMJH base medium + base enrichment + sodium pyruvate
B	EMJH base medium + base enrichment + sodium pyruvate + superoxide dismutase enzyme
C	EMJH base medium + fetal bovine serum + sodium pyruvate + superoxide dismutase

**All culture media were produced both with and without antibiotics.*

Molecular size was estimated by comparison with a 100-bp ladder.

Molecular Typing of the Isolates

To characterize the *Leptospira* strains, two molecular techniques were used. The MLVA identified isolates with five primer pairs for the VNTR loci 4, 7, 10, LB4, and LB5 as previously described (Salaün et al., 2006). For each of the five PCRs, the VNTR loci were used as positive controls for the reference strains of *L. interrogans* serovar Canicola serogroup canicola strain Canicola Hond Utrecht IV, *L. interrogans* serovar Hardjo serogroup sejroe genotype Hardjoprajitno strain Hardjoprajitno, and *L. borgpetersenii* serovar Hardjo serogroup sejroe genotype Hardjobovis strain Sponselee. After amplification, the sequencing of *secY* was used to identify and confirm genetic species, as previously described (Ahmed et al., 2006).

The products of the *secY* gene amplification were purified with a Purelink Genomic DNA extraction kit (Invitrogen Life Technologies, Eugene, OR, USA), quantified by a Qubit™ Fluorometer (Invitrogen Life Technologies, Eugene, OR, USA), and sequenced on a ABI3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using forward and reverse primers. Sequence quality was analyzed by the Phred program (http://asparagin.cenargen.embrapa.br/phph/). The consensus sequences were obtained by CAP3 software (http://asparagin.cenargen.embrapa.br/cgi-bin/phph/cap3.pl), and the identities were compared with the sequences in GenBank using the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The identity matrix was created in the BioEdit program with the alignment and phylogenetic tree developed by the MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets (Kumar et al., 2016).

Scanning Electron Microscopy (SEM)

After isolation, one isolate of each genotype (Hardjobovis and Hardjoprajitno) was subcultered in the three culture media formulations at 28°C for 7 days. Next, the cultures were centrifuged for 5 min at 2,000 rpm, resuspended in 100 μL of fixative (2.5% glutaraldehyde and 2% paraformaldehyde in 0.1 M sodium cacodylate buffer, pH 7.0) and transferred to 24-well polystyrene microtiter plates (Nunc, Roskilde, Denmark) with glass coverslips pre-coated with a thin layer of poly-L-lysine (Sigma Chemical Co, USA). After 1 h, the volume was adjusted to 500 μL of fixing solution to avoid cells adhering to the coverslips, and incubated at 25°C for 12 h. Samples were post-fixed in 1% OsO_4 (Electron Microscopy Sciences, Washington, PA, USA) and dehydrated in an ethanol series (30, 50, 70, 90, and 100°GL). Samples were critical-point dried with CO_2 (BALTEC CPD 030 Critical Point Dryer), coated with gold (BALTEC SDC 050 Sputter Coater), and observed under a SEM (FEI Quanta 200, Netherlands).

RESULTS

The combined formulations of EMJH liquid culture media allowed the growth of *Leptospira* serovar Hardjo strains from three farms (one strain of each farm). Posteriorly, these

three strains were characterized by molecular methods as *L. interrogans* genotype Hardjoprajitno (strain Londrina 53) and *L. borgpetersenii* genotype Hardjobovis (strains Londrina 49 and Londrina 54).

Table 2 shows that at the start of the experiment (12 ± 2 days), the first cells were unexpectedly seen in culture medium A with characteristics and movement similar to leptospires, while no cells were seen in culture media B or C. Thus, subculturing was performed using new tubes with the formulation media A, B, and C (A→A; A→B; A→C). After 21 ± 2 days, only subculture A→C presented leptospire cells, but they were few in number and with signs of suffering. On day 27 (±2), this tube was evaluated again and contained many leptospire cells; therefore, media A, B, and C were subcultered again (A→C→A; A→C→B; A→C→C). On day 34 (±3), there was no increase in cell number in subculture AC compared with subculture A→C→B, but there was an increased number of dead (unmoving) cells. In contrast, subculture A→C→B on day 34 (±3) presented many leptospiral cells and was transferred to new subcultures for medium B and C (A→C→B→B; A→C→B→C).

On day 41 (±2), the subculture A→C→B→C, unexpectedly presented more leptospiral cells per field than the subculture A→ C→ B→ B, and subculturing was performed on both tubes (A→C→B→C→B; A→C→B→C→C and A→C→B→B→B; A→C→B→B→C). On day 48 (±3), the subculture ACBC retained good cell growth as did subculture ACBCB. In contrast, subculture ACBCC had cells with less growth.

