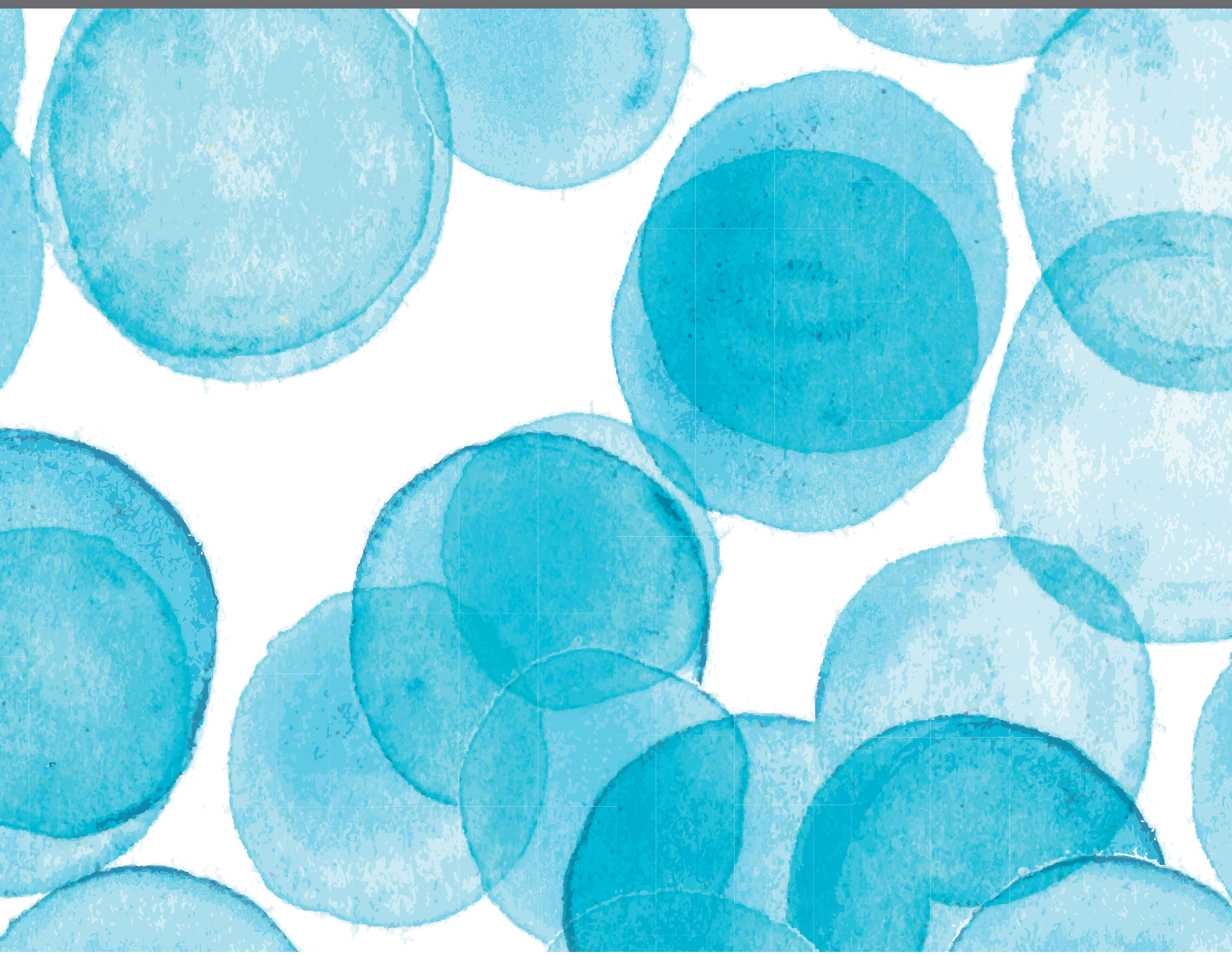
EXTREMOPHILES: MICROBIAL GENOMICS AND TAXOGENOMICS

EDITED BY: Rafael R. de la Haba, André Antunes and Brian P. Hedlund
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 978-2-88976-931-5
DOI 10.3389/978-2-88976-931-5

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: frontiersin.org/about/contact

EXTREMOPHILES: MICROBIAL GENOMICS AND TAXOGENOMICS

Topic Editors:
Rafael R. de la Haba, University of Sevilla, Spain
André Antunes, Macau University of Science and Technology, China
Brian P. Hedlund, University of Nevada, Las Vegas, United States

Citation: de la Haba, R. R., Antunes, A., Hedlund, B. P., eds. (2022). Extremophiles: Microbial Genomics and Taxogenomics. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-931-5

Table of Contents

06 ***Editorial: Extremophiles: Microbial Genomics and Taxogenomics***
Rafael R. de la Haba, André Antunes and Brian P. Hedlund

13 ***Diversity and Distribution of a Novel Genus of Hyperthermophilic* Aquificae *Viruses Encoding a Proof-Reading Family-A DNA Polymerase***
Marike Palmer, Brian P. Hedlund, Simon Roux, Philippos K. Tsourkas, Ryan K. Doss, Casey Stamereilers, Astha Mehta, Jeremy A. Dodsworth, Michael Lodes, Scott Monsma, Tijana Glavina del Rio, Thomas W. Schoenfeld, Emiley A. Eloe-Fadrosh and David A. Mead

31 ***Novel Extracellular Electron Transfer Channels in a Gram-Positive Thermophilic Bacterium***
Sergey N. Gavrilov, Daria G. Zavarzina, Ivan M. Elizarov, Tamara V. Tikhonova, Natalia I. Dergousova, Vladimir O. Popov, Jonathan R. Lloyd, David Knight, Mohamed Y. El-Naggar, Sahand Pirbadian, Kar Man Leung, Frank T. Robb, Maksim V. Zakhartsev, Orianna Bretschger and Elizaveta A. Bonch-Osmolovskaya

52 ***The Transcriptomic and Phenotypic Response of the Melanized Yeast* Exophiala dermatitidis *to Ionizing Particle Exposure***
Zachary Schultzhaus, Amy Chen, Igor Shuryak and Zheng Wang

70 **Gemmatimonas groenlandica *sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes***
Yonghui Zeng, Nupur, Naicheng Wu, Anne Mette Madsen, Xihan Chen, Alastair T. Gardiner and Michal Koblížek

88 ***Bioprospecting of Novel Extremozymes From Prokaryotes—The Advent of Culture-Independent Methods***
Maksim Sysoev, Stefan W. Grötzinger, Dominik Renn, Jörg Eppinger, Magnus Rueping and Ram Karan

104 ***Physicochemical and Microbial Diversity Analyses of Indian Hot Springs***
Manik Prabhu Narsing Rao, Zhou-Yan Dong, Zhen-Hao Luo, Meng-Meng Li, Bing-Bing Liu, Shu-Xian Guo, Wael N. Hozzein, Min Xiao and Wen-Jun Li

114 ***Taxogenomic and Comparative Genomic Analysis of the Genus* Saccharomonospora *Focused on the Identification of Biosynthetic Clusters PKS and NRPS***
Ninfa Ramírez-Durán, Rafael R. de la Haba, Blanca Vera-Gargallo, Cristina Sánchez-Porro, Scarlett Alonso-Carmona, Horacio Sandoval-Trujillo and Antonio Ventosa

132 **Methylobacterium ajmalii *sp. nov., Isolated From the International Space Station***
Swati Bijlani, Nitin K. Singh, V. V. Ramprasad Eedara, Appa Rao Podile, Christopher E. Mason, Clay C. C. Wang and Kasthuri Venkateswaran

146 ***Genomics, Exometabolomics, and Metabolic Probing Reveal Conserved Proteolytic Metabolism of* Thermoflexus hugenholtzii *and Three Candidate Species From China and Japan***
Scott C. Thomas, Devon Payne, Kevin O. Tamadonfar, Cale O. Seymour, Jian-Yu Jiao, Senthil K. Murugapiran, Dengxun Lai, Rebecca Lau, Benjamin P. Bowen, Leslie P. Silva, Katherine B. Louie, Marcel Huntemann, Alicia Clum, Alex Spunde, Manoj Pillay, Krishnaveni Palaniappan, Neha Varghese, Natalia Mikhailova, I-Min Chen, Dimitrios Stamatis, T. B. K. Reddy, Ronan O'Malley, Chris Daum, Nicole Shapiro, Natalia Ivanova, Nikos C. Kyrpides, Tanja Woyke, Emiley Eloe-Fadrosh, Trinity L. Hamilton, Paul Dijkstra, Jeremy A. Dodsworth, Trent R. Northen, Wen-Jun Li and Brian P. Hedlund

164 ***Temporal Microbial Community Dynamics Within a Unique Acid Saline Lake***
Noor-Ul-Huda Ghori, Michael J. Wise and Andrew S. Whiteley

178 ***Interplay of Various Evolutionary Modes in Genome Diversification and Adaptive Evolution of the Family* Sulfolobaceae**
Rachana Banerjee, Narendrakumar M. Chaudhari, Abhishake Lahiri, Anupam Gautam, Debaleena Bhowmik, Chitra Dutta, Sujay Chattopadhyay, Daniel H. Huson and Sandip Paul

192 ***Physiological and Genomic Characterization of a Hyperthermophilic Archaeon* Archaeoglobus neptunius *sp. nov. Isolated From a Deep-Sea Hydrothermal Vent Warrants the Reclassification of the Genus* Archaeoglobus**
Galina Slobodkina, Maxime Allioux, Alexander Merkel, Marie-Anne Cambon-Bonavita, Karine Alain, Mohamed Jebbar and Alexander Slobodkin

207 ***Effects of Simulated Microgravity on the Physiology of* Stenotrophomonas maltophilia *and Multiomic Analysis***
Xiaolei Su, Yinghua Guo, Tingzheng Fang, Xuege Jiang, Dapeng Wang, Diangeng Li, Po Bai, Bin Zhang, Junfeng Wang and Changting Liu

225 ***Description, Taxonomy, and Comparative Genomics of a Novel species,* Thermoleptolyngbya sichuanensis *sp. nov., Isolated From Hot Springs of Ganzi, Sichuan, China***
Jie Tang, Liheng Li, Meijin Li, Lianming Du, Md Mahfuzur R. Shah, Michal M. Waleron, Malgorzata Waleron, Krzysztof F. Waleron and Maurycy Daroch

243 ***Performance of Multiple Metagenomics Pipelines in Understanding Microbial Diversity of a Low-Biomass Spacecraft Assembly Facility***
Jason M. Wood, Nitin K. Singh, Lisa Guan, Arman Seuylemezian, James Nick Benardini and Kasthuri Venkateswaran

260 ***Phylogenomics of Haloarchaea: The Controversy of the Genera* Natrinema-Haloterrigena**
Rafael R. de la Haba, Hiroaki Minegishi, Masahiro Kamekura, Yasuhiro Shimane and Antonio Ventosa

284 **Flavobacterium flabelliforme *sp. nov. and* Flavobacterium geliluteum *sp. nov., Two Multidrug-Resistant Psychrotrophic Species Isolated From Antarctica***
Stanislava Králová, Hans-Jürgen Busse, Matěj Bezdíček, Megan Sandoval-Powers, Markéta Nykrýnová, Eva Staňková, Daniel Krsek and Ivo Sedláček

304 *The Piezo-Hyperthermophilic Archaeon* Thermococcus piezophilus *Regulates Its Energy Efficiency System to Cope With Large Hydrostatic Pressure Variations*
Yann Moalic, Jordan Hartunians, Cécile Dalmasso, Damien Courtine, Myriam Georges, Philippe Oger, Zongze Shao, Mohamed Jebbar and Karine Alain

315 *Genomic Insights Into New Species of the Genus* Halomicroarcula *Reveals Potential for New Osmoadaptative Strategies in Halophilic Archaea*
Ana Durán-Viseras, Cristina Sánchez-Porro and Antonio Ventosa

331 *Diversity and Distribution of Anaerobic Ammonium Oxidation Bacteria in Hot Springs of Conghua, China*
Lan Liu, Ai-Ping Lv, Manik Prabhu Narsing Rao, Yu-Zhen Ming, Nimaichand Salam, Meng-Meng Li, Ze-Tao Liu, Xiao-Tong Zhang, Jing-Yi Zhang, Wen-Dong Xian, Jian-Yu Jiao and Wen-Jun Li

344 *Characterization of a Novel Hot-Spring Cyanobacterium* Leptodesmis sichuanensis *sp. Nov. and Genomic Insights of Molecular Adaptations Into Its Habitat*
Jie Tang, Lian-Ming Du, Meijin Li, Dan Yao, Ying Jiang, Malgorzata Waleron, Krzysztof Waleron and Maurycy Daroch

358 *A Large-Scale Genome-Based Survey of Acidophilic Bacteria Suggests That Genome Streamlining Is an Adaption for Life at Low pH*
Diego Cortez, Gonzalo Neira, Carolina González, Eva Vergara and David S. Holmes

374 *Polyphasic Identification and Genomic Insights of* Leptothermofonsia sichuanensis *gen. sp. nov., a Novel Thermophilic Cyanobacteria Within Leptolyngbyaceae*
Jie Tang, Mahfuzur R. Shah, Dan Yao, Ying Jiang, Lianming Du, Kelei Zhao, Liheng Li, Meijin Li, Michal M. Waleron, Malgorzata Waleron, Krzysztof Waleron and Maurycy Daroch

389 *The Isolation and Characterization of Rare Mycobiome Associated With Spacecraft Assembly Cleanrooms*
Adriana Blachowicz, Snehit Mhatre, Nitin Kumar Singh, Jason M. Wood, Ceth W. Parker, Cynthia Ly, Daniel Butler, Christopher E. Mason and Kasthuri Venkateswaran

410 *Establishment of Genome Based Criteria for Classification of the Family Desulfovibrionaceae and Proposal of Two Novel Genera,* Alkalidesulfovibrio *gen. nov. and* Salidesulfovibrio *gen. nov.*
Mi-Jeong Park, Yun Jae Kim, Myeongkyu Park, Jihyun Yu, Teddy Namirimu, Yoo-Rim Roh and Kae Kyoung Kwon

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 02 August 2022
DOI 10.3389/fmicb.2022.984632

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Andreas Teske,
University of North Carolina at Chapel Hill, United States

*CORRESPONDENCE
Rafael R. de la Haba
rrh@us.es
André Antunes
aglantunes@must.edu.mo
Brian P. Hedlund
brian.hedlund@unlv.edu

SPECIALTY SECTION
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 02 July 2022
ACCEPTED 15 July 2022
PUBLISHED 02 August 2022

CITATION
de la Haba RR, Antunes A and Hedlund BP (2022) Editorial: Extremophiles: Microbial genomics and taxogenomics. *Front. Microbiol.* 13:984632. doi: 10.3389/fmicb.2022.984632

COPYRIGHT
© 2022 de la Haba, Antunes and Hedlund. 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.

Editorial: Extremophiles: Microbial genomics and taxogenomics

Rafael R. de la Haba[1]*, André Antunes[2,3]* and Brian P. Hedlund[4,5]*

[1]Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [2]State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology, Taipa, Macau SAR, China, [3]China National Space Administration (CNSA), Macau Center for Space Exploration and Science, Macau, Macau SAR, China, [4]School of Life Sciences, University of Nevada, Las Vegas, NV, United States, [5]Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, NV, United States

KEYWORDS

extremophiles, taxonomy, genomics, metagenomics, adaptation, Space Microbiology, extremozymes, biodiversity

Editorial on the Research Topic
Extremophiles: Microbial genomics and taxogenomics

Introduction

Extreme habitats exist across the globe and account for most of our planet's habitable zone by volume (Gold, 1992; Charette and Smith, 2010). They vary widely from a physical-chemical perspective as they include diverse types of extremes, such as temperature, pH, salinity, radiation, pressure, water activity, and nutrient availability. Organisms that thrive under conditions that are adverse or lethal for most organisms are called "extremophiles" (Brock, 1969).

All three domains of life are represented in each type of extreme environment. However, the vast majority of extremophiles are bacteria and archaea (Rampelotto, 2013), which is not surprising considering their remarkable diversity and adaptability. The discovery, study, and classification of novel microbial extremophiles is of enormous interest, given their potential biotechnological applications (Corral et al., 2019) and the ever-increasing expectation to find life on other planets, where conditions might be similar to those of extreme environments on Earth (Gómez et al., 2012).

Our understanding of extremophiles on Earth is rooted in studies of pure cultures, but has greatly expanded over the last few decades through both intellectual and technological advancements in molecular ecology, including several next-generation DNA sequencing and bioinformatics applications, which have evolved rapidly over the last decade. Recent developments increasing the fidelity and lowering the cost of both short- and long-read sequencing technologies hold promise for dense, accurate, and inexpensive DNA sequencing studies focused on both microbial pure cultures and

microbial communities (Slatko et al., 2018). In parallel, increases in computing power per Moore's law (Moore, 1965) and the development of new bioinformatic tools, machine learning algorithms, and data mining approaches now provide the necessary means to deal with such a huge amount of data (Gauthier et al., 2019; Goodswen et al., 2021).

Extreme environments possess a larger than expected microbial diversity, considering the harshness of the conditions that extremophiles need to cope with (Shu and Huang, 2022). Many efforts have been made to characterize and describe prokaryotic and eukaryotic extremophiles and their metabolic functions within these habitats using culture-dependent approaches (de la Haba et al., 2011; Nogi, 2011; Yumoto et al., 2011; Sorokin et al., 2014; Ventosa et al., 2014; Deming, 2019; Johnson and Aguilera, 2019; Nienow, 2019; Santos and Antón, 2019; Topçuoglu and Holden, 2019; Sood et al., 2021). Nevertheless, estimates using 16S rRNA gene sequences and fixed similarity cutoffs suggest that around 80 % of microbial genera remain uncultured in non-human-associated environments (Lloyd et al., 2018). Similarly, analysis of shotgun metagenomic data estimated that yet-uncultivated taxa account for ∼85 % of the phylogenetic diversity of prokaryotes (Nayfach et al., 2021). Thus, most of the diversity and functions of extremophiles remain poorly understood. Fortunately, high-performance DNA sequencing and computation-based analysis now allow the routine recovery of high-quality metagenome-assembled genomes (MAGs) (Yang et al., 2021) and single-cell genomes (Rinke et al., 2013), in addition to genomes of cultured microorganisms (Whitman et al., 2015). These approaches finally allow us to describe high-quality genomes from extremophilic "microbial dark matter," which serve as a foundation to guide both field and lab studies of the ecology and physiology of these organisms within the context of the environments and communities they inhabit.

The use of genomic data of both cultured and uncultured extremophiles is also valuable to infer evolutionary relationships and to establish a robust and accurate classification system (Parks et al., 2022), and a formalized code of nomenclature based on genome sequences as nomenclatural types will soon be available (Murray et al., 2020). Most existing taxa of extremophiles were proposed based on classical but increasingly old-fashioned polyphasic studies that included 16S rRNA gene phylogenetic analyses and wet-lab DNA-DNA hybridization approaches, which have limited resolution and suffer from problems with accuracy and reproducibility among laboratories, respectively. These approaches have therefore been overtaken by phylogenomic studies and *in silico* genome relatedness indices (Chun and Rainey, 2014; Jain et al., 2018). Therefore, the current taxonomy of extremophiles needs to be assessed using genomic information. The ever-increasing access to genomic databases and the ever-decreasing cost of next-generation sequencing technologies and computing makes this a particularly timely Research Topic and grants us an exceptional opportunity to address this issue.

The present Research Topic includes 25 manuscripts spanning the discovery and characterization of new microbial extremophiles, the biotechnological applications of their enzymes, the mechanisms of adaptation to harsh environments, their metabolic pathways and potential ecological roles, their evolution, systematics and biodiversity, as well as exciting research on Space Microbiology and the persistence of extremophiles in low-biomass spacecraft assembly cleanrooms. The Research Topic call was well received, with a total of 185 contributing authors from 18 different countries around the world (Australia, China, Czech Republic, Denmark, Egypt, France, Germany, India, Japan, Mexico, Norway, Poland, Russia, Saudi Arabia, Spain, South Korea, United Kingdom, and United States), which reflects global interest on extremophiles. At the time of writing, over 150,000 views of this Research Topic have been recorded. The high impact of this Research Topic has lead us to establish a community series and make a second call for manuscripts to be submitted to Community Series-Extremophiles: Microbial Genomics and Taxogenomics, Volume II.

Discovery of new extremophiles and extremozymes

Considering that only a small fraction of extremophilic taxa possess a cultured representative, it has never been more important to encourage the scientific community to isolate and describe not-yet-cultured extremophiles. New high-throughput cultivation approaches, referred as "culturomics" (Lagier et al., 2012), coupled with focused, genome-guided cultivation efforts (Tyson et al., 2005; Buessecker et al., 2022), will broaden our knowledge about extremophiles, as a major complement to metagenomics. This collection of articles contributes to this aim with the description and characterization of new extremophilic isolates belonging to both Bacteria and Archaea, as well as a novel viral genus.