At the end of the experiment, all medium B subcultures, particularly A→C→B→C→B→B, presented excellent growth and gradually adapted to the standard routine media used in the laboratory without pyruvate sodium, superoxide dismutase, and fetal bovine serum. The isolated strains were named Londrina 49, Londrina 53, and Londrina 54.

Electron microscopy revealed that the leptospiral cells in culture medium B had a morphology that was more elongated as well as more cells, suggesting more bacterial cell division rate (**Figures 1A,C,E**). In contrast, there were fewer cells in medium C and those cells had lost the typical spirochete morphology (corkscrew-shaped with hooked ends) (**Figures 1B,D,F**). Morphologically, medium A was similar to B, but did not have a high number of cells and shown few elongated cells in division (Supplementary Figure 1).

The Londrina 53 strain was characterized by MLVA and genetic sequencing, and was identified as *L. interrogans* serovar Hardjo genotype Hardjoprajitno (**Figures 2**, **3**). The other two isolated strains (Londrina 49 and Londrina 54) from the same culture medium formulation were molecularly characterized as *L. borgpetersenni* genotype Hardjobovis, which was previously published (Chideroli et al., 2016).

The phylogenetic tree for all isolates shows that the Londrina 53 strain was grouped in the same cluster as *L. interrogans* and had the sequence identity of the serovar Hardjo genotype Hardjoprajitno. The others strains (Londrina 49 and Londrina 54) remained in the same cluster as *L. borgpetersenii* (**Figure 4**).

DISCUSSION

Hardjo is the most prevalent serovar and causative agent of leptospirosis in dairy and beef cattle herds. It causes reproductive

TABLE 2 | Evaluation by dark field microscopy of the cultured *Leptospira* (Londrina 49, Londrina 53, and Londrina 54 strains) isolated from naturally infected bovine urine.

Evaluation date#	Culture medium*	Dark field microscopic evaluation	Subculture
0 day	A, B, C with antibiotics	NP	NP
1 day	A, B, C with antibiotics	NP	A, B, C without antibiotics
12 ± 2 day	A	Many cells with characteristics and movement similar to *Leptospira*	A→A A→B A→C
21 ± 2 day	A→C	Few structures with signs of suffering	NP
27 ± 2 day	A→C	Good number of leptospiral cells	A→C→A A→C→B A→C→C
34 ± 3 day	A→C	Stagnation of growth with no moving cells	NP
34 ± 3 day	A→C→B	Many cells similar to *Leptospira*	A→C→B→B A→C→B→C
41 ± 2 day	A→C→B→B	One *Leptospira* cell per field	A→C→B→B→B A→C→B→B→C
41 ± 2 day	A→C→B→C	Many cells similar to *Leptospira*	A→C→B→C→B A→C→B→C→C
48 ± 3 day	A→C→B→C	Good number of leptospiral cells	NP
48 ± 3 day	A→C→B→C→B	Good number of leptospiral cells	A→C→B→C→B→B A→C→B→C→B→C
48 ± 3 day	A→C→B→C→C	Few cells	NP
55 ± 4 day	A→C→B→C→B→B	Optimum growth and adaptation to medium B	A→C→B→C→B→B→B

*#On the dark field microscopic evaluation date, all tubes were evaluated, but only the significant data are presented in the table; *Numbers corresponding to the culture media designated in **Table 1**; → symbol represents a subculture for another medium; NP, not performed.*

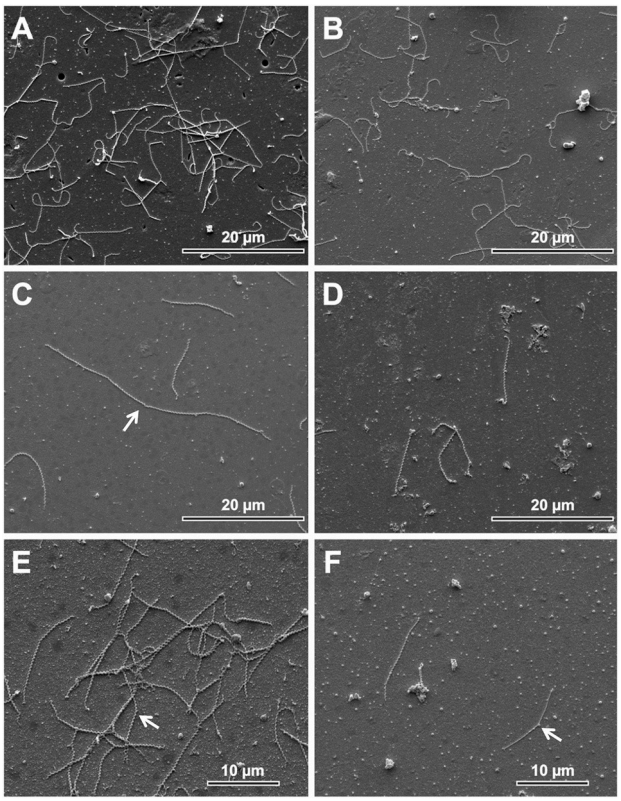


FIGURE 1 | Morphology of leptospire cell in the culture media by scanning electronic microscopy. **(A)** Culture media B with higher number of leptospiral cells of serovar Hardjo; **(B)** Culture media C with fewer number of leptospiral cells of serovar Hardjo; **(C)** culture media B with individualized cells and one elongated cell in division (arrow) of serovar Hardjo. **(D)** Culture media C with fewer cells of *Leptospira* serovar Hardjo in cell division; **(E)** Presence of typical spirochete morphology in culture media B with cells of *Leptospira* serovar Hardjo, Arrow—indicates spirochetal corkscrew morphology; **(F)** Lost of typical spirochete morphology in culture media C with cells of *Leptospira* serovar Hardjo, Arrow—indicates loss of spirochetal corkscrew morphology.

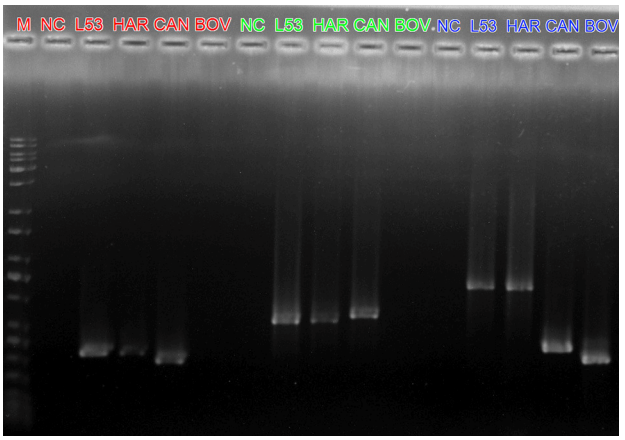


FIGURE 2 | Banding patterns of VNTR visualized in agarose gel. Lane M = 1 kb molecular ladder bp (Kasvi, Curitiba, PR, Brazil), L53 (Londrina 53 strain); HAR = reference sample serovar Hardjo strain Hardjoprajitno; CAN = reference sample serovar Canicola strain Hond Utrecht IV; BOV = reference sample serovar Hardjo strain Hardjobovis; NC = negative control. Locus colors: red (VNTR-4), green (VNTR-7), and blue (VNTR-10).

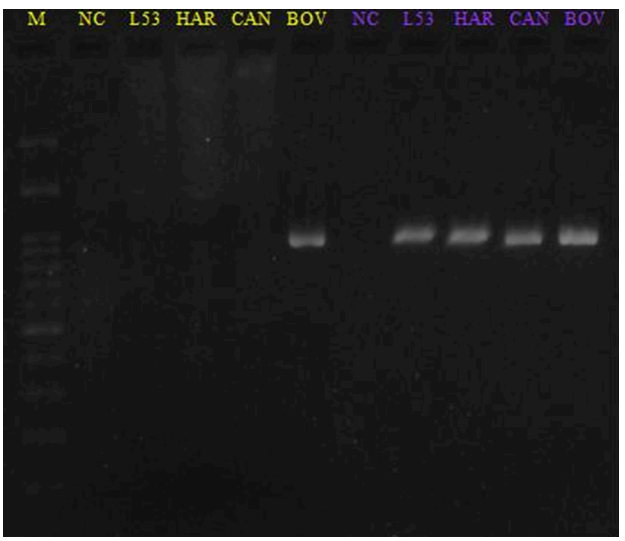


FIGURE 3 | Banding patterns of VNTR visualized in agarose gel. Lane M = molecular ladder bp, L53 (Londrina 53 strain); HAR = reference sample serovar Hardjo strain Hardjoprajitno; CAN = reference sample serovar Canicola strain Hond Utrecht IV; BOV = reference sample serovar Hardjo strain Hardjobovis; NC = negative control. Locus colors: yellow (VNTR-Lb4) and purple (VNTR-Lb5).

failure in livestock worldwide and results in substantial economic loss due to infertility and abortion (Ellis, 2015). In Latin America, few studies have reported recovery of serovar Hardjo (genotype Hardjoprajitno) in cattle (Aycardi et al., 1980; Salgado et al., 2015). In Brazil, two strains, Norma and 2012_OV5, were previously isolated from bovine and ewe, respectively, and characterized as belonging to *L. interrogans* genotype Hardjoprajitno (Cosate et al., 2012; Director et al., 2014). Recently, a Hardjo serovar isolated from bovine urine for the first time in Latin America, was molecularly characterized by MLVA and gene *secY* sequencing as *L. borgpetersenii* genotype Hardjobovis (Chideroli et al., 2016).