Three papers from the Duroch group describe new taxa of thermophilic cyanobacteria in the family *Leptolyngbiaceae*, including two new species, *Thermoleptolyngbya sichuanensis* (Tang et al.) and *Leptodesmis sichuanensis* (Tang, Du et al.), and a new genus and species, *Leptothermofonsia sichuanensis* (Tang, Shah et al.). As suggested by their species names, all three were isolated from hot springs in Sichuan Province, China; the morphological, physiological, genomic, and taxonomic data presented in this collection represent a significant expansion of the systematics of the family and of thermophilic cyanobacteria in general. The Research Topic also includes the proposal by Slobodkina et al. of a new member of the class *Archaeoglobi*, at that time represented by only eight validly published species names (Parte et al., 2020). The

new species *Archaeoglobus neptunius* was isolated from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge and possesses a lithoautotrophic mode of nutrition, although it is also able to grow chemoorganotrophically. This article points to polyphyletic origin of the species currently grouped into the genus *Archaeoglobus*, necessitating a future review of its taxonomic status.

In addition to thermophiles, new psychrotolerant bacterial species were proposed within this collection. The paper from Králová et al. describes two new psychrotolerant species of *Flavobacterium* isolated from Antarctica, namely, "*Flavobacterium flabelliforme*" and "*Flavobacterium geliluteum*", including three and five strains, respectively. As might be expected, both taxa harbor several cold-inducible or cold-adaptation related genes, but surprisingly they also accommodate numerous prophages and displayed a multidrug-resistant phenotype. Genome analysis revealed the presence of unknown biosynthetic gene clusters, which suggests a potential application to synthesize new biomolecules. Another paper combined high-throughput culturomics and mass-spectroscopy screening to successfully isolate only the second phototrophic member of the phylum *Gemmatimonadota*, "*Gemmatimonas groenlandica*", from a stream in Greenland (Zeng et al.). Although genes encoding the photosynthetic apparatus are closely related to that of *Gemmatimonas phototrophica* strain AP64^T, the new species is distinct based on average nucleic acid identity, pigment composition and absorption spectra, and high oxygen tolerance.

One significant driver of research in extremophiles is biotechnology, given their well-recognized roles as sources of new biomolecules and applications. Included in this volume, the review by Sysoev et al. provides an overview of culture-independent methods, including sequence-based metagenomics and single-cell genomics, for studying enzymes from extremophiles, with a focus on prokaryotes. Additionally, the authors provide a comprehensive list of extremozymes discovered *via* metagenomics and single-cell genomics, which is a useful resource for researchers both in academia and industry. Extremophilic viruses are also rich sources of commercially valuable enzymes, particularly nucleic acid-modifying enzymes. One such example is an unusual DNA polymerase A enzyme from an uncultivated, thermophilic virus that has been developed commercially as a high-affinity, thermophilic reverse-transcriptase (Schoenfeld et al., 2013). In this Research Topic, this *polA* gene and related genes were shown to be encoded by a new family and genus of thermophilic viruses, the genus *Pyrovirus*, that is inferred to infect diverse members of the *Aquificota* based on spacers within clustered regularly interspaced short palindromic repeats (Palmer et al.). The four species of *Pyrovirus* proposed based on complete or near-complete genomes is the first study of viruses infecting *Aquificota*, despite the dominance of *Aquificota* in many terrestrial and marine thermal systems.

Adaptation to extreme conditions, environmental role and metabolic functions of extremophiles

Mechanisms evolved by extremophiles not just to survive, but to thrive in inhospitable environments are fascinating. Many specialized mechanisms enabling growth under extreme conditions have already been studied while others remain elusive (Coker, 2019; Schmid et al., 2020). Some of the manuscripts in this volume expand on such strategies.

Three papers in the Research Topic focused on energy conservation, central carbon metabolism, and piezophily of thermophiles. The paper by Gavrilov et al. focused on mechanisms used by the thermophile *Carboxydothermus ferrireducens* to use extracellular amorphous ferrihydrite as a terminal electron acceptor for anaerobic respiration, producing large magnetite crystals. Genome sequencing combined with RNA-Seq and proteomic data derived from cultures using different electron acceptors were used to identify three constitutive c-type multiheme cytochromes, including OhmA, which was strongly bound to extracellular magnetite. Another paper by Thomas et al., focused on carbon metabolism in the thermophile *Thermoflexus hugenholtzii* JAD2^T, the only cultured representative of the *Chloroflexota* order *Thermoflexales*. Despite its chemoheterotrophic metabolism, *T. hugenholtzii* does not grow on any defined carbon sources, obscuring our understanding of its metabolism. Comparative genomics of the *T. hugenholtzii* genome with eight related MAGs from geothermal springs revealed a high abundance and strong conservation of peptidases among three species clusters, suggesting a conserved proteolytic metabolism. Exometabolomics and ^{13}C metabolic probing studies confirmed this metabolism for *T. hugenholtzii.* The metabolic probing confirmed glycolysis, tricarboxylic acid (TCA), and oxidative pentose-phosphate pathways, yet glycolysis and the TCA cycle were uncoupled. Microorganisms living at great depths in the sea must cope with variations in hydrostatic pressure. The classical stress response of piezophiles to changes in hydrostatic pressure consists of up- and down-regulation of several genes. Herein, Moalic et al. suggest that *Thermococcus piezophilus*, a piezohyperthermophilic archaeon with the widest range of tolerance to high pressure known so far, modulates its energy-conservation process by means of the control of the master transcriptional regulator SurR under non-optimal pressures conditions.

Two additional contributions focused on acidophily and halophily. Genome comparisons by Cortez et al. on bacterial acidophiles affiliated with different phyla and their neutrophilic counterparts demonstrated a correlation between decreases in genome size and adaptation to low-pH ecosystems. This genome streamlining of acidophilic bacteria is mainly due to reduction of average protein size and gene loss, with an unexpected constant number of paralogs compared with studies that claim a relatively

lower paralog frequency for other streamlined microorganisms (Giovannoni et al., 2005; Swan et al., 2013). According to these authors, other potential mechanisms for acid resistance might be the enrichment in cytoplasmic membrane proteins and those involved in energy conservation, DNA repair, and biofilm formation. The paper by Durán-Viseras et al. uses a culturomics approach to isolate new haloarchaea from the Odiel Saltmarsh, Spain, previously revealed to contain a high proportion of sequences not related to previously cultivated taxa (Vera-Gargallo and Ventosa, 2018). Their approach resulted in the isolation and description of three new species: *Halomicroarcula rubra*, *Halomicroarcula nitratireducens*, and *Halomicroarcula salinisoli*. Ensuing genomic analysis revealed the unexpected identification of complete pathways for the biosynthesis of the compatible solutes trehalose and glycine betaine, not identified before in any other haloarchaea. Although they use a salt-in strategy, this finding might suggest alternative osmoadaptation strategies within this group.

Phylogeny, evolution, classification, and biodiversity of extremophiles

A heated debate exists on the conditions of the environment where the last universal common ancestor (LUCA) of extant life could have inhabited. Although it is not yet clear if the LUCA was an extremophile (i.e., hyperthermophile), the scientific community agrees that extremophily emerged early in the diversification of prokaryotes (Catchpole and Forterre, 2019). Evolution of the thermoacidophilic archaeal family *Sulfolobaceae* has been a focus of research into this topic. The study by Banerjee et al. revealed a non-symmetric genome evolution within this family, with some species undergoing genome expansion whereas there was gene loss in others, even considering that they share a similar niche. They also detailed a high level of conservation for most of the autotropic pathways across this family.

In the genomic era, systematics has moved forward from gene-based phylogenies to genome-based classifications (Rosselló-Móra and Whitman, 2019; Parks et al., 2022). Thus, taxonomic arrangements supported by whole genome data analysis spanning several groups of extremophiles are proposed through this volume. The research conducted by de la Haba et al. sought to solve a well-known issue concerning the taxonomic status of the partially overlapping and contemporaneously described archaeal genera *Natrinema* and *Haloterrigena* (Minegishi et al., 2010; Papke et al., 2011). Phylogenomic analysis and the use of Overall Genome Relatedness Indexes allowed the authors to clearly differentiate the two genera, and justified transfer several species from *Haloterrigena* to *Natrinema*. In the domain *Bacteria*, the article by Park et al. proposed an average amino-acid identity (AAI) cut-off value of 63.43 ± 0.01% to delineate genera within the *Desulfovibrionaceae*, following up on recent rearrangements in the family based on a phylogenomic approach (Galushko and Kuever, 2020; Waite et al., 2020). As a result, two new genera (i.e., *"Alkalidesulfovibrio"* and *"Salidesulfovibrio"*), and one not yet validly published genus name ("*Psychrodesulfovibrio*") were proposed. On the other hand, Ramírez-Duran et al. revised the taxonomy of the genus *Saccharomonospora* using a combined approach based on 16S rRNA gene and core-genome sequence phylogenies as well as comparative genomics. Additionally, this article also identifies a wide variety of biosynthetic gene clusters, thus uncovering the potential biotechnological applications of this actinobacterial genus to produce unknown bioactive molecules.

The exploration of natural microbial communities inhabiting extreme environments was the focus of several studies in this Research Topic. These types of studies remain vibrant because they provide insights into the ecology of microorganisms in their natural habitats. As a first example, Liu et al. explored hot springs in Conghua, China, focusing on anaerobic ammonium oxidation (anammox). This is an important process of the nitrogen cycle, and anammox bacteria have been studied in a wide variety of environments. However, the distribution, diversity, and abundance of anammox bacteria in hot springs remains under-reported. Diverse putative anammox organisms were identified, especially "*Candidatus* Brocadia". In another contribution, Narsing Rao et al. combined culture-dependent and -independent approaches to study seven different hot spring in India. These locations were found to harbor novel microbial groups and some of which produced thermo-stable enzymes. Finally, a separate study focused on benthic microbial community dynamics in an acid saline lake, Lake Magic, Australia, by studying 16S rRNA gene amplicons for prokaryotes and internal transcribed spacer amplicons for fungi over an annual cycle (Ghori et al.). The annual cycle included a flooding stage, followed by evapo-concentration of solutes, which led to decreases in bacterial diversity and increases in halotolerant and acid-tolerant taxa, including phylotypes similar to those involved in sulfur cycling.

New frontiers

Research on extremophiles has traditionally been at the forefront of microbiology, boldly opening up new frontiers in our understanding of life and its limits. The reach and impact of such research is now extending beyond our own planet, with increasing relevance in the cross-disciplinary fields of Astrobiology and Space Microbiology. The study of extreme environments on Earth and their inhabitants is one of the main pillars of research supporting the search for life on Mars, the exooceans of the icy moons of the solar system, and beyond (e.g., Jebbar et al., 2020; Taubner et al., 2020; Changela et al., 2021). Studies of terran extremophiles is also increasingly seen

as useful for assisting space exploration in activities such as *in situ* resource utilization (ISRU; Cockell, 2010; Dhami et al., 2013; Changela et al., 2021).

Space itself can be seen as an extreme environment, with a combination of multiple conditions and with a range of effects on microbes that are not yet fully explored. Observed changes in the characteristics of pathogens and conditional pathogens in space, namely increased pathogenicity and virulence, are a significant concern when discussing long-term crewed mission or human presence in space (Simões and Antunes, 2021). The impact of the space environment seems to vary between species, so the need for further studies has been stressed by several authors.

The current volume of our special topic addresses this issue with a study by Su et al., focusing on the conditional pathogen *Stenotrophomonas maltophilia*, an understudied emerging Gram-stain-negative multidrug-resistant species. This paper constitutes the first general analysis of the phenotypic, genomic, transcriptomic, and proteomic changes in this species under simulated microgravity. Results from this investigation showed that microgravity led to an increased growth rate, enhanced biofilm formation, increased swimming motility, and metabolic alterations. This study further increases our understanding of the effects of space conditions on microbial physiology, metabolism, and pathogenicity and may help to provide new ideas for the prevention and treatment of infections in future missions.

Bijlani et al. describe in this article collection the new species *Methylobacterium ajmalii*, isolated from the International Space Station as part of an on-going microbial tracking experiment that retrieved a few strains from the family *Methylobacteriaceae* (Checinska Sielaff et al., 2019). In addition to the full characterization and description of this new species of *Methylobacterium* using polyphasic taxonomy, the paper further analyses whole genome sequencing data with identification of some specific features with potential relevance for biotechnology.

Another important link between microbiology and space exploration is the study of persistent species in spacecraft assembly cleanrooms. Despite their inhospitable conditions and intense precautions taken within these facilities, different types of microbes are known to persist and are a major source of concern regarding planetary protection measures focused on preventing potential biological contamination of other parts of the solar system (e.g., Rettberg et al., 2019). Within such studies, fungi are frequently overlooked despite the fact that they are known for their resilience and might pose a threat to closed habitats (biocorrosion) as well as their immunocompromised occupants (pathogenicity).

In this sense, Blachowicz et al. deal with isolation and characterization of rare mycobiome associated with spacecraft assembly cleanrooms. The study aimed to help address two current gaps in understanding fungi populations under such settings. On one hand, the majority of previous reports were based on cultivation-based approaches, while culture-independent mycobiome analyses were scarce. On the other, the authors highlighted the need for more efficient cultivation methods when analyzing cleanrooms, particularly as isolation is essential for thorough analysis and characterization of fungi. Metagenomic reads were dominated by *Ascomycota* and *Basidiomycota*, with significant differences between distant locations, and the detection of several potential novel species. This paper is presented as a first step toward characterizing cultivable and viable fungal populations in cleanrooms to assess fungal potential as biocontaminants during interplanetary exploration, in addition to other cleanroom settings, such as intensive care units, operating rooms, or in the semiconducting and pharmaceutical industries.

Schultzhaus et al. also focused on fungi with an eye toward extreme conditions relevant to space exploration, namely ionizing radiation. By studying the phenotypic and transcriptomic responses of the radioresistant yeast *Exophiala dermatitidis* to three different radiation sources (i.e., protons, deuterons, and α-particles), a common theme of induction of DNA repair and DNA replication genes and repression of ribosome and protein synthesis emerged. Yet, differences also emerged, most notably that particle irradiation resulted in greater changes in gene expression and an upregulation of genes involved in autophagy and protein catabolism.

This collection contains a manuscript by Wood et al. that highlights concerns about inconsistent results of different short-read metagenomic pipelines to assess the microbial diversity of the aforementioned cleanrooms. Although all four used pipelines detected extremophiles and spore-forming bacteria, they did not yield the same microbial profile based on the same raw dataset, which raises doubts of the reliability of using a single pipeline. The paper also presents a roadmap aimed at validating future studies on planetary protection-related microorganisms.

Final remarks

As evidenced by the success of this Research Topic, after over 50 years of research focused on extreme environments, extremophiles are still the subject of enormous interest for the scientific community and for biotechnology. The more we learn about extremophiles, the more fascinating they become, and each new discovery reveals more of their secrets, but also opens up new questions. After all, they are one of the most ancient dwellers on Earth and they will certainly remain here, and possibly beyond our planet, once all of us are gone.

Author contributions

RH, AA, and BH wrote the first draft of the editorial article, and they also reviewed and approved the final version of the manuscript.

Acknowledgments

We thank all contributing authors for submission of their articles to this Research Topic. We are grateful to reviewers for the valuable comments on the manuscript.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Brock, T. D. (1969). Microbial growth under extreme conditions. *Symp. Soc. Gen. Microbiol.* 19, 15–41.

Buessecker, S., Palmer, M., Lai, D., Dimapilis, J., Mayali, X., Mosier, D., et al. (2022). An essential role for tungsten in the ecology and evolution of a novel lineage of anaerobic, thermophilic Archaea. *Nat. Commun.* 13, 3773. doi: 10.1038/s41467-022-31452-8

Catchpole, R. J., and Forterre, P. (2019). The evolution of reverse gyrase suggests a nonhyperthermophilic Last Universal Common Ancestor. *Mol. Biol. Evol.* 36, 2737–2747. doi: 10.1093/molbev/msz180

Changela, H., Chatzitheodoridis, E., Antunes, A., Beaty, D., Bouw, K., Bridges, J. C., et al. (2021). Mars: new insights and unresolved questions. *Int. J. Astrobiol.* 20, 394–426. doi: 10.1017/S1473550421000276

Charette, M. A., and Smith, W. H. F. (2010). *Oceanography* 23, 112–114. doi: 10.5670/oceanog.2010.51

Checinska Sielaff, A., Urbaniak, C., Mohan, G. B. M., Stepanov, V. G., Tran, Q., Wood, J. M., et al. (2019). Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. *Microbiome* 7, 50. doi: 10.1186/s40168-019-0666-x

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

Cockell, C. S. (2010). Geomicrobiology beyond Earth: microbe–mineral interactions in space exploration and settlement. *Trends Microbiol.* 18, 308–314. doi: 10.1016/j.tim.2010.03.005

Coker, J. A. (2019). Recent advances in understanding extremophiles. *F1000Res* 8, 1917. doi: 10.12688/f1000research.20765.1

Corral, P., Amoozegar, M. A., and Ventosa, A. (2019). Halophiles and their biomolecules: recent advances and future applications in biomedicine. *Mar. Drugs* 18, 33. doi: 10.3390/md18010033

de la Haba, R. R., Sánchez-Porro, C., Marquez, M. C., and Ventosa, A. (2011). "Taxonomy of halophiles," in *Extremophiles Handbook*, ed K. Horikoshi (Tokyo: Springer), 255–308. doi: 10.1007/978-4-431-53898-1_13

Deming, J. W. (2019). "Extremophiles: cold environments," in *Encyclopedia of Microbiology (Fourth Edition)*, ed T. M. Schmidt (Cambridge, MA: Academic Press), 228–238. doi: 10.1016/B978-0-12-809633-8.13045-6

Dhami, N. K., Reddy, M. S., and Mukherjee, A. (2013). Biomineralization of calcium carbonates and their engineered applications: a review. *Front. Microbiol.* 4, 314. doi: 10.3389/fmicb.2013.00314

Galushko, A., and Kuever, J. (2020). "*Desulfovibrionaceae*," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed W. B. Whitman (Hoboken, NJ: John Wiley and Sons, Inc.), 1–13. doi: 10.1002/9781118960608.fbm00199.pub2

Gauthier, J., Vincent, A. T., Charette, S. J., and Derome, N. (2019). A brief history of bioinformatics. *Brief. Bioinform.* 20, 1981–1996. doi: 10.1093/bib/bby063

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309, 1242–1245. doi: 10.1126/science.1114057

Gold, T. (1992). The deep, hot biosphere. *Proc. Natl. Acad. Sci. U. S. A.* 89, 6045–6049. doi: 10.1073/pnas.89.13.6045

Gómez, F., Barták, M., and Bell, E. M. (2012). "Extreme environments on Earth as analogues for life on other planets: astrobiology," in *Life at extremes: environments, organisms and strategies for survival*, ed E. M. Bell (Wallingford: CABI Publishing), 522–536. doi: 10.1079/9781845938147.0522

Goodswen, S. J., Barratt, J. L. N., Kennedy, P. J., Kaufer, A., Calarco, L., and Ellis, J. T. (2021). Machine learning and applications in microbiology. *FEMS Microbiol. Rev.* 45, fuab015. doi: 10.1093/femsre/fuab015

Jain, C., Rodriguez,-R., L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 5114. doi: 10.1038/s41467-018-07641-9

Jebbar, M., Hickman-Lewis, K., Cavalazzi, B., Taubner, R. S., Simon, K. M. R., and Antunes, A. (2020). Microbial diversity and biosignatures: an icy moons perspective. *Space Sci. Rev.* 216, 10. doi: 10.1007/s11214-019-0620-z

Johnson, D. B., and Aguilera, A. (2019). "Extremophiles and acidic environments," in *Encyclopedia of Microbiology (Fourth Edition)*, ed T. M. Schmidt (Cambridge, MA: Academic Press), 206–227. doi: 10.1016/B978-0-12-809633-8.90687-3

Lagier, J. C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., et al. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18, 1185–1193. doi: 10.1111/1469-0691.12023

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. *mSystems* 3, e00055–e00018. doi: 10.1128/mSystems.00055-18

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

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics* 38, 114–117.