EMJH culture media (liquid or semi-solid) is widely used to isolate *Leptospira* (Rodríguez et al., 2002; González et al., 2006; Zacarias et al., 2008; Miraglia et al., 2009; Chakraborty et al., 2011). For many years, our leptospirosis research group has unsuccessfully attempted to isolate serovar Hardjo from bovine urine using unmodified EMJH, even after obtaining positive serology for this serovar (Hashimoto et al., 2017).

In the present study, the combined use of EMJH culture media supplemented with sodium pyruvate (culture medium A); sodium pyruvate and superoxide dismutase (culture medium B); and sodium pyruvate, superoxide dismutase and fetal bovine serum (culture medium C) was used to isolate, adapt, and maintain fastidious leptospires, such as serovar Hardjo, in the laboratory.

The two core ingredients (EMJH medium base and enrichment base) were sustained because they are present in a

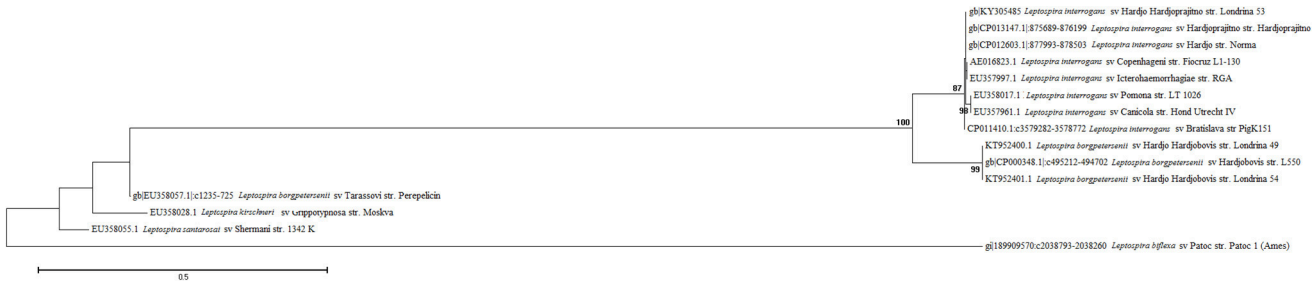


FIGURE 4 | Phylogenetic tree of *Leptospira* with partial *secY* gene sequences of the reference strains and a set of clinical isolates from bovine urine (Londrina 49, Londrina 53, and Londrina 54). The phylogenetic tree was drawn using the maximum likelihood method based on the Tamura 3-parameter model. Reliability of the branches was validated by generating 1,000 "bootstrap" replicates. Evolutionary analyses were conducted in MEGA7. The obtained sequences were deposited in the GenBank database with the following access codes: Londrina 49—KT952400, Londrina 53—KY305485, and Londrina 54 – KT952401.

commonly used medium for *Leptospira* isolation, and sodium pyruvate was included in all culture medium formulations because of a previous report that it enhances *Leptospira* growth when added to a solid medium (Johnson et al., 1973). Superoxide dismutase and fetal bovine serum were chosen to aid the growth of *Leptospira* serovar Hardjo, which has exigent nutritional requirements. Different combinations were made to verify the usefulness of each one. More specifically, superoxide dismutase was chosen because it eliminates the toxic radicals produced during spirochete metabolism that accumulate in the culture medium and the concentration was the same used in the culture media for *Treponema* which is a spirochete as well as leptospires (Austin et al., 1981; Cox et al., 1990). Fetal bovine serum was used to replace the rabbit serum present in the enrichment EMJH base (essential requirement for leptospire growth) because serovar Hardjo is thought to be better adapted to the bovine host. Indeed, future studies using a metabolomics approach could help us to better understand the molecules that are involved in adapting these bacteria to the culture media.

These results show that adding these components to the standard culture medium allowed rapid isolation of the Hardjo serovar. As shown in **Table 2**, the formulation with only sodium pyruvate (medium A) was where the first *Leptospira* cells were observed and where the primary isolation occurred. For leptospiral culture media, sodium pyruvate has previously been demonstrated to promote growth (Johnson et al., 1973). In a study on hydrogen peroxide damage in mammalian cell cultures, Giandomenico et al. (1997) found that sodium pyruvate was most effective in eliminating hydrogen peroxide and its toxic effects. Therefore, adding this component to the culture medium for isolating fastidious *Leptospira* serovars may be essential for its primary isolation.

After the first simultaneous subcultures with medium B and C (A→B and A→C), fetal bovine serum was found to be critical for initiating growth. However, when it was passed to a new medium with fetal bovine serum (A→C→C) there was no increase in cell growth. In contrast, subculturing the medium with fetal bovine serum to basal medium supplemented with superoxide dismutase without fetal bovine serum (A→C→B→C→B and A→C→B→C→B→B) increased the bacterial growth and final adaptation in the culture medium. In other words, the fetal bovine serum was important for initial growth but eventually, someway became detrimental, and the culture medium required only superoxide dismutase with an enrichment base of EMJH for final leptospiral adaptation. This result suggests that there is a distinct difference between the culture medium requirement for primary isolation and for the maintenance and adaptation of new isolated strains.

More importantly, the change in leptospiral morphological characteristics in the different media suggests that leptospires in the presence of fetal bovine serum will begin to show signs of suffering such as low growth rate, lower replication rate, loss of their corkscrew-shape and loss of their hooked ends, and will eventually die. A genetics study performed with spirochete non-motile mutants indicate that the periplasmic flagella were involved in spirochete motility. It also indicated that the structure of the flagella influenced the shape of the cell ends (Li et al., 2000). Thus, if the culture medium does not provide the elements necessary for bacterial growth and development, or becomes toxic, the cell metabolism and mobility decrease, and thus, lose their hooked ends (**Figure 1F**). In contrast, the presence of superoxide dismutase seemingly detoxifies the culture medium and allows bacterial development with increased cell size due the possible number of leptospires replicating (**Figure 1C**).

The ability to isolate and maintain *Leptospira* spp. is critical for both diagnostic and research purposes using molecular characterization to identify new isolates (Adler and de la Peña Moctezuma, 2010). To exchange information between laboratories, the MLVA molecular biology technique is efficient, with easy standardization, rapid clinical diagnosis, and can be applied in the field of epidemiology (Salaün et al., 2006; Slack et al., 2006).

Among the techniques used for molecular characterization of leptospiras, the MLST as the MLVA is a simple PCR based technique. The selected loci of MLST are generally the housekeeping genes, which evolve very slowly over an evolutionary time-scale (Enright and Spratt, 1999). However, this methodology depends of sequencing of seven genes which make the technique more expensive and time consuming. In this study, only the *secY* gene was used, which consists of conserved and variable regions with sufficient sequence heterogeneity to enable the phylogenetic classification of *Leptospira* genus (Victoria et al., 2008; Hamond et al., 2016). Another technique widely used for leptospires genotyping is PFGE, but this methodology requires

specific structure/equipaments that may not be available in all diagnosis and research laboratories.

Currently, the MLVA method is one practical alternative for differentiation and identification of the many pathogenic *Leptospira* serovar, including the differentiation of two serovar Hardjo genotypes (Salaün et al., 2006). Furthermore, identifying and typing new isolate strains is important for understanding disease epidemiology in the region, as well as developing diagnostic tools, effective vaccines, and prevention strategies for leptospirosis (Ahmed et al., 2011). The results obtained from the MLVA technique were corroborated by the sequential analysis of the partial *secY* gene, which confirmed the genetic species (Ahmed et al., 2006).

CONCLUSION

In this study, culture medium formulations were created to isolate fastidious leptospires of the serovar Hardjo genotypes Hardjobovis and Hardjoprajitno from urine of naturally infected bovine. Noteworthy, additional components were useful for the initial growth (sodium pyruvate and fetal bovine serum) and subsequent maintenance of leptospires (superoxide dismutase) adapted to the medium standard. With this strategy, using three formulations, we succeeded in isolating three pure strains of the serovar Hardjo. After isolation, the technique of MLVA associated with the partial sequencing of gene *secY* have been validated and suggested for molecular characterization of serovars such as Hardjo that may belong to different species. Additionally, an evaluation of leptospire cells in the three formulations by electronic microscopy showed differences in spirochete morphology based on the supplement used in each medium. The superoxide dismutase enzyme induced stretching and cell division; in contrast, cells in the fetal bovine serum medium were fewer in number and lost their corkscrew-shape and hooked ends. Finally, the culture strategies described in this study allowed the isolation and rapid molecular characterization of two serovar Hardjo genotypes inducing new insight into seroepidemiology, specific genotypes, and immunoprophylaxis for leptospirosis in dairy and beef cattle herds.

AUTHOR CONTRIBUTIONS

JdF, UP, and RC planned the project and designed the experiments. RC conducted the experiments and carried out the data analysis with help from JdF, UP, AFA, DG, and AAA. RC, SS, DG, AdO, and UP contributed reagents preparation and samples collection. RC wrote the manuscript, which was critically reviewed by JdF, AAA, AFA, and UP.

FUNDING

We acknowledge support with fellowships from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for RC. AFA and AAA are recipients of CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) fellowships.