Murray, A. E., Freudenstein, J., Gribaldo, S., Hatzenpichler, R., Hugenholtz, P., Kampfer, P., et al. (2020). Roadmap for naming uncultivated archaea and bacteria. *Nat. Microbiol.* 5, 987–994. doi: 10.1038/s41564-020-0733-x

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., et al. (2021). A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* 39, 499–509. doi: 10.1038/s41587-020-0718-6

Nienow, J. A. (2019). "Extremophiles: dry environments (including cryptoendoliths)," in *Encyclopedia of Microbiology (Fourth Edition)*, ed T. M. Schmidt (Cambridge, MA: Academic Press), 239–262. doi: 10.1016/B978-0-12-809633-8.13046-8

Nogi, Y. (2011). "Taxonomy of psychrophiles," in *Extremophiles Handbook*, ed K. Horikoshi (Tokyo: Springer), 777–792. doi: 10.1007/978-4-431-53898-1_38

Papke, R. T., White, E., Reddy, P., Weigel, G., Kamekura, M., Minegishi, H., et al. (2011). A multilocus sequence analysis approach to the phylogeny and taxonomy of the *Halobacteriales*. *Int. J. Syst. Evol. Microbiol.* 61, 2984–2995. doi: 10.1099/ijs.0.029298-0

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P. A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794. doi: 10.1093/nar/gkab776

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Rampelotto, P. H. (2013). Extremophiles and extreme environtments. *Life* 3, 482–485. doi: 10.3390/life3030482

Rettberg, P., Antunes, A., Brucato, J., Cabezas, P., Collins, G., Haddaji, A., et al. (2019). Biological contamination prevention for outer solar system moons of astrobiological interest: what do we need to know? *Astrobiology* 19, 951–974. doi: 10.1089/ast.2018.1996

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352

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

Santos, F., and Antón, J. (2019). "Extremophiles: hypersaline environments," in *Encyclopedia of Microbiology (Fourth Edition)*, ed T. M. Schmidt (Cambridge, MA: Academic Press), 270–275. doi: 10.1016/B978-0-12-809633-8.90673-3

Schmid, A. K., Allers, T., and DiRuggiero, J. (2020). SnapShot: microbial extremophiles. *Cell* 180, 818–818.e1. doi: 10.1016/j.cell.2020.01.018

Schoenfeld, T. W., Murugapiran, S. K., Dodsworth, J. A., Floyd, S., Lodes, M., Mead, D. A., et al. (2013). Lateral gene transfer of family A DNA polymerases between thermophilic viruses, *Aquificae*, and *Apicomplexa*. *Mol. Biol. Evol.* 30, 1653–1664. doi: 10.1093/molbev/mst078

Shu, W. S., and Huang, L. N. (2022). Microbial diversity in extreme environments. *Nat. Rev. Microbiol.* 20, 219–235. doi: 10.1038/s41579-021-00648-y

Simões, M. F., and Antunes, A. (2021). Microbial pathogenicity in space. *Pathogens* 10, 450. doi: 10.3390/pathogens10040450

Slatko, B. E., Gardner, A. F., and Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* 122, e59. doi: 10.1002/cpmb.59

Sood, U., Kumar, R., and Hira, P. (2021). Expanding culturomics from gut to extreme environmental settings. *mSystems* 6, e0084821. doi: 10.1128/mSystems.00848-21

Sorokin, D. Y., Berben, T., Melton, E. D., Overmars, L., Vavourakis, C. D., and Muyzer, G. (2014). Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18, 791–809. doi: 10.1007/s00792-014-0670-9

Swan, B. K., Tupper, B., Sczyrba, A., Lauro, F. M., Martinez-Garcia, M., González, J. M., et al. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci. U. S. A.* 110, 11463–11468. doi: 10.1073/pnas.1304246110

Taubner, R. S., Olsson-Francis, K., Vance, S. D., Ramkissoon, N. K., Postberg, F., de Vera, J. P., et al. (2020). Experimental and simulation efforts in the astrobiological exploration of exooceans. *Space Sci. Rev.* 216, 9. doi: 10.1007/s11214-020-0635-5

Topçuoglu, B. D., and Holden, J. F. (2019). "Extremophiles: hot environments," in *Encyclopedia of Microbiology (Fourth Edition)*, ed T. M. Schmidt (Cambridge, MA: Academic Press), 263–269. doi: 10.1016/B978-0-12-809633-8.90683-6

Tyson, G. W., Lo, I., Baker, B. J., Allen, E. E., Hugenholtz, P., and Banfield, J. F. (2005). Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* 71, 6319–6324. doi: 10.1128/AEM.71.10.6319-6324.2005

Ventosa, A., Fernández, A. B., León, M. J., Sánchez-Porro, C., and Rodriguez-Valera, F. (2014). The Santa Pola saltern as a model for studying the microbiota of hypersaline environments. *Extremophiles* 18, 811–824. doi: 10.1007/s00792-014-0681-6

Vera-Gargallo, B., and Ventosa, A. (2018). Metagenomic insights into the phylogenetic and metabolic diversity of the prokaryotic community dwelling in hypersaline soils from the Odiel Saltmarshes (SW Spain). *Genes* 9, 152. doi: 10.3390/genes9030152

Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., et al. (2020). Proposal to reclassify the proteobacterial classes *Deltaproteobacteria* and *Oligoflexia*, and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* 70, 5972–6016. doi: 10.1099/ijsem.0.004213

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

Yang, C., Chowdhury, D., Zhang, Z., Cheung, W. K., Lu, A., Bian, Z., et al. (2021). A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Comput. Struct. Biotechnol. J.* 19, 6301–6314. doi: 10.1016/j.csbj.2021.11.028

Yumoto, I., Hirota, K., and Yoshimune, K. (2011). "Environmental distribution and taxonomic diversity of alkaliphiles," in *Extremophiles Handbook*, ed K. Horikoshi (Tokyo: Springer), 55–79. doi: 10.1007/978-4-431-53898-1_4

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 November 2020
doi: 10.3389/fmicb.2020.583361

Diversity and Distribution of a Novel Genus of Hyperthermophilic *Aquificae* Viruses Encoding a Proof-Reading Family-A DNA Polymerase

Marike Palmer[1], Brian P. Hedlund[1,2]*, Simon Roux[3], Philippos K. Tsourkas[1,2], Ryan K. Doss[1], Casey Stamereilers[1], Astha Mehta[1], Jeremy A. Dodsworth[4], Michael Lodes[5], Scott Monsma[5], Tijana Glavina del Rio[3], Thomas W. Schoenfeld[6], Emiley A. Eloe-Fadrosh[3] and David A. Mead[7]**

[1] School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, NV, United States, [2] Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, Las Vegas, NV, United States, [3] Department of Energy Joint Genome Institute, Berkeley, CA, United States, [4] Department of Biology, California State University, San Bernardino, CA, United States, [5] Lucigen Corporation, Middleton, WI, United States, [6] Tamarack Bioscience, Inc., Beverly, MA, United States, [7] Varigen Biosciences Corporation, Madison, WI, United States

OPEN ACCESS

Edited by:
Richard Allen White III,
University of North Carolina at Charlotte, United States

Reviewed by:
Lin-Xing Chen,
University of California, Berkeley, United States
Julia M. Brown,
Bigelow Laboratory For Ocean Sciences, United States
Robert Reid,
University of North Carolina at Charlotte, United States

***Correspondence:**
Marike Palmer
marike.palmer@unlv.edu
Brian P. Hedlund
brian.hedlund@unlv.edu
David A. Mead
dmead@varigenbio.com

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

Received: 14 July 2020
Accepted: 16 October 2020
Published: 12 November 2020

Citation:
Palmer M, Hedlund BP, Roux S, Tsourkas PK, Doss RK, Stamereilers C, Mehta A, Dodsworth JA, Lodes M, Monsma S, Glavina del Rio T, Schoenfeld TW, Eloe-Fadrosh EA and Mead DA (2020) Diversity and Distribution of a Novel Genus of Hyperthermophilic Aquificae Viruses Encoding a Proof-Reading Family-A DNA Polymerase. Front. Microbiol. 11:583361. doi: 10.3389/fmicb.2020.583361

Despite the high abundance of *Aquificae* in many geothermal systems, these bacteria are difficult to culture and no viruses infecting members of this phylum have been isolated. Here, we describe the complete, circular dsDNA Uncultivated Virus Genome (UViG) of *Thermocrinis* Octopus Spring virus (TOSV), derived from metagenomic data, along with eight related UViGs representing three additional viral species. Despite low overall similarity among viruses from different hot springs, the genomes shared a high degree of synteny, and encoded numerous genes for nucleotide metabolism, including a PolA-type DNA polymerase polyprotein with likely accessory functions, a DNA Pol III sliding clamp, a thymidylate kinase, a DNA gyrase, a helicase, and a DNA methylase. Also present were conserved genes predicted to code for phage capsid, large and small subunits of terminase, portal protein, holin, and lytic transglycosylase, all consistent with a distant relatedness to cultivated *Caudovirales*. These viruses are predicted to infect *Aquificae*, as multiple CRISPR spacers matching the viral genomes were identified within the genomes and metagenomic contigs from these bacteria. Based on the predicted atypical bi-directional replication strategy, low sequence similarity to known viral genomes, and unique position in gene-sharing networks, we propose a new putative genus, "*Pyrovirus,*" in the order *Caudovirales*.

Keywords: viral metagenomics, bacteriophage, uncultivated virus genome, UViG, geothermal spring, DNA polymerase

INTRODUCTION

Viruses are the most abundant biological entities on Earth and are important drivers of genetic exchange, secondary production, and host metabolism on both local and global scales (Fuhrman, 1999; Suttle, 2007; Rohwer and Thurber, 2009; Breitbart et al., 2018). They also possess a high density of nucleic acid synthesis and modifying enzymes that are important sources of

enzymes for the biotechnology sector. Despite their importance, cultivation of viruses in the laboratory is limited, partly by challenges associated with cultivating their hosts. This problem is particularly true for viruses of thermophiles and hyperthermophiles because many hosts remain uncultured (Hedlund et al., 2015; Lloyd et al., 2018). Also, most thermophiles do not readily form lawns on solid media, which are typically exploited to screen for viruses. Although direct observation of filtrates from geothermal springs and enrichments has revealed a high diversity of virus morphotypes (Rice et al., 2001; Rachel et al., 2002), few thermophilic viruses have been studied in enrichment cultures and even fewer have been isolated in culture with their host. Currently, the NCBI Viral Genomes database lists 59 thermophilic archaeal viruses out of 95 total genomes, representing ten families (accessed 2/3/20)[1]; however, 49 of these infect members of the thermoacidophilic family *Sulfolobaceae*, leaving other archaeal thermophiles vastly under-explored. Similarly, only 15 of the 2,500 bacteriophage genomes represent thermophilic or hyperthermophilic viruses, representing only three virus families (accessed 2/3/20)[2]. Strikingly, although members of the phylum *Aquificae* (syn. *Aquificota*) predominate in many terrestrial and marine high-temperature ecosystems (Reysenbach et al., 2005; Spear et al., 2005; Takacs-Vesbach et al., 2013), to date, no cultivated viruses infecting *Aquificae* have been described.

Microbial ecologists have increasingly turned to cultivation-independent approaches to probe microbial diversity in nature. Although the low nucleic acid content and lack of universal conserved marker genes slowed the development of viral metagenomics, this field is now in full swing (Paez-Espino et al., 2016; Emerson et al., 2018; Koonin and Dolja, 2018). One of the early viral metagenomic investigations focused on Octopus Spring and other circumneutral pH springs in Yellowstone National Park (Schoenfeld et al., 2008), revealing 59 putative DNA polymerase (*pol*) genes, which were subsequently screened for heterologous activity in *Escherichia coli* (Moser et al., 2012). The most thermophilic of these enzymes, 3173 PolA, also demonstrated high-fidelity, thermostable reverse-transcriptase (RT) activity and strand-displacement activity and was subsequently marketed by Lucigen Corporation as a single-enzyme RT-PCR system called PyroPhage and RapidDxFire. That enzyme was further improved by molecular evolution and fusion of a high-performance chimeric variant of 3173 PolA with the 5′ to 3′ exonuclease domain of *Taq* polymerase to improve probe-based detection chemistries and enable highly sensitive detection of RNA (Heller et al., 2019).

A study of the diversity and evolution of 3173 PolA and related polymerases revealed clues about its complex evolutionary history (Schoenfeld et al., 2013). In addition to their discovery in viral metagenomes from hot springs, 3173 *polA*-like genes were also detected in two of the three families of *Aquificae*, where they have orthologously replaced host DNA *polA* genes, and phylogenetically diverse, non-thermophilic bacteria, where they appear to be transient alternative *polA* genes, presumably due to recombination following non-productive infections. Amazingly, 3173 *polA*-like genes are also known to encode thermophilic, nuclear-encoded, apicoplast-targeted polymerases in eukaryotic parasites in the *Apicomplexa* (e.g., *Plasmodium*, *Babesia*, and *Toxoplasma*) (Seow et al., 2005). The origin of these genes likely involved fixation of a progenitor sequence into the nuclear genome following endosymbiosis of a red alga (proto-apicoplast) containing a bacterial symbiont carrying a viral *polA* gene (Schoenfeld et al., 2013).

Recently, an Uncultivated Virus Genome (UViG) containing the 3173 *polA* gene was described from metagenomic data (Mead et al., 2017) with preliminary comparisons to a limited number of related genomes, although the sequence data were not published. This study comprehensively interrogated new and existing metagenomes from terrestrial geothermal springs with the goal of obtaining additional UViGs related to the previously identified viral genome and to address several outstanding questions regarding this group of viruses: (i) what is their genomic diversity and what features are conserved and variable?, (ii) what is their environmental distribution?, (iii) what are their cellular hosts, and (iv) how are they related to other viruses? We uncovered nearly complete UViGs from several Yellowstone geothermal springs and Great Boiling Spring (GBS), Nevada that range from 37,256 to 41,208 bp and encode 48 to 53 open reading frames (ORFs). The presence of fragments of these viral genomes in CRISPR arrays encoded by *Thermocrinis ruber* OC1/4^T, *Thermocrinis jamiesonii* GBS1^T, *Hydrogenobaculum* sp. 3684, and *Sulfurihydrogenibium yellowstonense* SS-5^T genomes, along with similarity between many viral genes and *Aquificaceae* genes, supports the previous hypothesis (Schoenfeld et al., 2013; Mead et al., 2017) that *Thermocrinis* and probably other *Aquificae* are putative hosts of these viruses. The high abundance of these viruses and their hosts suggests they may play an important role in chemolithotrophic productivity in geothermal springs globally, in addition to their role in evolution as a vector for horizontal gene transfer.

MATERIALS AND METHODS

Isolation of Uncultured Viral Particles From Octopus Hot Spring and Great Boiling Spring

Virus particles were isolated from Octopus Hot Spring in Yellowstone National Park (Permit # YELL-2007-SCI-5240), Wyoming (N 44.5342, W 110.79812) in 2007 and from Great Boiling Spring (GBS), Nevada, (N 44.6614, W 119.36622) in December 2008. Temperature at the time and location of sampling was 87°C at the outflow channel of Octopus Spring and 80–83°C in the source pool of Great Boiling Spring.

For Octopus Spring samples, thermal water (between 200 and 630 L) was filtered using a 100 kDa molecular weight cut-off (mwco) tangential flow filter (A/G Technology, Amersham Biosciences, GE Healthcare) and viruses and cells were

[1]http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239&host=archaea

[2]http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239&host=bacteria

concentrated to about 2 L. The resulting concentrates were filtered through a 0.2 μm tangential flow filter to remove microbial cells. The viral fractions were further concentrated to about 100 mL using a 100 kDa tangential flow filter and 40 mL of viruses were further concentrated to 400 μL and transferred to SM buffer (0.1 M NaCl, 8 mM $MgSO_4$, 50 mM Tris HCl, pH 7.5) by filtration in a 30 kDa mwco spin filter (Centricon, Millipore).

For the GBS viral sample, tangential-flow filtration with a 30 kDa molecular weight cutoff Millipore Prep/Scale TFF-6 filter (catalog # CDUF006TT) was used to concentrate ∼500 L of GBS water to ∼2 L. Filtration was done in December 2008 with water from the GBS "A" site (Cole et al., 2013) with a temperature of 80–83°C and pH of 7.15-7.2. The concentrated sample was stored on ice and transported to the laboratory, where it was pelleted by centrifugation at 4°C for 10 min at 10,000 × *g*. The supernatant was then further concentrated as above, and the cell pellet was stored at −80°C for DNA extraction.

Isolation of Viral and Planktonic Cell DNA

Serratia marcescens endonuclease (Sigma, 10 U) was added to both viral preparations described above to remove non-encapsidated (non-viral) DNA. The reactions were incubated at 23°C for 2 h. EDTA (20 mM), sodium dodecyl sulfate (SDS) (0.5%), and Proteinase K (100 U) were added and the reactions were incubated at 56°C. Subsequently, sodium chloride (0.7 M) and cetyltrimethylammonium bromide (CTAB) (1%) were added. The DNA was then extracted with chloroform, precipitated with isopropanol and washed with 70% ethanol. Yields of DNA ranged from 20 to 200 ng.

For preparation of cellular DNA from GBS, high molecular weight DNA was extracted from the pelleted cells essentially using the JGI bacterial DNA isolation CTAB protocol[3]. Briefly, this involved cell lysis with lysozyme (2.6 mg/mL), proteinase K (0.1 mg/mL), and SDS (0.5%), followed by purification of DNA by incubation with CTAB (1%) and sodium chloride (0.5 M), organic extraction, alcohol precipitation, treatment with RNase A (0.1 mg/mL), and an additional alcohol precipitation step.

Whole-Genome Amplification of Viral Metagenomic DNA

For the viral library that contained sequences of OS3173, a linker-based amplification method was used as described (Schoenfeld et al., 2008). For subsequent viral preparation isolated viral metagenomic DNA was amplified with an Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare, Piscataway, NJ, United States) following the manufacturer's protocol. Briefly, 9 μL sample buffer and 1 μL sample DNA were mixed and incubated at 95°C for 3 min and then placed on ice. Nine microliter reaction buffer and 1 μL enzyme were then mixed and combined with the 10 μL sample and incubated for 2 h at 30°C and 10 min at 65°C. The amplified DNA was then precipitated with 0.2 M NaCl and 70% ethyl alcohol and resuspended in 40 μL water. The amplified DNA was debranched by adding 10 μL of 5X S1 nuclease buffer and 2 μL S1 nuclease (200 U; Thermo Fisher Scientific Inc., Waltham, MA, United States), mixed and incubated at 25°C for 30 min and then 70°C for 10 min. The sample was reprecipitated twice with 0.2 M NaCl and 70% ethyl alcohol and resuspended in 20 μL water. Several amplification reactions were prepared and used for DNA sequence analysis and to construct a large insert library in order to capture regions of the viral replisome.

[3]https://jgi.doe.gov/user-programs/pmo-overview/protocols-sample-preparation-information/jgi-bacterial-dna-isolation-ctab-protocol-2012/

Metagenomic Sequencing and Assembly

The amplified Octopus Spring viral metagenomic DNA was sequenced using Roche 454 GS FLX Titanium chemistry at the Broad Institute (229,553 reads averaging 375 nucleotides each; 86,161,605 bases in total). The full read set was assembled *de novo* with CLC Genomics Workbench 8.0, using word size of 20 and bubble size of 375. A total of 5,143 contigs of length >500 were assembled with N50 = 1,818 bp, average length of 1,586 bp, maximum contig length of 35,614 bp, and total assembly length of 8,156,404 bp. Of the 229,553 original reads, 66% (152,673 reads) were incorporated into contig assemblies >500 bp.

Of the reads, 56.6% (86,379 reads) mapped to the largest contig at a stringency of 90%, which eventually was closed as Octopus Spring OS3173 virus, resulting in an average coverage of 913-fold. The OS3173 consensus viral sequence was finished by an iterative process of extending the ends of this viral scaffold with partially mapped reads until the extended consensus ends were found to overlap. This resulted in a 37,256 bp circular genome. A total of 99,924 reads were mapped to the finished genome (also at 90% stringency), and reads were found to map continuously across the joined overlap, consistent with a circular topology. Reads that did not map at 90% stringency were saved and remapped at relaxed stringency (80% identity over 80% length). These relaxed stringency reads were found to contain structural variants. The origin of the reported viral sequence was arbitrarily set to the beginning of the first ORF clockwise of the negative to positive GC skew transition (**Figure 1**). Viral contigs with lower coverage from the virus-enriched metagenome were obtained by reassembling the same reads using SPAdes v. 3.13.1 (Bankevich et al., 2012) with default parameters, except for the option "–only-assembler."