ACKNOWLEDGMENTS

We are grateful to Prof. Dr. Silvio Arruda Vasconcellos that provided *L. borgpetersenii* serovar Hardjo strain Hardjobovis used as positive control of MLVA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02155/full#supplementary-material

REFERENCES

Adler, B., Fainea, S., Christophera, W. L., and Chappel, R. J. (1986). Development of an improved selective medium for isolation of leptospires from clinical material. *Vet. Microbiol.* 12, 377–381. doi: 10.1016/0378-1135(86)90087-8

Adler, B., and de la Peña Moctezuma, A. (2010). Leptospira and leptospirosis. *Vet. Microbiol.* 140, 287–296. doi: 10.1016/j.vetmic.2009.03.012

Ahmed, A., Thaipadungpanit, J., Boonsilp, S., Wuthiekanun, V., Nalam, K., Spratt, B. G., et al. (2011). Comparison of two multilocus sequence based genotyping schemes for Leptospira species. *PLoS Negl. Trop. Dis.* 5:e1374. doi: 10.1371/journal.pntd.0001374

Ahmed, N., Devi, S. M., Valverde, M. L., Vijayachari, P., Machang'u, R. S., Ellis, W. A., et al. (2006). Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. *Ann. Clin. Microbiol. Antimicrob.* 5:28. doi: 10.1186/1476-0711-5-28

Austin, F. E., Barbieri, J. T., Corin, R. E., Grigas, K. E., and Cox, C. D. (1981). Distribution of superoxide dismutase, catalase, and peroxidase activities among *Treponema pallidum* and other spirochetes. *Infect. Immun.* 33, 372–379.

Aycardi, E. R., Myers, D. M., and Torres, B. (1980). A new Leptospiral serovar in the Tarassovi serogroup from Colombia. *Z. Vet. R. B* 27, 425–428.

Chakraborty, A., Miyahara, S., Villanueva, S. Y. A. M., Saito, M., Gloriani, N. G., and Yoshida, S. I. (2011). A novel combination of selective agents for isolation of *Leptospira* species. *Microbiol. Immunol.* 55, 494–501. doi: 10.1111/j.1348-0421.2011.00347.x

Bey, R. F., and Johnson, R. C. (1978). Protein-free and low-protein media for the cultivation of Leptospira. *Infect. Immun.* 19, 562–569.

Chideroli, R. T., Pereira, U. P., Nakamura, A. Y., Alfieri, A. A., Alfieri, A. F., and Freitas, J. C. (2016). Isolation and molecular characterization of *Leptospira borgpetersenii* serovar Hardjo strain Hardjobovis in the urine of naturally infected cattle in Brazil. *Genet. Mol. Res.* 15, 1–7. doi: 10.4238/gmr.15018473

Cosate, M. R. V., Barouni, A. S., Moreira, E. C., Veloso, I. F., Gomes, M. T. R., and Salas, C. E. (2012). Molecular characterization by LSSP-PCR and DNA sequencing of a pathogenic isolate of *Leptospira interrogans* from Brazil. *Zoonoses Public Health* 59, 379–388. doi: 10.1111/j.1863-2378.2012.01470.x

Cox, D. L., Riley, B., Chang, P., Sayahtaheri, S., Tassell, S., and Hevelone, J. (1990). Effects of molecular oxygen, oxidation-reduction potential, and antioxidants upon *in vitro* replication of *Treponema pallidum* subsp. pallidum. *Appl. Environ. Microbiol.* 56, 3063–3072.

de la Peña-Moctezuma, A., Bulach, D. M., Kalambaheti, T., and Adler, B. (1999). Comparative analysis of the LPS biosynthetic loci of the genetic subtypes of serovar Hardjo: *Leptospira interrogans* subtype Hardjoprajitno and *Leptospira borgpetersenii* subtype Hardjobovis. *FEMS Microbiol. Lett.* 177, 319–326. doi: 10.1111/j.1574-6968.1999.tb13749.x

Director, A., Penna, B., Hamond, C., Loureiro, A. P., Martins, G., Medeiros, M. A., et al. (2014). Isolation of *Leptospira interrogans* Hardjoprajitno from vaginal fluid of a clinically healthy ewe suggests potential for venereal transmission. *J. Med. Microbiol.* 63, 1234–1236. doi: 10.1099/jmm.0.065466-0

Ellis, W. A. (2015). Animal leptospirosis. *Curr. Top. Microbiol. Immunol.* 387, 99–137. doi: 10.1007/978-3-662-45059-8_6

Ellis, W. A., and Thiermann, A. B. (1986). Isolation of leptospires from the genital tracts of Iowa cows. *Am. J. Vet. Res.* 47, 1694–1696.

Enright, M. C., and Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol.* 7, 482–487. doi: 10.1016/S0966-842X(99)01609-1

Faine, S., Adler, B., Bolin, C., and Perolat, P. (1999). *Leptospira and Leptospirosis, 2nd Edn.* Melbourne, VIC: MedSci Press.

Favero, M., Pinheiro, S. R., Vasconcellos, S. A., Morais, Z. M., Ferreira, F., Neto, J. S. F., et al. (2001). *Leptospirose bovina* - variantes sorológicas predominantes em colheitas efetuadas no período de 1984 a 1997 em rebanhos de 21 estados do Brasil. *Arq. Inst. Biológico* 68, 29–35.