Both cellular and amplified viral metagenomes from GBS were sequenced at the DOE Joint Genome Institute using Roche 454 GS FLX Titanium chemistry. Double-stranded genomic DNA samples were fragmented via sonication to fragments ranging between approximately 400 and 800 bp. These fragments were end-polished and ligated to Y-shape adaptors during 454 Rapid Library Construction. Clonal amplification of the library fragments was then performed in bulk through hybridization of the fragments to microparticle beads and subsequent emulsion-based PCR. Beads containing amplified DNA fragments were loaded into wells of a Pico Titer Plate (PTP) so that each well contained a single bead, followed by sequentially flowing sequencing reagents over the PTP. For the water-borne cell metagenome, a total of 355,082 reads were obtained ranging in length from 56 to 2,049 nucleotides, with an average read length of 554 nucleotides, producing 196,771,207 bases in total. During preprocessing through the DOE-JGI Metagenome Annotation

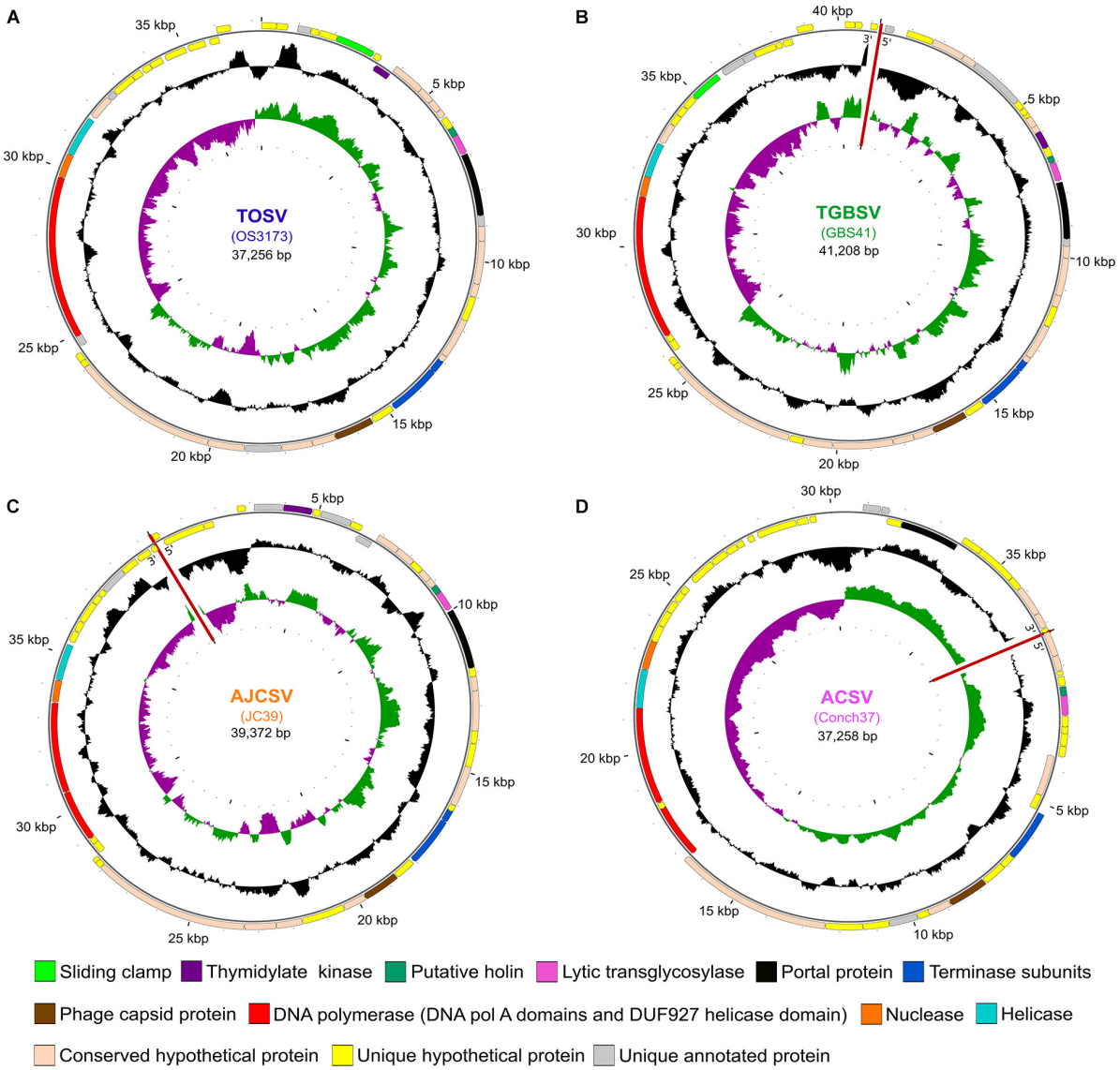


FIGURE 1 | Map of four representative UViGs. The uncultivated viral genomes were recovered from metagenomes from **(A)** Octopus Spring (TOSV); **(B)** Great Boiling Spring (TGBSV); **(C)** Joseph's Coat Spring (AJCSV); and **(D)** Conch Spring (ACSV). Outer circles show ORFs and selected annotation features, with arrows in the putative direction of transcription. Middle circles show the GC content and the inner circles show the GC skew. The sequences of GBS41, JC39, and Conch37 could not be circularized as indicated with red lines. Maps have been rotated to reflect the orientation of OS3173. TOSV, TGBSV, AJCSV, and ACSV are represented by OS3173, GBS41, JC39, and Conch37, respectively.

Pipeline [MAP; (Huntemann et al., 2016)][4], 454 reads shorter than 150 bp and longer than 1,000 bp were removed. The remaining reads were assembled with SPAdes v 3.6.1 (Bankevich et al., 2012), to a total of 315,164 contigs or sequences resulting in a total assembled size of 131,296,876 bases. Gene calling on the assembled sequences were done through the DOE-JGI MAP (Huntemann et al., 2016). Through this pipeline, CRISPR array prediction was also done and a total of 508 CRISPR arrays were found to be present in the GBS cell metagenome. After binning with the DOE-JGI binning pipeline, a single *T. jamiesonii* MAG was recovered. For the amplified viral metagenome or GBS virus-enriched metagenome, a total of 787,720 reads were sequenced ranging between 53 and 1,200 nucleotides for a total read library size of 392,631,172 bases. Read processing and assembly was also performed through the DOE-JGI MAP, in the same manner as the cellular metagenome. The virus-enriched metagenome had a total assembled size of 27,375,388 bases, which was divided over 55,185 contigs. In contrast to the cellular metagenome, only 137 RNA genes were predicted for this metagenome, supporting a low level of cellular contamination, and 74,087 protein-coding genes were predicted. A total of 60 CRISPR arrays were predicted.

[4]https://img.jgi.doe.gov/m/doc/MetagenomeAnnotationSOP.pdf

Functional Annotation

Open reading frames in OS3173 were identified by the GeneMarkS heuristic algorithm (Besemer et al., 2001). Open reading frames identified by GeneMarkS were submitted to NCBI BLASTP (Altschul et al., 1990) using default settings for comparison with proteins in the public database. The ORFs for all other UViGs were retained as identified through the DOE-JGI MAP (Huntemann et al., 2016).

Putative protein functions for four representative UViGs were inferred from searches against the NCBI non-redundant (nr) protein database with BLASTP using default settings[5], NCBI Conserved Domain Database (CDD)[6] with CD-Search, UniProtKB with HMMer[7], and CDD, Protein Data Bank (PDB), SCOPe 70 and Pfam with HHPred[8]. An *E*-value cutoff of $1e^{-10}$ was used for all tools. For each tool, the result with the lowest *E*-value that was not a "hypothetical protein" was chosen as the putative function predicted by that tool (Stamereilers et al., 2018). In some instances, putative function was assigned by synteny and gene length (e.g., small subunit of the terminase, holin).

In order to compile a composite annotation for all four of the UViGs used as representatives of the four PolA species (i.e., "*Pyrovirus*"), all manual annotations were combined with functional annotations determined via the DOE-JGI MAP. Bidirectional BLASTP (Altschul et al., 1990) analyses were performed between all four viral representatives with default settings. Genes that were bidirectional best hits were considered homologous and robust annotations (separately identified as having the same function in at least two of the four UViGs) were transferred to all homologs. Where homologous genes had no functional annotation, or contradicting annotations between the reference sequences, the respective genes were denoted as encoding conserved hypothetical proteins.

Single-Gene Trees

In order to place the viral sequences identified to be close relatives of OS3173 into phylogenetic context, two single-gene phylogenetic analyses were conducted on the protein sequences of firstly, the PolA from all viral scaffolds, together with the 3173 PolA-like sequences from Schoenfeld et al. (2013), and secondly, the large subunit sequence of terminase. For the PolA phylogeny, the 3173 PolA-like sequences of *Thermocrinis* species were used for outgroup purposes based on previous studies (Schoenfeld et al., 2013). In contrast, the terminase large subunit phylogeny was unrooted, and reference sequences of Chelikani et al. (2014), were used to infer the potential DNA packaging strategy of these viruses (Chelikani et al., 2014). Due to the variability present in these viral genes, the protein sequences were aligned based on structurally homologous protein domains with DASH (Rozewicki et al., 2019) in MAFFT v. 7 (Katoh et al., 2019)[9], with default settings. The appropriate protein model of evolution was determined for the respective alignments with ProtTest 3.4 (Darriba et al., 2011) and maximum likelihood analyses were conducted with RAxML v. 8.20 (Stamatakis, 2014). Branch support for the phylogenies was inferred from 1,000 bootstrap pseudoreplicates.

Prediction of Protein Domains

For the prediction of protein domains from the 3173 PolA-like sequences, a search of domain profiles based on hidden Markov Models was conducted through the EMBL-EBI hmmsearch tool[10] against the pfam database (El-Gebali et al., 2019). Protein family domains were predicted for all 3173 PolA-like protein sequences used in this study to determine whether the DUF 927 helicase and DNA pol A exo domains are fused to the pol A domain of the 3173 PolA-like proteins to form a polyprotein. Transmembrane domains for putative holins present in the four representative genomes from the proposed genus "*Pyrovirus*" were predicted through the TMHMM server[11].

Genome Maps

Genome maps for the four reference sequences were constructed with CGView (Grant and Stothard, 2008)[12]. The GC content and skew for each genome was calculated with a step size of 1 bp using a sliding window of 500 bp. Protein-coding sequences were colored based on the homology inferences from the synteny analyses and the composite annotations for each genome. Breaks in the UViG sequences that were not circularized, i.e., GBS41, JC39, and Conch37, were indicated with red lines in all three tracks of the genome maps. The genome maps were rotated to align with that of OS3173 for easier visualization.

Relative Abundance of Viral Contigs in Viromes

From the metagenomes analyzed, viral genomes were predicted with VirSorter v. 1.0.5 (Roux et al., 2015), Earth's Virome pipeline (Paez-Espino et al., 2016), and Inovirus detector pipeline v. 1.0 (Roux et al., 2019b)[13]. From the respective viral-enriched metagenomes, 372 contigs were obtained with 42 contigs ≥10,000 bp. Dereplication was done with an Average Nucleotide Identity (ANI) of 95% over an alignment fraction of 85% to obtain 320 non-redundant contigs. Contig coverage was estimated by mapping reads from individual metagenomes to the 320 non-redundant viral contigs using BBMap v. 38.67[14]. Only reads that mapped at ≥95% nucleotide identity were considered and contig coverage was set at 0 if less than 70% of the contig's length was covered by metagenomic reads, or as the average read depth per position otherwise, as typical for UViG analysis (Roux et al., 2019a).

[5]http://blast.ncbi.nlm.nih.gov

[6]http://ncbi.nlm.nih.gov/Structure/cdd

[7]http://hmmer.org

[8]https://toolkit.tuebingen.mpg.de/tools/hhpred

[9]https://mafft.cbrc.jp/alignment/server/

[10]https://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch

[11]http://www.cbs.dtu.dk/services/TMHMM/

[12]http://stothard.afns.ualberta.ca/cgview_server/

[13]https://bitbucket.org/srouxjgi/inovirus/src/master/

[14]https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner

Viral Classification

All contigs ≥10,000 bp obtained from the virus-enriched metagenomes, together with the four representative UViGs, were used as input with the viral reference sequence database (RefSeq v94), to automatically delineate genus-level groups based on shared gene content in vContact2 using default parameters (Jang et al., 2019). The resulting gene-sharing network was viewed and edited in Cytoscape 3.7.2[15], using a prefuse force directed layout.

Proteomic Tree and Synteny Analyses

In order to confirm the relationships among the seven UViGs, a proteomic tree was constructed with ViPtree (Nishimura et al., 2017)[16]. This Neighbor-Joining (NJ) tree is constructed by computing genome-wide tBLASTx similarity scores (McGinnis and Madden, 2004) among all submitted and all reference viral sequences. From these similarity scores a distance matrix was obtained and used for constructing a BIONJ tree. Based on previous results, the nucleic acid type was specified as dsDNA, with prokaryotes indicated as the potential hosts. Gene predictions as performed above were used for the UViGs. This process was repeated for the 10 UViGs with the highest coverage in the two virus-enriched metagenomes (i.e., from Octopus Spring and Great Boiling Spring). For depicting synteny, the genome alignments based on tBLASTx analyses, as inferred with ViPtree, was used.

Host Identification for Abundant Viruses in Great Boiling Spring and Octopus Spring

The ten viruses with the highest coverage in Great Boiling Spring and Octopus Spring, respectively, were identified from the viral metagenomes. A two-pronged approach was employed to identify potential hosts for these viruses. The first approach consisted of identifying potential prophages in bacterial and archaeal genomes, while the second approach consisted of identifying CRISPR spacers in host genomes matching the viral sequences.

For the identification of potential prophages matching the viral sequences, BLASTn analyses were conducted with the 10 viruses with the highest coverage in each spring to the DOE JGI/IMG isolate genome database (Chen et al., 2019), as well as the NCBI Whole Genome Shotgun (WGS) and RefSeq Genomic (refseq_genomic) databases.

For the second approach, CRISPR clusters were used from all metagenomes, single-amplified genomes (SAGs) and isolate genomes, available on IMG for Octopus Spring and Great Boiling Spring. All CRISPR spacer regions available on IMG for these genomes were used for further analysis. Those SAGs and isolate genomes that did not have CRISPR prediction results available on IMG were analyzed with CRISPRCasFinder (Couvin et al., 2018)[17]. All predicted spacer regions were then compared to the ten most covered virus sequences in each spring using BLASTn [BLAST v.2.2.31; (Altschul et al., 1990)] with custom settings (-word_size 7 -gapopen 10 -gapextend 2 -penalty -1 -outfmt 6 -dust no). For the spacer comparisons from the metagenomes, only spacer regions with matches over 100% of the length of the spacer were considered to limit the vast number of sequences interrogated, while matches over 80% of the length of the spacers were considered for SAGs and isolate genomes. Resulting BLAST hits were then further limited to those with a percentage identity of ≥80% and an Expect(*e*)-value of ≤0.00001.

For the CRISPR spacer detection of the four representative UViGs to *Hydrogenobaculum* sp. 3684 (Romano et al., 2013), *S. yellowstonense* SS-5^T (Reysenbach et al., 2009), *T. ruber* OC1/4^T, and *T. jamiesonii* GBS1^T, these microbial isolate genomes were subjected to CRISPR array prediction with CRISPRCasFinder. The resulting CRISPR arrays with a confidence level of three or above were further analyzed. All predicted spacer sequences were subjected to BLASTn analyses against OS3173, GBS41, JC39, and Conch37 as described above.

Recruitment Plots

To visualize the level of variability within the viral populations and coverage across the UViGs for Octopus Spring and Great Boiling Spring, raw sequence reads were recruited to the UViGs of OS3173 and GBS41. The UViGs were used to construct BLAST databases using makeblastdb in BLAST v. 2.2.31. Following this, BLASTn analyses were conducted with each UViG database as reference and their respective metagenomic reads from which they were assembled, as query. Default settings for BLAST analyses were used apart from specifying tabular format for the data output (-outfmt 6), reporting a single HSP per subject sequence (-max_hsps 1), and keeping a single alignment per subject sequence (-max_target_seqs 1). The BLAST results were formatted with BlastTab.catsbj.pl[18] limiting the identity of hits to report to 30%, and these data was then subjected to recruitment plot construction with enve.recplot2 in the Enveomics Collection (Rodriguez-R and Konstantinidis, 2016)[19] in RStudio v. 3.6.1. To compare obtained recruitment plots to the genomic architecture of the UViGs, annotated UViGs were visualized with Geneious R7 (Biomatters) and edited in Inkscape v. 0.92.

Sequence Accession Numbers

The individual sequence reads from the 2007 Octopus hot spring viral sample can be accessed at https://www.imicrobe.us/#/samples/345. The quality-filtered reads for the Octopus Spring metavirome and the Great Boiling Spring metavirome have been submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) with sequencing run accession numbers SRR12281643 and SRR12248450, respectively. The accession numbers for the twelve "*Pyrovirus*" contigs can be found in the DOE-JGI IMG/M

[15]http://cytoscape.org

[16]https://www.genome.jp/viptree/

[17]https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index

[18]http://enveomics.blogspot.com/2013/01/blasttabcatsbjpl.html

[19]https://github.com/lmrodriguezr/enveomics

TABLE 1 | Distribution of OS3173-like *polA* genes in metagenomic databases.

	Hot spring[a]	Temp. (°C)	pH	% AA ID[b]	Largest scaffolds (kbp)	Genbank or IMG accession
Yellowstone National Park	Octopus	85	8.0	82–90	37, 28	MK783188.1, JGI20132J14458_1000016
	Conch	85	8.8	66–91	37, 32	Ga0080008_153848, Ga0080008_158027
	Joseph's Coat	74	2.5	25–37	39	Ga0080003_1000231
	Bath	85	8.0	70–94	1	2007311021
	Black Pool	73	8.0	56–89	8	Ga0111098_10004
	Calcite	75	7.8	39–49	32	YNPsite12_CeleraDRAF_ scf1119014592999
	Bechler	81	7.8	83–92	0.7	YNPsite13_CeleraDRAF_29640
United States Great Basin	Great Boiling	80	6.4	35–50	41	Ga0097684_1000009
	Sandy's West	86.6	7.0	34–57	7	Ga0105155_1001723
	Little Hot Creek	82	6.8	33–50	2	Ga0105158_1016092

[a] *Metagenomes from YNP (Inskeep et al., 2013; Takacs-Vesbach et al., 2013; Inskeep and Jay, unpl.).*
[b] *Range of amino acid identities to the full-length OS3173 Pol based on tBLASTx.*

(Chen et al., 2019) website[20] under IMG Scaffold ID numbers found in **Table 1**. The four representative UViGs (for the species TOSV, TGBSV, AJCV, and ACSV) have also been submitted to the NCBI[21] as third party assemblies under the nucleotide database, with accession numbers MK783188.1 (OS3173) and BK013345-7 (Conch37, JC39, and GBS41).

RESULTS AND DISCUSSION

Dominant UViGs From Octopus Spring and Great Boiling Spring Encode an Unusual DNA Polymerase

Viral particles were isolated from Octopus Spring in Yellowstone National Park and Great Boiling Spring (GBS) in the United States Great Basin by sequential tangential-flow filtration (Schoenfeld et al., 2008) and used for metagenomic sequencing (Mead et al., 2017). In parallel, the cell fraction from GBS was also used for metagenomic sequencing. Forty-three percent of the reads from the Octopus Spring virus-enriched metagenome assembled into a single contig herein called OS3173 [representative of the proposed species *Thermocrinis* Octopus Spring virus or TOSV, equivalent to the term OS3173 used previously (Mead et al., 2017)]. The OS3173 genome was 37,256 bp and encoded 48 predicted open reading frames (**Figure 1A** and **Supplementary Figure S1A**), as detailed below. Metagenomic sequence coverage was high (mean 913X) and uniform across the OS3173 genome above 95% nucleotide identity, and read depth was low at lower identity (**Supplementary Figures S1B,C**). Together, these data indicate that TOSV was likely the dominant virus present at the time and place of sampling. Among the 48 predicted genes was a full-length, PolA-type DNA polymerase polyprotein nearly identical to 3173 PolA (OCT-3173; **Figure 2**), a portion of which was previously discovered via Sanger sequencing of metagenomic clone libraries (Schoenfeld et al., 2008). The near-complete absence of any TOSV reads from a pink streamer microbial metagenome dominated by *Thermocrinis* from the outflow of Octopus Spring (Takacs-Vesbach et al., 2013) suggests viral activity is temporally or spatially variable in that environment, or that the virus has a lytic lifestyle that results in lysed cells that are rapidly cleared from the pink streamer community in the outflow channel.

Other viral contigs with lower coverage present in the Octopus Spring virus-enriched metagenome (**Figure 3** and **Supplementary Figures S2, S3**) were similar to *Pyrobaculum* Spherical Virus (PSV) (Häring et al., 2004), a member of the *Globuloviridae*, which was previously described in Octopus Spring viral metagenomes (Schoenfeld et al., 2008; Mead et al., 2017), or distantly related to *Siphoviridae* viruses infecting mesophilic *Actinobacteria* or *Leptospira* (**Supplementary Figure S3** and **Supplementary File S1**).

A similar viral contig encoding a 3173 PolA-like protein (**Figure 2**), GBS41, representative of the putatively named species *Thermocrinis* Great Boiling Spring virus (TGBSV), was obtained from the GBS cell metagenome. The GBS41 genome is 41,208 bp and encodes 53 putative open reading frames (**Figure 1B** and **Supplementary Figure S4A**). Metagenomic sequence coverage was low across the majority of the genome (mean 15.4X), yet it was highly variable in the intergenic regions on either end of the linear contig (**Supplementary Figures S4B,C**). Metagenomic reads were also recruited from the GBS virus-enriched metagenome at 50.4X coverage, where TGBSV had the highest coverage (**Supplementary File S1**), although the *de novo* assembly was not complete. In contrast to Octopus Spring, the high recruitment of viral reads from the GBS cellular metagenomes suggests active infection of *T. jamiesonii* in GBS during the time of sampling. This hot spring community is primarily planktonic and the long residence time of the GBS source pool (Costa et al., 2009) may allow for the capture of active viral infections, unlike the rapidly flowing streamer community analyzed in Octopus spring.

Other contigs from the GBS virus-enriched metagenome (**Supplementary Figures S5, S6**) were distantly related to viruses from halophilic *Euryarchaeota*, various *Sulfolobales* viruses, and PSV (**Supplementary Figure S6**). *Pyrobaculum* is

[20] http://img.jgi.doe.gov/m
[21] https://www.ncbi.nlm.nih.gov/

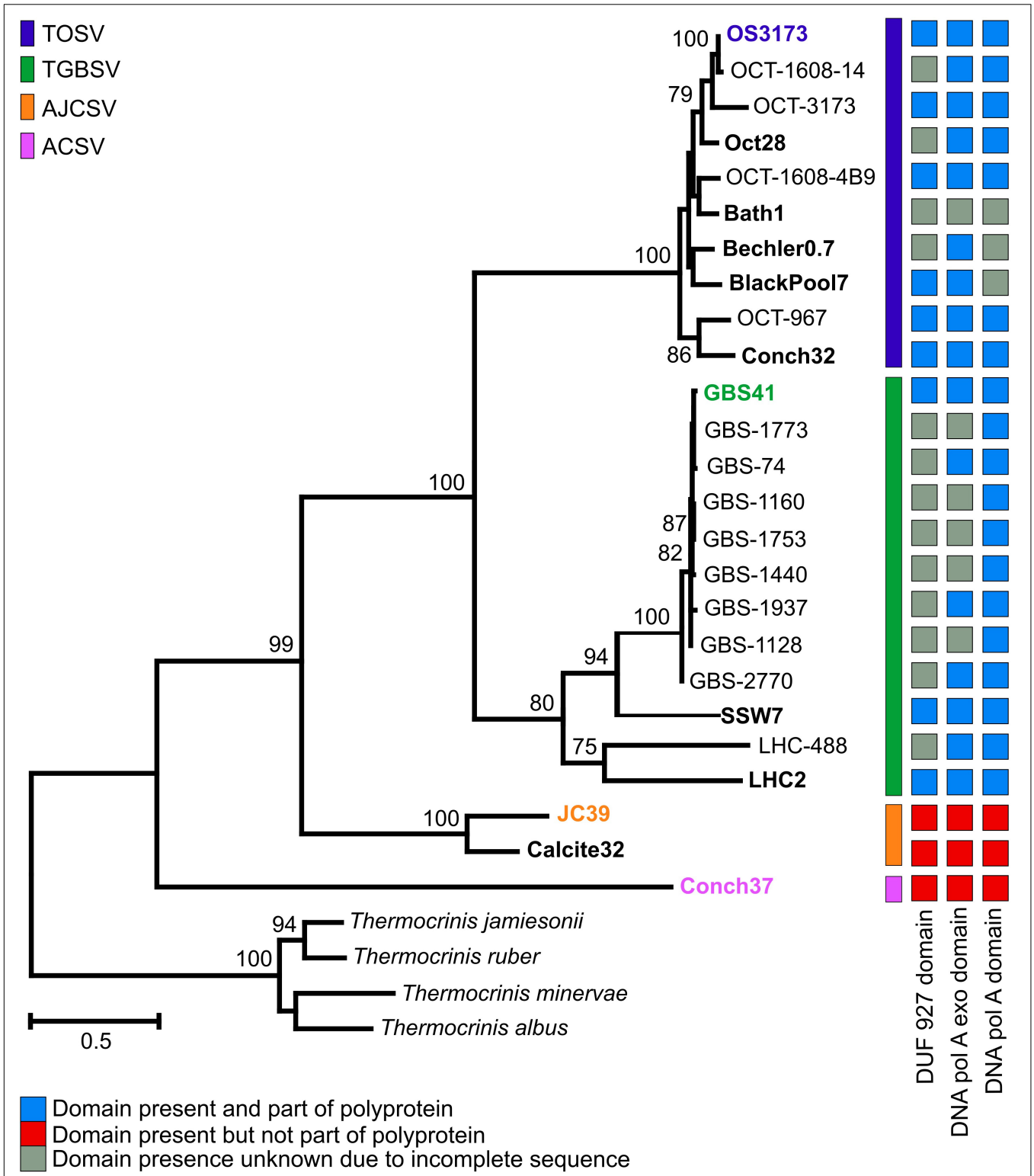


FIGURE 2 | Phylogeny and structure of 3173 PolA-like proteins. Maximum-likelihood phylogeny of near full-length 3173 PolA-like proteins, with bootstrap values above 70% from 1,000 pseudoreplicates indicated. OCT, Oct or OS, Octopus Spring; Conch, Conch Spring; Bath, Bath Spring; Bechler, Bechler Spring; BlackPool, Black Pool; GBS, Great Boiling Spring; SSW, Sandy's Spring West; LHC, Little Hot Creek; JC, Joseph's Coat Spring; Calcite, Calcite Spring. The presence of helicase, exonuclease, and polymerase domains are indicated, where known, and whether these domain forms part of the polyprotein. The exonuclease and polymerase domains of AJCSV and ACSV are present and fused as a single ORF. The scale bar indicates the number of amino acid changes per site. Taxa indicated in bold represent contigs obtained from metagenomes, while the representative UViG of species group is colored in the corresponding color.

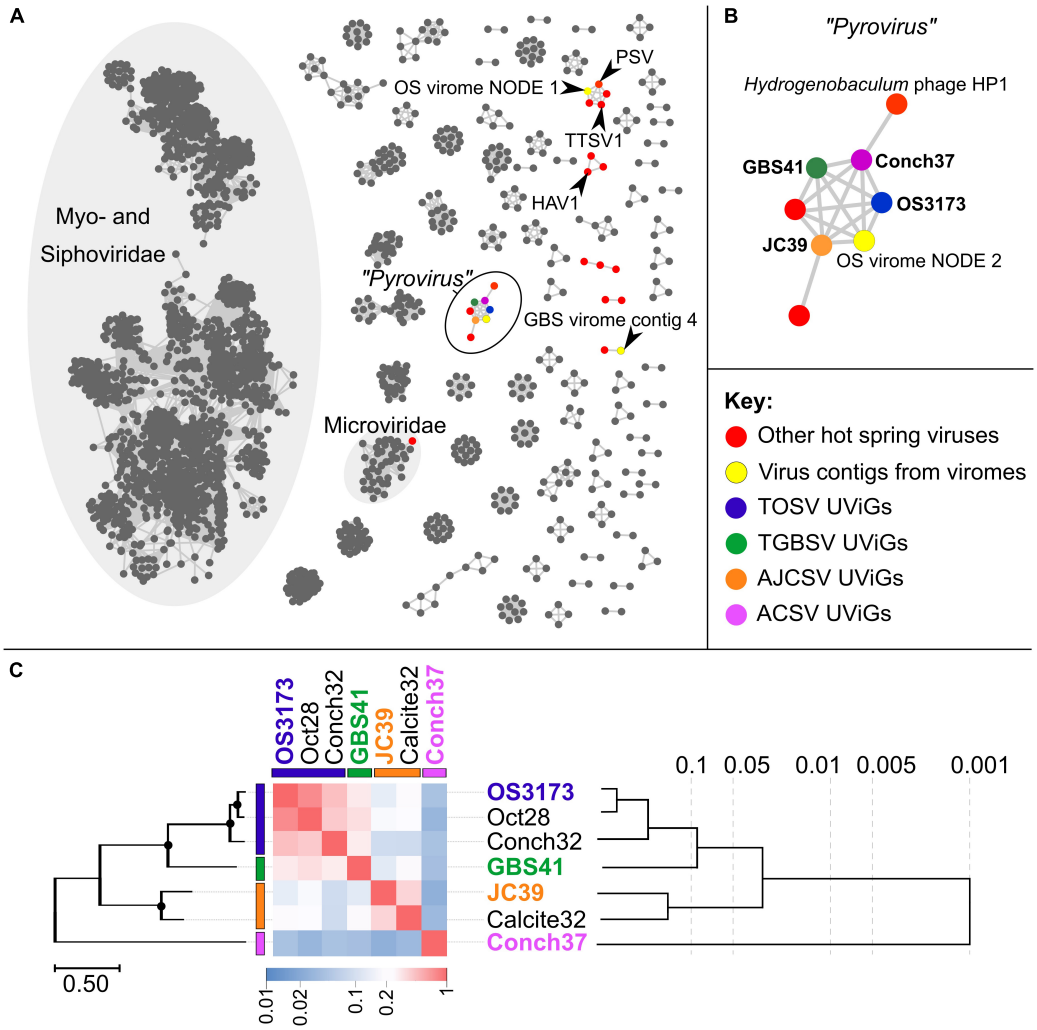


FIGURE 3 | Relatedness inferred from gene-content between the OS3173-like UViGs. **(A)** Gene-sharing network inferred by vContact2 and visualized with Cytoscape 3.7.2. Nodes in the network represent cultivated or uncultivated viral genomes, while edges represent shared gene content between nodes. Viral contigs identified from hot spring microbial genomes are indicated in red, while contigs from either the Octopus Spring or Great Boiling Spring virus-enriched metagenomes are indicated in yellow (see **Supplementary Figures S2, S3, S5, S6** and **Supplementary File S1**). The four representative UViGs are color-coded in their respective species colors. PSV, *Pyrobaculum* spherical virus; TTSV1, *Thermoproteus tenax* spherical virus 1; HAV1, Hyperthermophilic Archaeal virus 1. **(B)** Component of the gene-sharing network [circled in black on **(A)**] representing "*Pyrovirus,*" connecting the four representative UViGs together with two viruses peripheral to the network, *Hydrogenobaculum* phage HP1 and another uncultivated virus from a pink streamer microbial community metagenome from Octopus Spring, with each only sharing gene content with a single representative of the larger group. One additional viral contig of the Octopus Spring virus-enriched metagenome was connected to the genus-level group "*Pyrovirus*" (OS virome NODE 2). **(C)** Genomic relatedness among the seven related UViGs based on normalized tBLASTx scores across the genomes (heatmap) with the PolA phylogeny depicted on the left of the figure and a BioNJ phylogeny inferred from tBLASTx scores on the right. Bootstrap values above 95% on the PolA phylogeny are indicated with circles at nodes. The phylogeny based on normalized tBLASTx scores of these UViGs, and their placement within the dsDNA viral reference sequences database, is indicated in **Supplementary Figure S6**. OS3173, GBS41, JC39, and Conch37 represent TOSV, TGBSV, AJCSV, and ACSV, respectively.

relatively abundant in GBS (Costa et al., 2009; Cole et al., 2013); however, *Sulfolobales* are not known to occur at GBS, as no high-temperature, low-pH habitat is known to exist there. Due to the small size of these contigs and large genetic distance to characterized relatives, these relationships are highly uncertain.

The virus-enriched metagenomes from Octopus Spring and GBS are summarized in **Supplementary File S1**, including read recruitment, vContact 2.0 files, and CRISPR spacer matches of the 10 viral contigs with the highest coverage from these metagenomes.

Recovery of OS3173-Like Genomes From Yellowstone and Great Basin Spring Metagenomes

To assess the distribution and diversity of OS3173-like viruses, the full-length 3173 *polA* gene of TOSV was used to

recruit homologs *in silico* from public databases. In total, 12 unique contigs containing 3173 *polA*-like genes were obtained from whole-community and virus-enriched metagenomes from Yellowstone National Park and Great Basin hot springs (**Table 1** and **Figure 2** and **Supplementary Figure S7**). Thus, all UViGs from sources other than the Octopus Spring and Great Boiling Spring viromes were obtained from publicly available data. The Yellowstone springs from which OS3173-like UViGs were recovered, specifically Octopus Spring, Conch Spring, Joseph's Coat Spring, and Calcite Spring, span several geothermal areas; the pH ranges from 2.5 in this specific Joseph's Coat spring (JC2E) to 8.8 at Conch spring, originate from boiling or near-boiling spring discharge, and are known to host abundant populations of *Aquificae* (Reysenbach et al., 1994, 2000; Inskeep et al., 2010; Takacs-Vesbach et al., 2013) (Inskeep unpl.). In the Great Basin, Great Boiling Spring and Sandy's Spring West are only ∼1 km apart (Costa et al., 2009), but Little Hot Creek is ∼380 km away, and each is separated from the Yellowstone springs by >1,200 km. These springs also share a circumneutral pH, near-boiling sources, and abundant *Aquificae* populations (Costa et al., 2009; Vick et al., 2010; Cole et al., 2013).

Phylogenetic analysis of the near-complete 3173 PolA-like proteins revealed four well-supported groups (representing putative novel species) that were mostly site-specific (**Figure 2**), except that one of two Pols from Conch Spring grouped with several from Octopus Spring in TOSV, whereas a distinct Conch Spring Pol split off at the most basal node in the phylogeny (*Aquificae* Conch Spring Virus, ACSV). Additionally, the Pols from the two springs, Joseph's Coat Spring and Calcite Spring, grouped together in *Aquificae* Joseph's Coat Spring Virus (AJCSV). The Pols from Great Basin springs were monophyletic and distinct from the Yellowstone Pols, forming TGBSV, following a pattern seen for several thermophilic bacteria and archaea (Miller-Coleman et al., 2012; Dodsworth et al., 2015; Zhou et al., 2020). All the full-length 3173 PolA-like proteins contained a 3′-5′ proofreading exonuclease and DNA polymerase (3′exo/pol) domain, as is typical of many bacterial PolAs. Several also contained putative helicase domains (DUF927), described later in detail; however, this domain was fused to form a putative polyprotein in TOSV and TGBSV, or alternatively present as a separate open reading frame in the three most divergent Pols, all from springs north of Yellowstone Lake (AJCSV and ACSV) (**Figure 2**). Each of the metagenomes contained only one of the Pol variants, except for the previously mentioned Conch Spring Pols.

Seven of the 12 contigs containing the genes encoding the 3173 PolA-like proteins were >20 kbp and were thus considered UViGs (**Figure 4** and **Table 2** and **Supplementary Table S1**). All seven UViGs were compared by tBLASTx to identify other regions of homology and assess genomic synteny (**Figure 4**). Within the groups previously identified by the Pol phylogeny, shared gene content and synteny were both high. Between the groups, shared gene content and synteny was considerably lower, reflecting low average amino acid identities (**Figure 3C**); however, some of the core genes were organized similarly even in the most distant genomes, including the polymerase/helicase, terminase subunits, and phage capsid protein, described in detail below.

For the classification of these seven UViGs, vContact2 was used to delineate genus-level groups for four representatives, one from each species group in the Pol phylogeny, consisting of TOSV (OS3173), TGBSV (GBS41), AJCSV (JC39), and ACSV (Conch37) (**Figure 1** and **Table 2**). These four representatives were selected because they were considered complete or near-complete and thus allow more robust classification through gene-sharing networks (Jang et al., 2019). The four representative UViGs were connected as a single component of the gene-sharing network (**Figure 3A**), with representatives from all four groups forming a single putative genus (proposed "*Pyrovirus*") (**Figure 3B**). One peripheral member of the "*Pyrovirus*" network (i.e., shared gene content detected only with ACSV) was *Hydrogenobaculum* phage HP1 (**Figures 3A,B** and **Supplementary Figures S8, S9**; Gudbergsdóttir et al., 2016), a 19,351 bp UViG recovered from a metagenome from Grensdalur, Iceland that was assigned to *Hydrogenobaculum* based on CRISPR spacer matches to genomes from cultivated *Hydrogenobaculum* strains. A second peripheral member of the "*Pyrovirus*" network (below the "*Pyrovirus*" group, shared gene content only detected to AJCSV; **Figure 3B**) was obtained from a microbial metagenome of a pink streamer community from Octopus Spring. The gene-sharing network also illuminated some other viral contigs from the Octopus Spring and GBS viromes, belonging to gene-sharing sub-networks with PSV and *Thermoproteus tenax* spherical virus 1 (TTSV) (Ahn et al., 2006), Hyperthermophilic archaeal virus 1 (HAV) (Garrett et al., 2010), and *Microviridae*, among other isolated clusters. No genomes belonging to the primary *Myoviridae* or *Siphoviridae* networks were present in the hot spring metagenomes, reflecting the unique gene content of hyperthermophilic viruses.

Bi-Directional Genome Replication Strategy and Unique Genomic Features

The four representative genomes (OS3173, GBS41, JC39, and Conch37) ranged from 37,256 to 41,208 bp in length, ranged in GC content from 34.0 to 37.1%, and encoded 48 to 53 open reading frames, with coding fraction ranging from 94.5 to 96.5% (**Table 2**, annotations found in **Supplementary Files S2, S3**). The OS3173 contig assembled into a circular genome, whereas the other genomes could not be circularized (**Table 3**), possibly due to lower coverage or incomplete assembly owing to population heterogeneity. For now, it is uncertain whether the genomes represent circular genomes or linear genomes with terminal repeats.

For all genomes, the transcriptional orientation of the ORFs is generally divided into a 23–26 kb set of contiguous genes on the same strand (clockwise in **Figure 1**), encoding 32–37 genes, and a smaller block on the other strand (counterclockwise in **Figure 1**), encoding 13–18 genes. As with most viral genomes, most genes are located in large blocks on the same strand. In each genome there are two to four instances of changes of strand involving one to two genes, except GBS41, which consists exclusively of two large gene blocks, one on each strand. In the Conch37 genome, there are two instances of a change of strand, each consisting of two genes. In each genome, a small (750–1,350 bp)

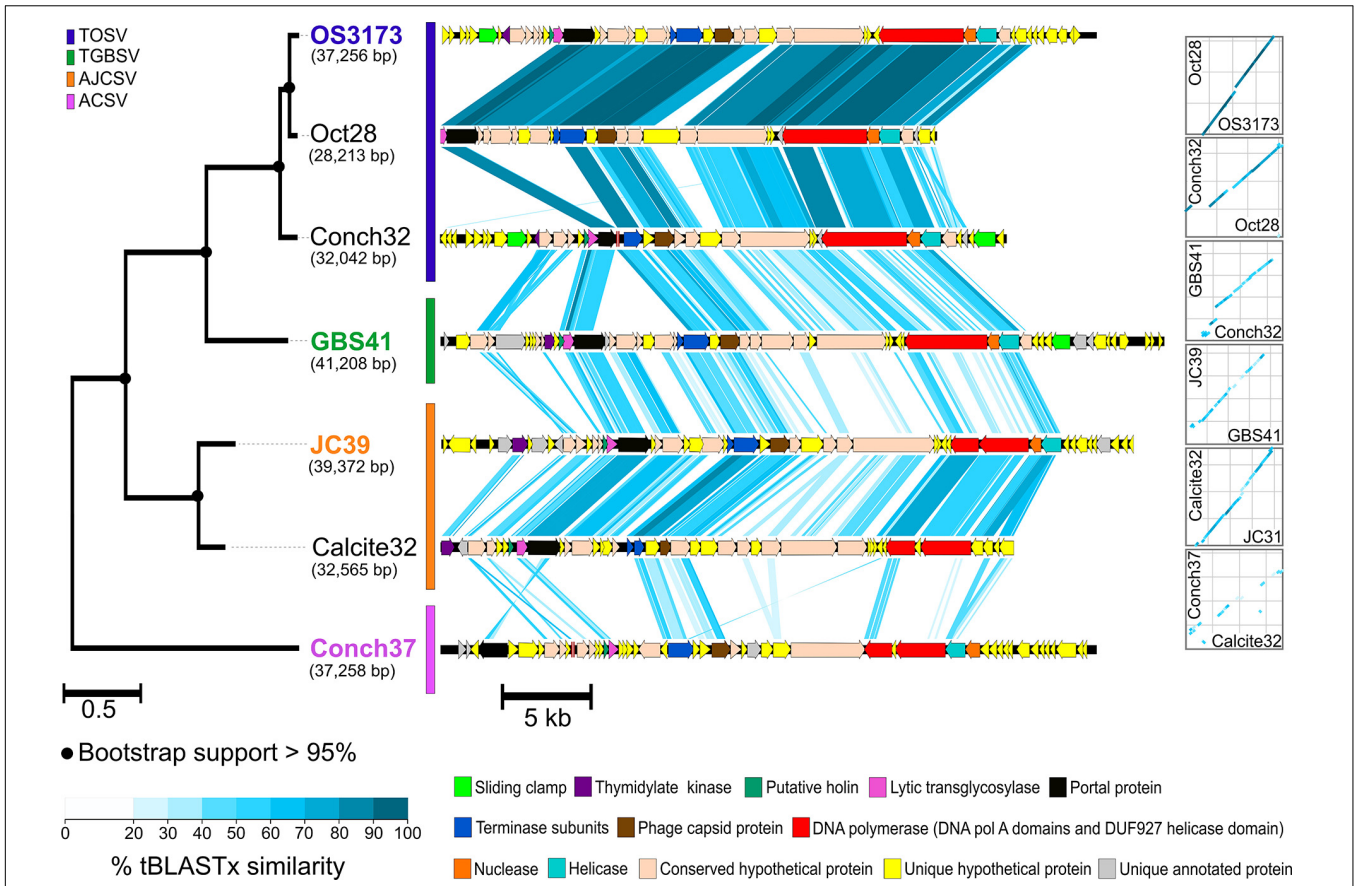


FIGURE 4 | Synteny and amino acid identity across OS3173-like viral genomes. Synteny based on pairwise tBLASTx similarity across OS3173-like viral genomes, showing high overall synteny and few inversions, despite low amino acid identity between groups. The PolA phylogeny for the seven UViGs is indicated on the left of the figure. The representative taxa of each PolA-based species group are indicated in the species' corresponding color. Bootstrap support above 95% is indicated at the nodes. On the right of the figure dot plots representing overall genomic synteny between the UViGs are indicated. OS3173, GBS41, JC39, and Conch37 represent TOSV, TGBSV, AJCSV, and ACSV, respectively.

TABLE 2 | Summary of genomic features from four representative viral UViGs.

UViG	Source	Length	%GC	Number of genes	% Coding	Annotated proteins	% Annotated proteins
TOSV (OS3173)	Octopus Spring, WY	37,265	37.1%	49	95.1%	21	35%
TGBSV (GBS41)	Great Boiling Spring, NV	41,208	36.9%	53	94.5%	19	36%
AJCSV (JC39)	Joseph's Coat Spring, WY	39,372	34.0%	51	96.5%	17	33%
ACSV (Conch37)	Conch Spring, WY	37,258	35.5%	50	94.8%	15	30%

intergenic region separated the sets of divergently transcribed genes, and this intergenic region also marked a strong divergence in GC skew. These features suggest bidirectional DNA replication beginning in the intergenic region around 36,429 bp of OS3173 and the corresponding regions of the other viral genomes. These intergenic regions also contained repetitive elements predicted to form stem-loop structures, consistent with secondary structures typical of origins of replication. Many bacterial genomes are replicated bidirectionally, and their genomes have a G > C bias in the leading strand of replication and a C > G bias in the lagging strand (Képès et al., 2012); however, dsDNA phage do not typically replicate bidirectionally (Weigel and Seitz, 2006), and in this regard we suggest these viral genomes replicate more like mini bacterial genomes than typical phage genomes. Examples of dsDNA viruses replicating bidirectionally include T7, lambda, and P4 phages, while T4 and P1 have bidirectional replication phases (Weigel and Seitz, 2006). However, cultivation of one of the viruses would be necessary to test this hypothesis.