Figueiredo, A. D. O., Pellegrin, A. O., Gonçalves, V. S. P., Freitas, B., Monteiro, L. A. R. C., and de Oliveira, J. M. (2009). Prevalência e fatores de risco para a leptospirose em bovinos de Mato Grosso do Sul. *Pesqui. Vet. Bras.* 29, 375–381. doi: 10.1590/S0100-736X2009000500003

Flint, S. H., and Liardet, D. M. (1980). Isolation of *Leptospira interrogans* serovar Hardjo from bovine urine. *N. Z. Vet. J.* 28, 55. doi: 10.1080/00480169.1980.34693

Giandomenico, A. R., Cerniglia, G. E., Biaglow, J. E., Stevens, C. W., and Koch, C. J. (1997). The importance of sodium pyruvate in assessing damage produced by hydrogen peroxide. *Free Radic. Biol. Med.* 23, 426–434. doi: 10.1016/S0891-5849(97)00113-5

González, A., Borrero, R., Ruiz, J., Batista, N., Fernández, Y., Valdés, Y., et al. (2006). Medio EMJH modificado para el cultivo de *Leptospira interrogans* serogrupo Ballum. *Rev. Argent. Microbiol.* 38, 61–68.

Hamond, C., Pestana, C., Medeiros, M., and Lilenbaum, W. (2016). Genotyping of Leptospira directly in urine samples of cattle demonstrates a diversity of species and strains in Brazil. *Epidemiol. Infect.* 144, 72–75. doi: 10.1017/S0950268815001363

Hashimoto, V. Y., Chideroli, R. T., Ribeiro, J., Alfieri, A. A., Costa, G. M., Freitas, J. C., et al. (2017). Serological and molecular findings in diagnosis of leptospirosis serovar hardjo in a dairy bovine herd. *Semin. Cien. Agrar.* 38, 3155–3164. doi: 10.5433/1679-0359.2017v38n5p3155

Hashimoto, V. Y., Dias, J. A., Spohr, K. A. H., Silva, M. C. P., Andrade, M. G. B., Müller, E. E., et al. (2012). Prevalência e fatores de risco associados á *Leptospira* spp. em rebanhos bovinos da região centro-sul do estado do Paraná. *Pesqui. Vet. Bras.* 32, 99–105. doi: 10.1590/S0100-736X2012000200001

Herrmann, J. L., Bellenger, E., Perolat, P., Baranton, G., and Saint Girons, I. (1992). Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification. *J. Clin. Microbiol.* 30, 1696–1702.

Johnson, R. C., and Gary, N. D. (1962). Nutrition of Leptospira pomona. 1. A chemically defined substitute for rabbit serum ultrafiltrate. *J. Bacteriol.* 83, 668–672.

Johnson, R. C., and Rogers, P. (1964). 5-Fluorouracil as a selective agent for growth of Leptospirae. *J. Bacteriol.* 87, 422–426.

Johnson, R. C., Walby, J., Henry, R. A., and Auran, N. E. (1973). Cultivation of parasitic leptospires: effect of pyruvate. *Appl. Microbiol.* 26, 118–119.

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

Leonard, F. C., Quinn, P. J., Ellis, W. A., and O'Farrell, K. (1992). Duration of urinary excretion of leptospires by cattle naturally or experimentally infected with *Leptospira interrogans* serovar hardjo. *Vet. Rec.* 131, 435–439. doi: 10.1136/vr.131.19.435

Li, C., Motaleb, A., Sal, M., Goldstein, S. F., and Charon, N. W. (2000). Spirochete periplasmic flagella and motility. *J. Mol. Microbiol. Biotechnol.* 2, 345–354.

Majed, Z., Bellenger, E., Postic, D., Pourcel, C., Baranton, G., and Picardeau, M. (2005). Identification of variable-number tandem-repeat loci in *Leptospira interrogans* sensu stricto. *J. Clin. Microbiol.* 43, 539–545. doi: 10.1128/JCM.43.2.539-545.2005

Miraglia, F., De Moraes, Z. M., Melville, P. A., Dias, R. A., and Vasconcellos, S. A. (2009). Emjh medium with 5-fluorouracil and nalidixic acid associated with serial dilution technique used to recover leptospira spp from experimentally contaminated bovine semen. *Braz. J. Microbiol.* 40, 189–193. doi: 10.1590/S1517-83822009000100033

Myers, D. M. (1975). Efficacy of combined furazolidone and neomycin in the control of contamination in Leptospira cultures. *Antimicrob. Agents Chemother.* 7, 660–671. doi: 10.1128/AAC.7.5.666