The presence of polymerase-, nuclease/ recombinase-, and helicase-annotated genes in the smaller, counterclockwise set of genes in all four genomes suggests these genes might be transcribed earlier than the mainly structural genes in the larger, clockwise-facing block (**Figure 1** and **Supplementary File S2**). However, some genes encoding proteins associated with nucleotide metabolism were located among the clockwise-facing genes, including a DNA Pol III beta subunit (sliding

TABLE 3 | Minimum information about uncultivated virus genomes (MIUViG) for the four representative UViGs.

Metadata	TOSV (OS3173)	TGBSV (GBS41)	AJCSV (JC39)	ACSV (Conch37)
Source of UViG	Viral fraction metagenome (virome)	Metagenome (not viral targeted)	Metagenome (not viral targeted)	Metagenome (not viral targeted)
Sequencing approach	454 GS FLX Titanium	454 GS FLX Titanium	Illumina HiSeq 2000, 2500	Illumina HiSeq 2000, 2500
Assembly software	CLC Genomics 8.0 (word size = 20, bubble size = 375), SPAdes v3.13.1	SPAdes v 3.6.1	SPAdes v 3.10.0 (–meta –only-assembler -k 21, 33, 55, 77, 99, 127)	SPAdes v 3.10.0 (–meta –only-assembler -k 21, 33, 55, 77, 99, 127)
Viral identification software	VirSorter, Earth's Virome pipeline, Inovirus detector pipeline	VirSorter, Earth's Virome pipeline, Inovirus detector pipeline	VirSorter, Earth's Virome pipeline, Inovirus detector pipeline	VirSorter, Earth's Virome pipeline, Inovirus detector pipeline
Predicted genome type	dsDNA	dsDNA	dsDNA	dsDNA
Predicted genome structure	Non-segmented	Non-segmented	Non-segmented	Non-segmented
Detection type	Independent sequence (UViG)	Independent sequence (UViG)	Independent sequence (UViG)	Independent sequence (UViG)
Assembly quality	Finished	High-quality draft	High-quality draft	High-quality draft
Number of contigs	1	1	1	1

clamp) in OS3173; a thymidylate kinase in OS3173, GBS41, and JC39; and several genes that were found in only one of the four genomes, including site-specific DNA methylase (JC39), ribonucleotide reductase beta subunit (JC39), ATPase/kinase (JC39), and methyltransferases (Conch37). The location of these genes among the clockwise-facing part of the genomes and variability of these genes among the four UViGs suggest a variable and complex transcriptional/replication lifecycle for these viruses, or alternatively, that some nucleotide modification may be required during the lytic phase of infection.

Several genes encoding enzymes putatively involved in nucleic acid metabolism or DNA replication bear similarity to those in other viruses. ORF 3 of OS3173 encodes a 119-amino acid protein with some similarity to a *Sulfolobus* virus anti-CRISPR protein (Acr) (Athukoralage et al., 2020) that is highly conserved in diverse viruses and plasmids (Keller et al., 2007; Larson et al., 2007; Athukoralage et al., 2020). OS3173 and GBS41 both encode a putative sliding clamp beta subunit of DNA polymerase III, but they both lack an obvious clamp loader. Whether the viral replicase uses the host clamp loader or encodes an unrecognized clamp loader is unknown. Other viruses, including bacteriophage T4, encode sliding clamps, which have been shown to greatly increase processivity and the rate of replication (Trakselis et al., 2001). OS3173, GBS41, and JC39 each encode putative thymidylate kinases. Thymidylate kinases are encoded by a variety of viruses, including T4 and herpes simplex type 1 viruses. They are part of the nucleotide salvage pathway, typically have broad substrate activity, and are popular targets for antiviral drugs as they are often required for viability (Xie et al., 2019). ORF 5 in JC39 encodes a putative site-specific DNA methylase. Viral genome methylation is a common epigenetic defense against host restriction-modification systems. Two putative methyltransferases of unknown activity are encoded by ORF 41 and ORF 42 of Conch37.

The counterclockwise-oriented genes included three major replicase-associated proteins that were conserved in all four UViGs: an ATP-dependent helicase (ORF 38 in OS3173), a nuclease/recombinase (ORF 37 in OS3173), and a large polyprotein encoding a Pol A with functionally active polymerase activity (OS3173 Pol) (ORF 36 in OS3173). The helicase genes contain two P-loop-containing nucleoside triphosphate hydrolase domains related to the DEAD-like helicase superfamily, but the similarity to functionally characterized homologs is low. The Cas4-RecB-like nuclease (ORF 37 in OS3173) belongs to the PD-(D/E)XK nuclease superfamily, and may function as a single-stranded DNA-specific nuclease during replication and/or recombination, as these functions have been demonstrated for similar enzymes encoded by thermophilic archaeal viruses (Gardner et al., 2011; Guo et al., 2015).

In OS3173, ORF36 encodes the 1,606-amino acid polyprotein (OS3173 Pol), which was used to identify this group of viruses in the metagenomes (**Figure 2**). The amino-terminal region has conserved motifs that suggest primase and/or helicase function, including DUF927 (conserved domain with carboxy terminal P-loop NTPase) and COG5519 [Superfamily II helicases associated with DNA replication, recombination, and repair (Marchler-Bauer et al., 2010)]. Consensus Walker A and Walker B motifs suggest NTP binding and hydrolysis likely associated with helicase activity (Walker et al., 1982). As reported previously (Schoenfeld et al., 2013), the viral *polA* genes are similar to the single genomic *polA* of *Aquificaceae* and *Hydrogenothermaceae*, as well as genes found as additional *polA* copies in a variety of other bacterial genomes, and to the nuclear-encoded, apicoplast-targeted DNA polymerases of several *Apicomplexa* species, typified by the Pfprex protein of *Plasmodium falciparum*. That enzyme is also encoded as a polyprotein that is proteolytically processed to a polymerase that is optimally active at 75°C (Seow et al., 2005), much higher than would be encountered during the *Plasmodium* life cycle, but similar to the optimal growth temperature of *Thermocrinis* and the geothermal springs sampled in this study, implying lateral gene transfer (Schoenfeld et al., 2013). Understanding the biochemical functions of the rest of the ORF 36 domains could reveal new thermostable accessory proteins for DNA amplification.

Most of the clockwise-facing genes that were annotated suggest these UViGs are dsDNA tailed viruses belonging to the *Caudovirales*. Independent evidence that these viruses have dsDNA genomes comes from the initial study reporting the OS3173 PolA (Schoenfeld et al., 2008), because the viral DNA was amplified using a linker-dependent PCR method that is specific for dsDNA. Furthermore, OS3173 ORF 25, along with corresponding genes in the other UViGs, was annotated as a terminase large subunit, and ORF 24 was inferred to be a terminase small subunit, based on location immediately upstream of the terminase large subunit, open reading frame length (∼300–400 bp), and a similar isoelectric point as other terminases. The terminase small subunit protein allows recognition of the packaging site in the viral genome through specific binding of the DNA (Kala et al., 2014). Terminase large subunit phylogenies have previously been used to infer the viral DNA packaging mechanism (Chelikani et al., 2014; Merrill et al., 2016); however, the terminase large subunits from this group of viruses were distant from those of well-studied viruses, so the DNA packaging mechanism could not be inferred by this method (**Supplementary Figure S10**). Downstream of the putative terminase subunits in all genomes is a putative phage capsid protein at ORF 27 in OS3173. ORF 16 in OS3173 was annotated as a portal protein, which forms dodecameric rings that play critical roles in virion assembly, DNA packaging, and DNA injection in *Caudovirales* (Prevelige and Cortines, 2018). Additionally, GBS41 encodes a putative prohead protease (ORF 1), a WAIG tail domain protein (ORF 3), and a T7 tail fiber protein homolog (ORF 5), further supporting a relationship to *Caudovirales* and suggesting it encodes tail fibers typical of many *Caudovirales*. ORF 15 in OS3173 was annotated as a lytic transglycosylase (lysin) based on the presence of a lysozyme-like domain. ORF 14 in OS3173 was annotated as a holin based on the presence of three transmembrane domains, its small size (270 bp), and its location immediately upstream of ORF15. Also, the overlapping of open reading frames between ORFs 13, 14, and 15, suggests an anti-holin, holin, lysin operon, as found in numerous viruses. Together, these enzymes form the lysis cassette, which is common in *Caudovirales* bacteriophage, but for which there is limited knowledge in viruses of *Archaea* (Prangishvili, 2013; Saier and Reddy, 2015). No lysogeny-related genes [e.g., integrases, excisionases or Cro/CI genes (Lima-Mendez et al., 2011; Shao et al., 2017)] were identified from these UViGs, suggesting a purely lytic lifestyle. This was also supported by the fact that no prophage sequences could be identified from any microbial genomes. As most of the clockwise-facing genes appear to be involved in viral packaging and lysis, these genes are predicted to be transcribed later than the counterclockwise-facing genes, as the lysis cassette is typically the last to be transcribed (Labrie et al., 2004; Young, 2014).

Each of the UViGs encode numerous hypothetical proteins with no predicted function (∼70%; including hits to known hypothetical proteins as well as those with no homology to known proteins; **Supplementary File S2**), as is common in bacterial and archaeal viruses (Gardner et al., 2011; Gong et al., 2017; Zhang et al., 2018; Wang et al., 2019). Several of these were conserved among the genomes, but others were unique to each genome, or have diverged sufficiently that primary sequence conservation is difficult to discern. Many of the hypothetical proteins are related to genes found in different members of the *Aquificae* (**Supplementary File S2**), consistent with the previous hypothesis that *Thermocrinis* and possibly other *Aquificae* are the putative hosts for these viruses.

Putative Hosts Belong to the *Aquificae*

Arrays of Clustered Regularly Interspaced Palindromic Repeats (CRISPRs) and related Cas (CRISPR associated) genes found in many bacterial and archaeal genomes (Grissa et al., 2007) provide a means to infer virus-host relationships (Heidelberg et al., 2009; Snyder et al., 2010; Anderson et al., 2011; Gudbergsdóttir et al., 2016; Roux et al., 2019b), as the CRISPR spacers provide a record of foreign nucleic acids that have been recorded by the CRISPR-Cas system. To determine the potential host range of these UViGs, genomes derived from isolates of *Hydrogenobaculum* sp. 3684, *S. yellowstonense* SS-5^T, *T. ruber* OC1/4^T, and *T. jamiesonii* GBS1^T were screened for CRISPR arrays to identify spacers matching the UViGs. Of the four bacterial genomes analyzed, *S. yellowstonense* SS-5^T contained the highest number of CRISPR arrays at 19, with four to 41 spacer regions in each, while *T. jamiesonii* GBS1^T only contained four CRISPR arrays, with four to 15 spacer regions per array. Each of these host genomes had a number of spacer sequences with significant homology to the OS3173, GBS41, and JC39 genomes (**Figure 5** and **Supplementary Figure S11**). These matches were grouped into high confidence matches (i.e., 100% length of the spacer aligned with >90% identity) and lower confidence matches (<100% length of spacer aligned and/or <90% identity). No significant matches of spacer sequences from the putative hosts were detected for Conch37. The six CRISPR spacers from the *T. ruber* OC1/4^T genome that had homology to the viral genomes were all somewhat distant and classified as lower confidence matches (80–95% nucleic acid identity, **Supplementary Figure S11**), which is reasonable considering that this organism was isolated from samples collected from Octopus Spring in 1994 (Huber et al., 1998), and the samples from which the UViGs were assembled were collected in 2007 and 2012. Furthermore, metagenomic studies of the pink streamer community in Octopus Spring revealed three dominant *Thermocrinis* populations, but each was distinct from *T. ruber* OC1/4^T (Takacs-Vesbach et al., 2013); thus, it is possible that the *T. ruber* OC1/4^T genotype is rarely encountered by members of TOSV. To assess this possibility, we analyzed CRISPR arrays contained within metagenome-assembled genomes (MAGs) of *Thermocrinis* and other *Aquificae*, from Octopus Spring and other springs, as well as arrays in unbinned contigs to identify potential matches to the virus sequences. The MAGs did not contain any CRISPR arrays, likely because contigs including these arrays cannot be binned reliably, presumably due to the non-native nucleotide word frequency associated with the foreign-derived CRISPR spacers (data not shown). However, unbinned contigs from the metagenomes that contained high confidence CRISPR spacer matches showed homology to *Thermocrinis* genomes (**Supplementary File S1**). Similar to matches to the *T. ruber* OC1/4^T isolate genome, CRISPR spacers of

the *Hydrogenobaculum* sp. 3684 and *S. yellowstonense* SS-5^T genomes showed lower confidence matches to the viral genomes (81–92%) apart from one high confidence match between JC39 and *Hydrogenobaculum* sp. 3684 (**Supplementary Figure S11C**). By comparison, *T. jamiesonii* GBS1^T, contained three arrays with five CRISPR spacers in total with high confidence matches and significant identity to the GBS41 genome (**Supplementary Figure S11B**), providing strong evidence of the virus-host relationship.

The CRISPR spacers mapped to several different genes in the "*Pyrovirus*" genomes; however, the C-terminus of the PolA was targeted by spacers in each virus, albeit mostly low confidence matches, and another two spacers mapped to the central portion of the PolA gene in GBS41, suggesting that the C-terminus of the PolA might be a functionally important antiviral target for the host (**Figure 5** and **Supplementary Figure S11**). Accordingly, the C-terminal-encoding portion of the *polA* gene was among the most highly conserved regions of the genomes (**Figure 4**). The capsid protein gene of OS3173 and JC39 matched with low confidence to several spacers, but this was not the case for GBS41. The terminase large subunit in JC39 had matches to multiple CRISPR spacers, although this was not observed in the other two UViGs.

Thermocrinis is the dominant member of the pink streamer community in Octopus Spring (Reysenbach et al., 1994; Takacs-Vesbach et al., 2013; Colman et al., 2016) and the planktonic community in GBS (Cole et al., 2013); thus, it is reasonable to hypothesize that the natural host for the dominant viruses in these springs is *Thermocrinis*, as supported by shared gene content and CRISPR spacer matches. *Thermocrinis* is also extremely abundant in Little Hot Creek (Vick et al., 2010). Thus, we suggest that TOSV and TGBSV, all encoding the polyprotein (**Figure 2**), associate with *Thermocrinis* as their putative host. These viruses are typified by OS3173 and GBS41, with the complete UViG of OS3173 serving as the reference genome for the group.

In contrast, *Sulfurihydrogenibium* is the dominant *Aquificae* at Calcite Spring (Reysenbach et al., 2000; Inskeep et al., 2010) and the sample from Joseph's Coat Spring (JC2E) contains abundant *Hydrogenobaculum* spp. (Inskeep unpl.). We suggest that *Sulfurihydrogenibium* and/or *Hydrogenobaculum* are the most likely hosts for AJCSV and ACSV, especially as multiple homologous matches to spacer regions were obtained to both these potential hosts with the JC39 UViG. *Hydrogenobaculum* forms a distinct clade from *Thermocrinis*, *Hydrogenobacter*, *Aquifex*, and *Hydrogenivirga* within the *Aquificaceae*, and predominates in low pH springs (pH < 4.0) (Inskeep et al., 2013; Takacs-Vesbach et al., 2013). *Sulfurihydrogenibium* belongs to the sister family, *Hydrogenothermaceae*, and predominates in circumneutral springs (pH 6.5–7.8) (Takacs-Vesbach et al., 2013) and grows in a wide pH range in the lab (pH 5.0–8.8) (O'Neill et al., 2008). In this regard, it is noteworthy that some geothermal springs are poorly buffered and can change from circumneutral to highly acidic in both space and time, depending on the amounts and sources of geothermal and meteoric water that pool, and particularly on the source of sulfide, which can be oxidized to sulfuric acid by sulfide- and sulfur-oxidizing microorganisms (Nordstrom et al., 2009). Thus, it is possible that AJCSV and/or ACSV viruses encounter and infect *Sulfurihydrogenibium* in circumneutral regions of the springs and *Hydrogenobaculum* in highly acidic regions, explaining the nearly equal numbers of CRISPR spacer matches in each organism. Additionally, the gene-sharing network and a neighbor-joining tree based on amino acid identity both suggested a distant relationship to *Hydrogenobaculum* phage 1 (**Figures 3A,B** and **Supplementary Figures S8, S9**; Gudbergsdóttir et al., 2016), a 19,351 bp UViG recovered from a metagenome from Grensdalur, Iceland that was assigned to *Hydrogenobaculum* based on CRISPR spacer matches to genomes from cultivated *Hydrogenobaculum* strains. Since the exact hosts of these viruses are not conclusive, we suggest the names *Aquificae* Joseph's Coat Spring Virus (JC39, high-quality draft genome) and *Aquificae* Conch Spring

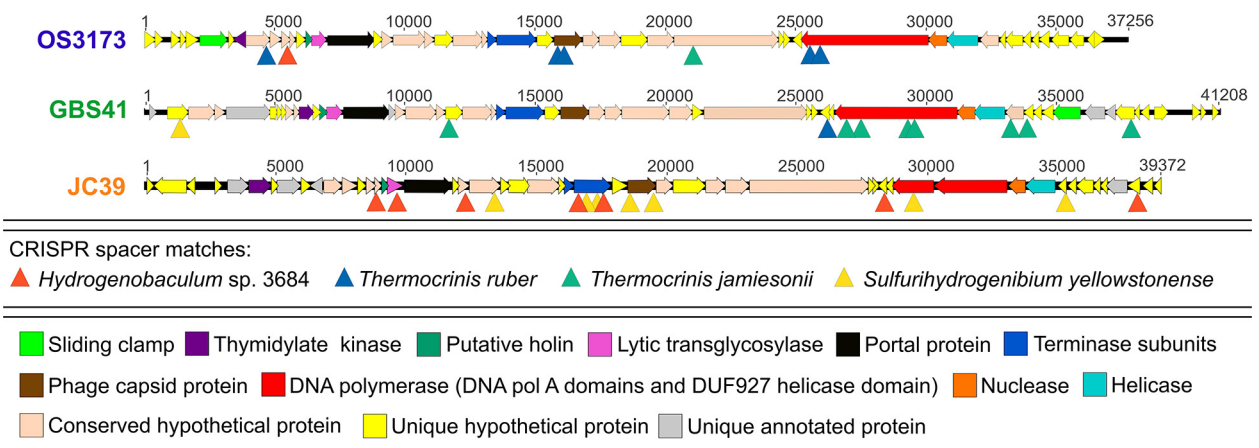


FIGURE 5 | CRISPR spacers from *Aquificae* genomes matching the "*Pyrovirus*" genomes. Linearized maps of OS3173, GBS41, and JC39 genomes with sites matching *Thermocrinis ruber* OC1/4^T, *Thermocrinis jamiesonii* GBS1^T, *Hydrogenobaculum* sp. 3684 CRISPR and *Sulfurihydrogenibium yellowstonense* SS-5^T spacer sequences are denoted by triangles. OS3173, GBS41, and JC39 represent TOSV, TGBSV, and AJCSV, respectively. Detailed information on the CRISPR spacers can be found in **Supplementary Figure S11**.

Virus (Conch37, high-quality draft genome) to represent the best genomes of these species.

DESCRIPTION OF PROPOSED VIRUSES

(Py.ro.vi'rus. Gr. n. *pur,* fire; N.L. neut. n. Pyrovirus, "fire virus," a thermophilic virus).

Based on the data presented here, we propose the following names and taxonomic relationships. Multiple genomic features suggest the seven novel UViGs belong to the order *Caudovirales*. The low overall sequence similarity and distinct placement of these taxa in gene-sharing networks suggest these viruses belong to an unclassified viral family and represent one putative genus-level group.

The proposed genus "*Pyrovirus*" accommodates TOSV, TGBSV, AJCSV, and ACSV, with the complete genome of OS3173 serving as the reference species for the genus. Members of this genus are predicted to infect *Aquificae* and are abundant in terrestrial geothermal springs. The estimated size of genomes in this genus range from 37 to 42 kb. The genomes contain genes encoding a thymidylate kinase, a holin, a lytic transglycosylase, a portal protein, large and small subunits of terminase, phage capsid protein, DNA polymerase A (with fused or unfused DUF927 helicase domain), a nuclease and a helicase. Members of this genus are proposed to employ a complex bidirectional replication strategy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

BH, JD, TS, and DM: conceptualization. MP, BH, SR, PT, and DM: data curation. MP, SR, RD, CS, AM, SM, and TG: formal analysis. MP, SR, PT, JD, ML, SM, and TG: investigation. MP, BH, SR, PT, JD, ML, SM, TG, TS, and DM: methodology. BH, SR, PT, EE-F, and DM: resources. BH and PT: supervision. MP, SR, and PT: validation. MP: visualization. BH, ML, SM, TS, and DM: writing – original draft. MP, BH, SR, PT, RD, CS, AM, JD, ML, SM, TG, TS, EE-F, and DM: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

We thank the Gordon and Betty Moore Foundation for funding the sequence of the viral metagenome from Octopus Spring. This research was supported by the United States National Science Foundation grant DEB 1557042, United States Department of Energy grant DE-EE-0000716, and the Joint Genome Institute at the DOE (CSP-182). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The publication fees for this article were supported by the UNLV University Libraries Open Article Fund.

ACKNOWLEDGMENTS

The authors would like to acknowledge Matt Henn at the Broad Institute for 454 sequencing of the viral metagenome from Octopus Spring, and W. P. Inskeep and Z. J. Jay (Montana State University) for comparison of virus sequences to metagenomes from Yellowstone National Park (YNP), critically reading the manuscript, and providing information on the precise sampling location and conditions associated with the Yellowstone metagenomes. This manuscript was submitted to bioRxiv as a preprint with the doi: 10.1101/2020.02.27.968263 (Palmer et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.583361/full#supplementary-material

REFERENCES

Ahn, D.-G., Kim, S.-I., Rhee, J.-K., Kim, K. P., Pan, J.-G., and Oh, J.-W. (2006). TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote *Thermoproteus tenax*. *Virology* 351, 280–290. doi: 10.1016/j.virol.2006.03.039

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

Anderson, R. E., Brazelton, W. J., and Baross, J. A. (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol. Ecol.* 77, 120–133. doi: 10.1111/j.1574-6941.2011.01090.x

Athukoralage, J. S., McMahon, S. A., Zhang, C., Grüschow, S., Graham, S., Krupovic, M., et al. (2020). An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* 577, 572–575. doi: 10.1038/s41586-019-1909-5

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

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607

Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3, 754–766. doi: 10.1038/s41564-018-0166-y

Chelikani, V., Ranjan, T., and Kondabagil, K. (2014). Revisiting the genome packaging in viruses with lessons from the "Giants". *Virology* 466-467, 15–26. doi: 10.1016/j.virol.2014.06.022

Chen, I.-M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/M v. 5.0: an integrated data management and comparative analysis

system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, D666–D677. doi: 10.1093/nar/gky901

Cole, J. K., Peacock, J. P., Dodsworth, J. A., Williams, A. J., Thompson, D. B., Dong, H., et al. (2013). Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J.* 7, 718–729. doi: 10.1038/ismej.2012.157

Colman, D. R., Jay, Z. J., Inskeep, W. P., Jennings, R. D., Maas, K. R., Rusch, D. B., et al. (2016). Novel, deep-branching heterotrophic bacterial populations recovered from thermal spring metagenomes. *Front. Microbiol.* 7:304.

Costa, K. C., Navarro, J. B., Shock, E. L., Zhang, C. L., Soukup, D., and Hedlund, B. P. (2009). Microbiology and geochemistry of Great Boiling and Mud Hot Springs in the United States Great Basin. *Extremophiles* 13, 447–459. doi: 10.1007/s00792-009-0230-x

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

Dodsworth, J. A., Ong, J. C., Williams, A. J., Dohnalkova, A. C., and Hedlund, B. P. (2015). *Thermocrinis jamiesonii* sp. nov., a thiosulfate-oxidizing, autotropic thermophile isolated from a geothermal spring. *Int. J. Syst. Evol. Microbiol.* 65, 4769–4775. doi: 10.1099/ijsem.0.000647

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. B., et al. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* 3, 870–880. doi: 10.1038/s41564-018-0190-y

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548. doi: 10.1038/21119

Gardner, A. F., Prangishvili, D., and Jack, W. E. (2011). Characterization of *Sulfolobus islandicus* rod-shaped virus 2 gp19, a single-strand specific endonuclease. *Extremophiles* 15:619. doi: 10.1007/s00792-011-0385-0

Garrett, R. A., Prangishvili, D., Shah, S. A., Reuter, M., Stetter, K. O., and Peng, X. (2010). Metagenomic analyses of novel viruses and plasmids from a cultured environmental sample of hyperthermophilic neutrophiles. *Environ. Microbiol.* 12, 2918–2930. doi: 10.1111/j.1462-2920.2010.02266.x

Gong, Z., Wang, M., Yang, Q., Li, Z., Xia, J., Gao, Y., et al. (2017). Isolation and complete genome sequence of a novel *Pseudoalteromonas* phage PH357 from the Yangtze River Estuary. *Curr. Microbiol.* 74, 832–839. doi: 10.1007/s00284-017-1244-8

Grant, J. R., and Stothard, P. (2008). The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36(Suppl. 2), W181–W184. doi: 10.1093/nar/gkn179

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. doi: 10.1186/1471-2105-8-172

Gudbergsdóttir, S. R., Menzel, P., Krogh, A., Young, M., and Peng, X. (2016). Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. *Environ. Microbiol.* 18, 863–874. doi: 10.1111/1462-2920.13079

Guo, Y., Kragelund, B. B., White, M. F., and Peng, X. (2015). Functional characterization of a conserved archaeal viral operon revealing single-stranded DNA binding, annealing and nuclease activities. *J. Mol. Biol.* 427, 2179–2191. doi: 10.1016/j.jmb.2015.03.013

Häring, M., Peng, X., Brügger, K., Rachel, R., Stetter, K. O., Garrett, R. A., et al. (2004). Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the *Globuloviridae*. *Virology* 323, 233–242. doi: 10.1016/j.virol.2004.03.002

Hedlund, B. P., Murugapiran, S. K., Alba, T. W., Levy, A., Dodsworth, J. A., Goertz, G. B., et al. (2015). Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. *Curr. Opin. Microbiol.* 25, 136–145. doi: 10.1016/j.mib.2015.06.008

Heidelberg, J. F., Nelson, W. C., Schoenfeld, T., and Bhaya, D. (2009). Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS One* 4:e4169. doi: 10.1371/journal.pone.0004169

Heller, R. C., Chung, S., Crissy, K., Dumas, K., Schuster, D., and Schoenfeld, T. W. (2019). Engineering of a thermostable viral polymerase using metagenome-derived diversity for highly sensitive and specific RT-PCR. *Nucleic Acids Res.* 47, 3619–3630. doi: 10.1093/nar/gkz104

Huber, R., Eder, W., Heldwein, S., Wanner, G., Huber, H., Rachel, R., et al. (1998). *Thermocrinis ruber* gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone National Park. *Appl. Environ. Microbiol.* 64, 3576–3583. doi: 10.1128/AEM.64.10.3576-3583.1998

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Tennessen, K., et al. (2016). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Standards Genomic Sci.* 11:17. doi: 10.1186/s40793-016-0138-x

Inskeep, W. P., Jay, Z. J., Tringe, S. G., Herrgard, M., and Rusch, D. B. (2013). The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front. Microbiol.* 4:67. doi: 10.3389/fmicb.2013.00067

Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., et al. (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* 5:e9773. doi: 10.1371/journal.pone.0009773

Jang, H. B., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* 37, 632–639. doi: 10.1038/s41587-019-0100-8

Kala, S., Cumby, N., Sadowski, P. D., Hyder, B. Z., Kanelis, V., Davidson, A. R., et al. (2014). HNH proteins are a widespread component of phage DNA packaging machines. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6022–6027. doi: 10.1073/pnas.1320952111

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinformatics* 20, 1160–1166. doi: 10.1093/bib/bbx108

Keller, J., Leulliot, N., Cambillau, C., Campanacci, V., Porciero, S., Prangishvili, D., et al. (2007). Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. *Virology* 4:12. doi: 10.1186/1743-422X-4-12

Képès, F., Jester, B. C., Lepage, T., Rafiei, N., Rosu, B., and Junier, I. (2012). The layout of a bacterial genome. *FEBS Lett.* 586, 2043–2048. doi: 10.1016/j.febslet.2012.03.051

Koonin, E. V., and Dolja, V. V. (2018). Metaviromics: a tectonic shift in understanding virus evolution. *VIrus Res.* 246, A1–A3. doi: 10.1016/j.virusres.2018.02.001

Labrie, S., Vukov, N., Loessner, M. J., and Moineau, S. (2004). Distribution and composition of the lysis cassette of *Lactococcus lactis* phages and functional analysis of bacteriophage ul36 holin. *FEMS Microbiol. Lett.* 233, 37–43. doi: 10.1016/j.femsle.2004.01.038

Larson, E. T., Eilers, B. J., Reiter, D., Ortmann, A. C., Young, M. J., and Lawrence, C. M. (2007). A new DNA binding protein highly conserved in diverse crenarchaeal viruses. *Virology* 363, 387–396. doi: 10.1016/j.virol.2007.01.027

Lima-Mendez, G., Toussaint, A., and Leplae, R. (2011). A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. *Res. Microbiol.* 162, 737–746. doi: 10.1016/j.resmic.2011.06.006

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. *mSystems* 3:e00055-18. doi: 10.1128/mSystems.00055-18

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39(Suppl. 1), D225–D229. doi: 10.1093/nar/gkq1189

McGinnis, S., and Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32(Suppl. 2), W20–W25. doi: 10.1093/nar/gkh435

Mead, D. A., Monsma, S., Mei, B., Gowda, K., Lodes, M., and Schoenfeld, T. W. (2017). "Functional metagenomics of a replicase from a novel hyperthermophilic *Aquificales* Virus," in *Functional Metagenomics: Tools and Applications*, eds T. C. Charles, M. R. Liles, and A. Sessitsch (Cham: Springer International Publishing), 217–242. doi: 10.1007/978-3-319-61510-3_13

Merrill, B. D., Ward, A. T., Grose, J. H., and Hope, S. (2016). Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. *BMC Genomics* 17:679. doi: 10.1186/s12864-016-3018-2

Miller-Coleman, R. L., Dodsworth, J. A., Ross, C. A., Shock, E. L., Williams, A. J., Hartnett, H. E., et al. (2012). Korarchaeota diversity, biogeography, and abundance in Yellowstone and Great Basin hot springs and ecological niche modeling based on machine learning. *PLoS One* 7:e35964. doi: 10.1371/journal.pone.0035964

Moser, M. J., DiFrancesco, R. A., Gowda, K., Klingele, A. J., Sugar, D. R., Stocki, S., et al. (2012). Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS One* 7:e38371. doi: 10.1371/journal.pone.0038371

Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017). ViPTree: the viral proteomic tree server. *Bioinformatics* 33, 2379–2380. doi: 10.1093/bioinformatics/btx157

Nordstrom, K. D., McCleskey, B. R., and Ball, J. W. (2009). Sulfur geochemistry of hydrothermal waters in Yellowstone National Park: IV Acid–sulfate waters. *Appl. Geochem.* 24, 191–207. doi: 10.1016/j.apgeochem.2008.11.019

O'Neill, A. H., Liu, Y., Ferrera, I., Beveridge, T. J., and Reysenbach, A.-L. (2008). Sulfurihydrogenibium rodmanii sp. nov., a sulfur-oxidizing chemolithoautotroph from the Uzon Caldera, Kamchatka Peninsula, Russia, and emended description of the genus *Sulfurihydrogenibium*. *Int. J. Syst. Evol. Microbiol.* 58, 1147–1152. doi: 10.1099/ijs.0.65431-0

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., et al. (2016). Uncovering Earth's virome. *Nature* 536, 425–430. doi: 10.1038/nature19094

Palmer, M., Hedlund, B. P., Roux, S., Tsourkas, P. K., Doss, R. K., Stamereilers, C., et al. (2020). Diversity and Distribution of a Novel Genus of Hyperthermophilic *Aquificae* Viruses Encoding a Proof-reading Family-A DNA Polymerase. *bioRxiv* [Preprint]. doi: 10.1101/2020.02.27.968263

Prangishvili, D. (2013). The wonderful world of archaeal viruses. *Ann. Rev. Microbiol.* 67, 565–585. doi: 10.1146/annurev-micro-092412-155633

Prevelige, P. E., and Cortines, J. R. (2018). Phage assembly and the special role of the portal protein. *Curr. Opin. Virol.* 31, 66–73. doi: 10.1016/j.coviro.2018.09.004

Rachel, R., Bettstetter, M., Hedlund, B. P., Häring, M., Kessler, A., Stetter, K. O., et al. (2002). Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch. Virol.* 147, 2419–2429. doi: 10.1007/s00705-002-0895-2

Reysenbach, A. L., Banta, A., Civello, S., Daly, J., Mitchell, K., Lalonde, S., et al. (2005). "The *Aquificales* of Yellowstone National Park," in *Geothermal Biology and Geochemistry in Yellowstone National Park*, eds W. P. Inskeep and T. R. McDermott (Bozeman: Montana State University Thermal Biology Institute), 129–142.

Reysenbach, A. L., Ehringer, M., and Hershberger, K. (2000). Microbial diversity at 83°C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and "Korarchaeota" coexist. *Extremophiles* 4, 61–67. doi: 10.1007/s007920050008

Reysenbach, A.-L., Hamamura, N., Podar, M., Griffiths, E., Ferreira, S., Hochstein, R., et al. (2009). Complete and draft genome sequences of six members of the *Aquificales*. *J. Bacteriol.* 191, 1992–1993. doi: 10.1128/JB.01645-08

Reysenbach, A. L., Wickham, G. S., and Pace, N. R. (1994). Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl. Environ. Microbiol.* 60, 2113–2119. doi: 10.1128/aem.60.6.2113-2119.1994

Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001). Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13341–13345. doi: 10.1073/pnas.231170198

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 4:e1900v1901. doi: 10.7287/peerj.preprints.1900v1

Rohwer, F., and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* 459, 207–212. doi: 10.1038/nature08060

Romano, C., D'Imperio, S., Woyke, T., Mavromatis, K., Lasken, R., Shock, E. L., et al. (2013). Comparative genomic analysis of phylogenetically closely related *Hydrogenobaculum* sp. isolates from Yellowstone National Park. *Appl. Environ. Microbiol.* 79, 2932–2943. doi: 10.1128/AEM.03591-12

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., et al. (2019a). Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* 37, 29–37. doi: 10.1038/nbt.4306

Roux, S., Hallam, S. J., Woyke, T., and Sullivan, M. B. (2015). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* 4:e08490. doi: 10.7554/eLife.08490

Roux, S., Krupovic, M., Daly, R. A., Borges, A. L., Nayfach, S., Schulz, F., et al. (2019b). Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* 4, 1895–1906. doi: 10.1038/s41564-019-0510-x

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., and Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* 47, W5–W10. doi: 10.1093/nar/gkz342

Saier, M. H. Jr., and Reddy, B. L. (2015). Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. *J. Bacteriol.* 197, 7–17. doi: 10.1128/JB.02046-14

Schoenfeld, T., Patterson, M., Richardson, P. M., Wommack, K. E., Young, M., and Mead, D. (2008). Assembly of viral metagenomes from yellowstone hot springs. *Appl. Environ. Microbiol.* 74, 4164–4174. doi: 10.1128/AEM.02598-07

Schoenfeld, T. W., Murugapiran, S. K., Dodsworth, J. A., Floyd, S., Lodes, M., Mead, D. A., et al. (2013). Lateral gene transfer of family A DNA polymerases between thermophilic viruses, *Aquificae*, and *Apicomplexa*. *Mol. Biol. Evol.* 30, 1653–1664. doi: 10.1093/molbev/mst078

Seow, F., Sato, S., Janssen, C. S., Riehle, M. O., Mukhopadhyay, A., Phillips, R. S., et al. (2005). The plastidic DNA replication enzyme complex of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 141, 145–153. doi: 10.1016/j.molbiopara.2005.02.002

Shao, Q., Trinh, J. T., McIntosh, C. S., Christenson, B., Balázsi, G., and Zeng, L. (2017). Lysis-lysogeny coexistence: prophage integration during lytic development. *Microbiol. Open* 6:e00395. doi: 10.1002/mbo3.395

Snyder, J. C., Bateson, M. M., Lavin, M., and Young, M. J. (2010). Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl. Environ. Microbiol.* 76, 7251–7258. doi: 10.1128/AEM.01109-10

Spear, J. R., Walker, J. J., McCollom, T. M., and Pace, N. R. (2005). Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2555–2560. doi: 10.1073/pnas.0409574102

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

Stamereilers, C., Fajardo, C. P., Walker, J. K., Mendez, K. N., Castro-Nallar, E., Grose, J. H., et al. (2018). Genomic analysis of 48 *Paenibacillus larvae* bacteriophages. *Viruses* 10:377. doi: 10.3390/v10070377

Suttle, C. A. (2007). Marine viruses – major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750

Takacs-Vesbach, C., Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Rusch, D. B., Tringe, S. G., et al. (2013). Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three *Aquificales* lineages. *Front. Microbiol.* 4:84. doi: 10.3389/fmicb.2013.00084

Trakselis, M. A., Alley, S. C., Abel-Santos, E., and Benkovic, S. J. (2001). Creating a dynamic picture of the sliding clamp during T4 DNA polymerase holoenzyme assembly by using fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8368–8375. doi: 10.1073/pnas.111006698

Vick, T. J., Dodsworth, J. A., Costa, K. C., Shock, E. L., and Hedlund, B. P. (2010). Microbiology and geochemistry of Little Hot Creek, a hot spring environment in the Long Valley Caldera. *Geobiology* 8, 140–154. doi: 10.1111/j.1472-4669.2009.00228.x

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951. doi: 10.1002/j.1460-2075.1982.tb01276.x

Wang, D., Jiang, Y., Xiao, S., Wang, M., Liu, Q., Huang, L., et al. (2019). Characterization and genome analysis of a novel *Alteromonas* phage JH01 isolated from the Qingdao Coast of China. *Curr. Microbiol.* 76, 1256–1263. doi: 10.1007/s00284-019-01751-3

Weigel, C., and Seitz, H. (2006). Bacteriophage replication modules. *FEMS Microbiol. Rev.* 30, 321–381. doi: 10.1111/j.1574-6976.2006.00015.x

Xie, Y., Wu, L., Wang, M., Cheng, A., Yang, Q., Wu, Y., et al. (2019). Alpha-Herpesvirus thymidine kinase genes mediate viral virulence and are potential therapeutic targets. *Front. Microbiol.* 10:941. doi: 10.3389/fmicb.2019.00941

Young, R. (2014). Phage lysis: three steps, three choices, one outcome. *J. Microbiol.* 52, 243–258. doi: 10.1007/s12275-014-4087-z

Zhang, F., Huang, K., Yang, X., Sun, L., You, J., Pan, X., et al. (2018). Characterization of a novel lytic podovirus O4 of *Pseudomonas aeruginosa*. *Arch. Virol.* 163, 2377–2383. doi: 10.1007/s00705-018-3866-y

Zhou, E.-M., Adegboruwa, A. L., Mefferd, C. C., Bhute, S. S., Murugapiran, S. K., Dodsworth, J. A., et al. (2020). Diverse respiratory capacity among *Thermus* strains from US Great Basin hot springs. *Extremophiles* 24, 71–80. doi: 10.1007/s00792-019-01131-6

Conflict of Interest: ML and SM were employed by the company Lucigen Corporation. TS was employed by the company Tamarack Bioscience, Inc. DM was employed by the company Varigen Biosciences Corporation.

The remaining 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 © 2020 Palmer, Hedlund, Roux, Tsourkas, Doss, Stamereilers, Mehta, Dodsworth, Lodes, Monsma, Glavina del Rio, Schoenfeld, Eloe-Fadrosh and Mead. 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: 11 January 2021
doi: 10.3389/fmicb.2020.597818

Novel Extracellular Electron Transfer Channels in a Gram-Positive Thermophilic Bacterium

Sergey N. Gavrilov[1], Daria G. Zavarzina[1], Ivan M. Elizarov[1], Tamara V. Tikhonova[2], Natalia I. Dergousova[2], Vladimir O. Popov[2,3], Jonathan R. Lloyd[4], David Knight[5], Mohamed Y. El-Naggar[6], Sahand Pirbadian[6], Kar Man Leung[6], Frank T. Robb[7], Maksim V. Zakhartsev[8], Orianna Bretschger[9] and Elizaveta A. Bonch-Osmolovskaya[1,10]*

[1] Winogradsky Institute of Microbiology, FRC Biotechnology Russian Academy of Sciences, Moscow, Russia, [2] Bach Institute of Biochemistry, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia, [3] Kurchatov Complex NBICS-Technologies, National Research Center "Kurchatov Institute," Moscow, Russia, [4] Dalton Nuclear Institute, FSE Research Institutes, The University of Manchester, Manchester, United Kingdom, [5] Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom, [6] University of Southern California, Los Angeles, CA, United States, [7] School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States, [8] Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway, [9] J. Craig Venter Institute, La Jolla, CA, United States, [10] Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia

OPEN ACCESS

Edited by:
Brian P. Hedlund, University of Nevada, Las Vegas, United States

Reviewed by:
Bradley Gary Lusk, Science The Earth Mesa, United States
Jessica A. Smith, Central Connecticut State University, United States

***Correspondence:**
Sergey N. Gavrilov
sngavrilov@gmail.com

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

Received: *22 August 2020*
Accepted: *04 December 2020*
Published: *11 January 2021*

Citation:
Gavrilov SN, Zavarzina DG, Elizarov IM, Tikhonova TV, Dergousova NI, Popov VO, Lloyd JR, Knight D, El-Naggar MY, Pirbadian S, Leung KM, Robb FT, Zakhartsev MV, Bretschger O and Bonch-Osmolovskaya EA (2021) Novel Extracellular Electron Transfer Channels in a Gram-Positive Thermophilic Bacterium. Front. Microbiol. 11:597818. doi: 10.3389/fmicb.2020.597818

Biogenic transformation of Fe minerals, associated with extracellular electron transfer (EET), allows microorganisms to exploit high-potential refractory electron acceptors for energy generation. EET-capable thermophiles are dominated by hyperthermophilic archaea and Gram-positive bacteria. Information on their EET pathways is sparse. Here, we describe EET channels in the thermophilic Gram-positive bacterium *Carboxydothermus ferrireducens* that drive exoelectrogenesis and rapid conversion of amorphous mineral ferrihydrite to large magnetite crystals. Microscopic studies indicated biocontrolled formation of unusual formicary-like ultrastructure of the magnetite crystals and revealed active colonization of anodes in bioelectrochemical systems (BESs) by *C. ferrireducens*. The internal structure of micron-scale biogenic magnetite crystals is reported for the first time. Genome analysis and expression profiling revealed three constitutive *c*-type multiheme cytochromes involved in electron exchange with ferrihydrite or an anode, sharing insignificant homology with previously described EET-related cytochromes thus representing novel determinants of EET. Our studies identify these cytochromes as extracellular and reveal potentially novel mechanisms of cell-to-mineral interactions in thermal environments.