Pailhoriès, H., Buzelé, R., Picardeau, M., Robert, S., Mercier, E., Mereghetti, L., et al. (2015). Molecular characterization of *Leptospira* sp by multilocus variable number tandem repeat analysis (MLVA) from clinical samples: a case report. *Int. J. Infect. Dis.* 37, 119–121. doi: 10.1016/j.ijid.2015.06.026

Perolat, P., Lecuyer, I., Postic, D., and Baranton, G. (1993). Diversity of ribosomal DNA fingerprints of Leptospira serovars provides a database for subtyping and species assignation. *Res. Microbiol.* 144, 5–15. doi: 10.1016/0923-2508(93)90210-S

Picardeau, M. (2013). Diagnosis and epidemiology of leptospirosis. *Med. Mal. Infect.* 43, 1–9. doi: 10.1016/j.medmal.2012.11.005

Pourcel, C., Vidgop, Y., Ramisse, F., Vergnaud, G., and Tram, C. (2003). Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. *J. Clin. Microbiol.* 41, 1819–1826. doi: 10.1128/JCM.41.5.1819-1826.2003

Ralph, D., McClelland, M., Welsh, J., Baranton, G., and Perolat, P. (1993). Leptospira species categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes. *J. Bacteriol.* 175, 973–981. doi: 10.1128/jb.175.4.973-981.1993

Robertson, A., Boulanger, P., and Mitchell, D. (1964). Isolation and identification of a leptospire of the hebdomadis serogroup (L hardjo) from cattle in Canada. *Can. J. Comp. Med.* 28, 13–18.

Rodríguez, A. G., Santiesteban, N. B., Abreu, Y. V., and González, M. G. (2002). Crecimiento, virulencia y antigenicidad de *Leptospira interrogans* serovar mozdok en medio EMJH modificado. *Rev. Cuba. Med. Trop.* 54, 32–36.

Salaün, L., Mérien, F., Gurianova, S., Baranton, G., and Picardeau, M. (2006). Application of multilocus variable-number tandem-repeat analysis for molecular typing of the agent of leptospirosis. *J. Clin. Microbiol.* 44, 3954–3962. doi: 10.1128/JCM.00336-06

Salgado, M., Otto, B., Moroni, M., Sandoval, E., Reinhardt, G., Boqvist, S., et al. (2015). Isolation of *Leptospira interrogans* serovar Hardjoprajitno from a calf with clinical leptospirosis in Chile. *BMC Vet. Res.* 11:66. doi: 10.1186/s12917-015-0369-x

Silva, F. J., Conceição, W. L. F., Fagliari, J. J., Girio, R. J. S., Dias, R. A., Borba, M. R., et al. (2012). Prevalência e fatores de risco de leptospirose bovina no Estado do Maranhão. *Pesqui. Vet. Bras.* 32, 303–312. doi: 10.1590/S0100-736X2012000400006

Slack, A. T., Symonds, M. L., Dohnt, M. F., and Smythe, L. D. (2006). Identification of pathogenic Leptospira species by conventional or real-time PCR and sequencing of the DNA gyrase subunit B encoding gene. *BMC Microbiol.* 6:95. doi: 10.1186/1471-2180-6-95

Staneck, J. L., Henneberry, R. C., and Cox, C. D. (1973). Growth requirements of pathogenic leptospira. *Infect. Immun.* 7, 886–897.

Thiermann, A. B., Handsaker, A. L., Moseley, S. L., and Kingscote, B. (1985). New method for classification of leptospiral isolates belonging to serogroup pomona by restriction endonuclease analysis: serovar kennewicki. *J. Clin. Microbiol.* 21, 585–587.

Victoria, B., Ahmed, A., Zuerner, R. L., Ahmed, N., Bulach, D. M., and Hartskeerl, R. A. (2008). Conservation of the S10-spc- a locus within otherwise highly plastic genomes provides phylogenetic insight into the genus Leptospira. *PLoS ONE* 3:e2752. doi: 10.1371/journal.pone.0002752

Wuthiekanun, V., Amornchai, P., Langla, S., Oyuchua, M., Day, N. P. J., and Limmathurotsakul, D. (2014). Maintenance of *Leptospira* species in Leptospira Vanaporn Wuthiekanun agar. *J. Clin. Microbiol.* 52, 4350–4352. doi: 10.1128/JCM.02273-14

Zacarias, F. G. D. S., Vasconcellos, S. A., Anzai, E. K., Giraldi, N., De Freitas, J. C., and Hartskeerl, R. (2008). Isolation of Leptospira serovars Canicola and Copenhageni from cattle urine in the State of Parana, Brazil. *Braz. J. Microbiol.* 39, 744–748. doi: 10.1590/S1517-83822008000400028

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 Chideroli, Gonçalves, Suphoronski, Alfieri, Alfieri, de Oliveira, de Freitas and Pereira. 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