Keywords: multiheme cytochromes, thermophilic prokaryotes, iron reduction, electrogenesis, biogenic magnetite crystals, Gram-positive bacteria

INTRODUCTION

Extracellular electron transfer (EET) is a unique energy conservation mechanism of prokaryotic organisms that allows them to exploit a wide variety of redox-active minerals for energy generation and to maintain direct interspecies electron transfer in complex sedimentary milieus (Shi et al., 2016). Most studies of electron exchange between microbial cells and insoluble electron acceptors are based on the data obtained for diderm Gram-negative Fe(III) reducing microorganisms. Moreover, key genes and gene products enabling the EET to solid-phase Fe(III) minerals have

been extensively characterized only in model species of mesophilic Proteobacteria belonging to the genera *Geobacter* and *Shewanella* (Shi et al., 2016).

Extracellular electron transfer processes drive important biogeochemical transformations in thermal sediments where insoluble Fe(III) compounds are abundant electron acceptors that can energize microbial life adapted to high temperature growth and survival (Kashefi et al., 2002). Thermophilic dissimilatory metal-reducers are dominated by monoderm prokaryotes (Firmicutes and hyperthermophilic Archaea), which represent many deep phylogenetic lineages (see Lusk, 2019 and the references therein). Their monolayer cell envelope obviates the need for soluble diffusible periplasmic electron shuttles and short-circuits their respiratory electron transport chains for access to such insoluble electron acceptors as Fe(III) or Mn(VI) oxides, ferruginous silicates, etc.

Current reviews highlight the key role of three groups of multiheme *c*-type cytochromes in EET in diderm bacteria (Shi et al., 2016; White et al., 2016). These include quinol-oxidizing cytochromes in the cytoplasmic membrane, electron-shuttling periplasmic cytochromes, and terminal respiratory oxidoreductases, transferring electrons directly to an insoluble acceptor on the cell surface (Xiong et al., 2006; Zhang et al., 2008). Two routes are generally proposed for the terminal electron transfer step (Shi et al., 2016; White et al., 2016). The first pathway is mediated by porin–cytochrome complexes linking intracellular electron shuttles with the cytochromes on the cell surface. Such complexes of Pcc- or Mtr-type are widespread among various iron-cycling prokaryotes (White et al., 2016), although it is still unclear whether these complexes are universal. The second pathway for the terminal step of EET is based on electrically conductive cell appendages ("nanowires"), that have been described as electron transferring type-IV pili, or "e-pili" (see Walker et al., 2018 for review), and later on argued to be stacks of cytochromes in *Geobacter* spp. (Filman et al., 2019; Wang et al., 2019) or cytochrome-functionalized membrane extensions in *Shewanella oneidensis* (Lovley and Malvankar, 2015; Subramanian et al., 2018). Cytochrome redox proteins and nanowires form multiple chimeric electron conduits in the genera *Geobacter* and *Shewanella* to support diverse electron transfer processes to different electron acceptors (Nevin et al., 2009; Coursolle and Gralnick, 2010; Kavanagh et al., 2016). EET determinants in the wide variety of prokaryotes with monolayer cell envelope are apparently highly diverse. However, limited information exists on them. Physiological mechanisms for Fe(III) respiration in the monoderm prokaryotes are cursorily reported for three Firmicutes, including two thermophiles (see White et al., 2016; Lusk, 2019 for reviews), and three hyperthermophilic archaea of the family *Archaeoglobaceae* (Manzella et al., 2015; Mardanov et al., 2015; Smith et al., 2015). The major role of multiheme *c*-type cytochromes in EET of all these organisms has been described, but specific electron transferring proteins involved in energy generation and their interactions with each other still remain poorly characterized. Recent reports on the electrically conductive archaellum from *Methanospirillum hungatei* (Walker et al., 2019) and multiheme *c*-type cytochromes involved in direct electron exchange between ANME archaea and sulfate-reducing bacteria (Krukenberg, 2018) support the notion that additional classes of electron transferring proteins remain to be discovered.

Significantly, EET pathways determine electrocatalytic activity of particular organisms and of whole microbial communities. Interaction with conductive minerals (such as magnetite) in natural environments and with charged electrodes in bioelectrochemical systems (BESs) allows prokaryotes to use these solid-state surfaces as electron acceptors or donors in the processes of microbial exoelectrogenesis or electrotrophy, respectively. Electrocatalytic activity has been well documented for a variety of pure, mixed, and enrichment prokaryotic cultures, as well as for natural sediments (Dopson et al., 2016; Nealson, 2017), but very little is known about the electrogenic activity in thermophiles. The information on their mechanisms of electrogenesis still remains limited to the detection of e-pili-like appendages in *Thermincola ferriacetica* biofilms (Lusk, 2019), heterologous expression of e-pili genes from *Calditerrivibrio nitroreducens* (Walker et al., 2018), and biochemical characterization of recombinant EET-related multiheme cytochromes OcwA and TherjR_0333 from "*Thermincola potens*" (Costa et al., 2015, 2019). OcwA, associated with the cell surface and possessing iron-reducing activity, revealed structural similarity to octaheme cytochromes involved in nitrate and sulfate reduction. Considering this fact and wide potential window of the cytochrome's activity, the authors proposed its multifunctional role in the respiratory metabolism and evolutionary closeness to the octahemes but not the known terminal reductases involved in EET (Costa et al., 2019).

Here we report exoelectrogenic and magnetite-crystallizing activity of the Gram-positive dissimilatory Fe(III)-reducing bacterium *Carboxydothermus ferrireducens*, and identify the determinants of these processes at both genomic and protein levels.

Carboxydothermus ferrireducens is an anaerobic, thermophilic bacterium isolated from a hot spring at Yellowstone National Park (Slobodkin et al., 1997). The organism belongs to the clostridial family *Thermoanaerobacteraceae* according to rRNA-based taxonomy or together with its closest relatives comprises a separate deep lineage "*Carboxydothermales*" in the class "*Desulfotomaculia*" proposed by genome-based phylogeny (Parks et al., 2018). *C. ferrireducens* couples the oxidation of organic and inorganic electron donors to the reduction of a variety of electron acceptors, including soluble and insoluble Fe(III) forms and sparingly soluble U(VI) minerals, as well as bicarbonate (in homoacetogenic growth driven by Wood-Ljungdahl pathway). The organism has a typical Gram-positive cell wall, enclosed by an *S*-layer. Our previous studies revealed that direct cell-to-mineral contact is the major physiological strategy for ferrihydrite reduction in the organism, and this strategy is promoted by cell surface-associated *c*-type cytochromes (Slobodkin et al., 1997, 2006; Gavrilov et al., 2012; Toshchakov et al., 2018). The organism does not produce any soluble electron shuttles or iron chelators which could serve as the mediators of EET (Gavrilov et al., 2012). In the current study, we describe several novel secreted multiheme cytochromes, fairly distantly related to previously

reported determinants of EET, and provide evidence for their metabolic significance and differential involvement in electron transfer chains (ETCs) to different insoluble electron acceptors, Fe(III) oxide ferrihydrite, and the anodes of BESs.

MATERIALS AND METHODS

Cultivation Conditions

Carboxydothermus ferrireducens strain DSM 11255^T was sustained in an anaerobic bicarbonate-buffered medium with poorly crystalline Fe(III) oxide ferrihydrite containing 90 mM Fe(III) (Slobodkin et al., 1997). For the experiments with different electron acceptors, the same type media lacking ferrihydrite and containing either ferric citrate [20 mM Fe(III)], sodium fumarate (20 mM) (Gavrilov et al., 2012), or none of these electron acceptors but instead a stainless steel or graphite electrode (see below) were used. Glycerol (33 mM) was used as an electron donor and a carbon source and yeast extract (100 mg/L)—as a growth factor throughout all the cultivation experiments. Cultures were maintained in Hungate tubes at 65°C, biomass for proteomic profiling and cell fractionation was obtained in the same conditions in 100 mL and 1 L serum bottles, respectively, the bottles were sealed with butyl rubber stoppers and screw caps. Current production by *C. ferrireducens* was studied using dual-chamber BESs, total volume 300 mL (Glass Reactor Set 948113, Adams & Chittenden Scientific Glass, Berkeley, CA, United States), as described previously (Bretschger et al., 2007). In all the cultivation vessels, except BESs, 70% of the vessel volume was filled with liquid medium, and 30%—with extra pure CO_2 gas. In the BESs, anodic chambers were completely filled with the liquid medium lacking any electron acceptors prior to inoculation.

Inoculation Techniques

All the transfers were made anaerobically with syringes and needles, 1% inoculum volume was used throughout. For proteomic profiling across different electron acceptors, the cultures were two times subsequently transferred from Hungate tubes with ferrihydrite medium to the tubes containing the media with corresponding electron acceptors (20 mM fumarate or ferric citrate) and third transfers from Hungate tubes to 100 mL serum bottles were used for biomass harvesting. BESs initially contained the medium without electron acceptors and were inoculated with ferrihydrite-grown cultures with residual magnetite. The inocula were taken with syringes and needles after preliminary separation of cells from the bulk of magnetite with a neodymium hand magnet attached to the Hungate tube. The total Fe content of the medium in BESs after inoculation comprised ca. 600 μM [200 μM Fe(II) and 400 μ M Fe(III)].

Determination of Growth and Fe(III) Reduction Kinetics

Growth with soluble electron acceptors was determined by direct cell counting under a CX41 phase-contrast microscope (Olympus). For counting the cells growing with ferrihydrite, subsamples of the cultures were diluted with the mixture of ammonium oxalate and oxalic acid to dissolve iron-containing minerals as previously described (Lovley and Phillips, 1988). Growth of electrogenic cells was estimated by visual observations of biofilms appearance on BES's anodes as well as by direct counting of planktonic cells in the anolyte. Iron(III) reduction kinetics was traced by determination of HCl-extractable Fe(II) species in culture subsamples with ferrozine as previously described (Gavrilov et al., 2012).

SEM and FIB Studies

For scanning electron microscopy (SEM) studies, ferrihydrite-grown cultures in stationary growth phase were used, and subsamples were taken anaerobically with syringes and needles. These subsamples were fixed in 2.5% glutaraldehyde before being subjected to dehydration series (25, 50, 75, 90, and 100% v:v ethanol, 15 min each treatment) and critical point drying (Autosamdri 815 critical point drier, Tousimis Inc., Rockville, MD, United States). The samples were then imaged at 10 keV using a JEOL JSM 7001F low vacuum field emission SEM. Focused ion beam (FIB) milling and associated SEM imaging of magnetite crystals were performed using a JEOL JIB-4500 FIB/SEM.

Electrocatalytic Activity Determination

Two bioelectrochemical reactors were used: one operated on microbial fuel cell (MFC) mode, and another one operated on potentiostatic mode. All the BES parts (except membranes and reference electrodes) were autoclaved at 121°C for 20 min and further assembled and filled with pre-sterilized culture medium aseptically in UV-irradiated anaerobic glove box (98% N_2, 2% H_2). Nafion 242 proton-exchange membranes (DuPont) and Ag/AgCl reference electrodes (+200 mV vs. SHE, RE-5B, BASi, IN, United States) were UV-irradiated for 20 min before BESs assembling. In MFC reactors, graphite felt electrodes (ca. 80 cm^2 surface area; GF-S6-06; Electrolytica) bonded to platinum wires (0.3 mm; Alfa-Aesar) were used, while potentiostatic mode reactors were assembled with meshy stainless steel electrodes (240 mesh, 80 cm^2 surface area) bonded to 0.4 mm titanium wires (both from Joyetech, China) to simplify shearing of biofilms and extracellular proteins from anodes. The culture medium lacking external electron acceptors was used in anodic chambers of both BESs types. Cathodic chambers in potentiostatic mode BESs were filled with the same sterile culture medium. In MFCs, cathodic chambers were filled with potassium ferricyanide solution. MFCs were incubated for 3 days with open circuit in the bicarbonate-buffered medium to pregrow biofilms on graphite surface, and were further operated with a 200 Ω resistor load. Before closing the circuit, samples of the pregrown *C. ferrireducens* biofilms were collected anaerobically, stained with SYBR Green fluorescent DNA stain (Sigma–Aldrich) according to the manufacturer's recommendations, and imaged using a Leica confocal laser microscope. The potentiostat mode BESs were turned on right after inoculation with anodes poised at −60 mV vs. SHE. Anode potential was controlled using the reference electrodes with IPC-micro potentiostats (Cronas, Russia). Three different potentiostatic BESs were inoculated

from the same source culture and started simultaneously. Potentiostatic BESs were operated continuously for 9 days with phase-contrast microscopic control of cell growth (see above) until appearance of transparent lysing planktonic cells in the anodic chamber that corresponded to the formation of visible biofilms on the anodic surface. After the end of incubation, subsamples of anodes with biofilms were imaged using a Leica confocal laser microscope as described above. The biofilms were used to obtain biomass of electrogenic cells for proteomic studies. The biofilms were sheared from stainless steel anodes by rigorous shaking in the culture broth, and the cells were further separated by centrifugation at 16,000 *g* 15 min.

Genome Analysis

The closed genome sequence of *C. ferrireducens* was obtained at the DOE Joint Genome Institute by Pacbio sequencing. Screening of the *C. ferrireducens* genome for multiheme cytochromes and their sequence analysis was performed as previously described (Toshchakov et al., 2018) using reported cytochromes, involved in EET in *Geobacter sulfurreducens* (Butler et al., 2010; Aklujkar et al., 2013), *S. oneidensis* (Coursolle and Gralnick, 2010; Schicklberger et al., 2013), and "*T. potens*" (Carlson et al., 2012), as queries, heme-binding motifs were predicted as previously described (Mardanov et al., 2015).

Fractionation of Ferrihydrite-Grown Cultures

Cells were cultivated in 1 L bottles and separated from magnetite, produced by the end of logarithmic growth, by rigorous hand shaking of culture bottles for 5 min and further centrifugation of the broth at 1000 *g* 5 min. Hereafter, all the centrifugation steps were performed at +4°C. The resulting supernatant with magnetite-free cell suspension was further fractionated, and the mineral pellet was washed three times with PBS-Tween buffer (pH 7.5) and finally, with 0.1% SDS. The resulting SDS outwash of magnetite was separated by centrifugation (16,000 *g*, 15 min), concentrated 20 times using Microcon YM-3 centrifugal filter devices (3 kDa cut-off membrane, Millipore Co.) and used for further MS-identification of the proteins stuck to magnetite.

The magnetite-free cell suspension was halved. One portion was used for shearing the cell surface proteins according to the previously described procedure (Mehta et al., 2005). The resulting PBS-buffered protein suspension was centrifuged for 30 min at 16,000 *g* to remove debris, and the obtained supernatant was used for identification of cell surface-associated proteins. Whole cells from another portion of magnetite-free cell suspension were pelleted at 16,000 *g* for 15 min, resuspended at pH 8.0 in 50 mM Tris-HCl buffer, then ultrasonicated, and further centrifuged at 100,000 *g* for 40 min in order to separate soluble and insoluble proteins.

Identification of Target Proteins in Subcellular Fractions

All protein samples from subcellular fractions were first routinely analyzed with SDS-PAGE and the gels were stained with Coomassie brilliant blue R-250 for proteins. For identification of the proteins associated with the cell surface or magnetite crystals, all the protein bands were prepared for further LC-MS/MS analysis using nanoAcquity LC (Waters) coupled to a LTQ Velos Pro (Thermo Fisher Scientific) mass spectrometer. To identify heme-containing proteins in subcellular fractions and cell extracts, parallel preparations with the same protein load were run on the gels and separately stained with Coomassie for proteins or benzidine for hemes according to Francis and Becker (1984). Coomassie-stained protein bands corresponding to heme bands were excised from gels and dehydrated using acetonitrile followed by vacuum centrifugation. Dried gel pieces were reduced with 10 mM dithiothreitol and alkylated with 55 mM iodoacetamide. Gel pieces were then washed alternately with 25 mM ammonium bicarbonate followed by acetonitrile. This was repeated, and the gel pieces were dried by vacuum centrifugation. Samples were digested with trypsin overnight at 37°C. Peptide mixtures of tryptic digests were separated by liquid chromatography with a gradient from 92% solution A [0.1% formic acid (FA) in water] and 8% solution B (0.1% FA in acetonitrile) to 67% A and 33% B, in 44 min at 300 nL min^{-1}, using a 0.25 × 75 mm, 1.7 μm, ethylene bridged hybrid C18 analytical column (Waters). Peptides were selected for fragmentation automatically by data-dependent analysis and subjected to MALDI-TOF MS on an UltrafleXtreme mass-spectrometer (Bruker). Mass spectra were analyzed as follows. The peak lists identified from MS spectra were subjected to the Peptide Mass Fingerprint (PMF) search at MASCOT Server v2.6 against *C. ferrireducens* amino acid FASTA dataset (2554 protein coding sequences; 739,902 residues, IMG database genome ID 2844895905). The common contaminants database was included into the search. The following parameters were employed for the PMF search: protease—trypsin; max missed cleavages—2; fixed modifications—carbamidomethyl; variable modifications—oxidation; mass values—monoisotopic; protein mass—unrestricted; peptide mass tolerance ± 100 ppm; peptide charge state 1^+; number of queries 103. The peak lists identified from MS/MS spectra were subjected to the MS/MS Ion search against the same FASTA database with the same parameters but peptide mass tolerance adjusted to ± 1.2 Da; fragment mass tolerance ± 100 ppm; instrument type—MALDI-TOF-TOF. Protein scores were calculated as $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. The scores greater than 47 were regarded significant (p-value < 0.05).

Shotgun Proteomic Analysis

For proteomic analysis, biomass was harvested from 100 mL cultures by centrifugation at 16,000 *g* for 15 min. Cell lysis, reduction, alkylation, and digestion of the proteins were performed as follows. Sodium deoxycholate (SDC) lysis, reduction, and alkylation buffer (pH 8.5) containing 100 mM TRIS, 1% (w/v) SDC, 10 mM TRIS(2-carboxyethyl)phosphine, and 40 mM 2-chloroacetamide was added to a biomass sample. The sample was sonicated and boiled for 10 min, protein concentration in the sample was routinely determined by Bradford assay, and equal volume of trypsin solution (pH 8.5) in 100 mM TRIS was added in a 1:100 (w/w) ratio. After overnight digestion at 37°C, peptides were acidified by 1% trifluoroacetic

acid (TFA) for styrenedivinylbenzene reverse-phase sulfonate (SDB-RPS)-binding, and 20 μg was loaded on two 14-gauge StageTip plugs. Equal volume of ethyl acetate was added, and the StageTips were centrifuged at 300 *g*. After washing the StageTips with a 100 μL of 1% TFA/ethyl acetate mixture and 100 μL of 0.2% TFA, peptides were eluted by 60 μL acetonitrile/ammonia (80/5%, v/v) mixture. The collected material was vacuum-dried and stored at −80°C. Before analysis, peptides were dissolved in 2% acetonitrile/0.1% TFA buffer and sonicated for 2 min. LC-MS/MS analysis was commercially performed in the core facility center "Bioorganic" (IBCh RAS, Moscow, Russia) using the Q Exactive HF benchtop Orbitrap mass spectrometer coupled to the Ultimate 3000 Nano LC System via a nanoelectrospray source (all from Thermo Fisher Scientific). Dissolved peptides were analyzed using the HPLC system configured in a trap-elute mode. Approximately 1 μg of tryptic peptide digests was loaded on an Acclaim PepMap 100 (100 μm × 2 cm) trap column and separated on an Acclaim PepMap 100 (75 μm × 50 cm) column (both from Thermo Fisher Scientific). Peptides were loaded in solvent A (0.2% FA) and eluted at a flow rate of 350 nL min^{-1} with a multistep linear gradient of solvent B (0.1% FA, 80% acetonitrile): 4–6% B for 5 min; 6–28% B for 91 min; 28–45% B for 20 min; 45–99% B for 4 min; 99% B for 7 min; 99–4% B for 1 min. After each gradient, the column was washed with 96% buffer B for 9 min. Column was kept at 40°C. Peptides were analyzed on a mass spectrometer with one full scan (375–1,400 *m*/*z*, *R* = 60,000 at 200 *m*/*z*) at a target of 3e6 ions and max ion fill time 30 ms, followed by up to 15 data-dependent MS/MS scans with higher-energy collisional dissociation (target 1e5 ions, max ion fill time 50 ms, isolation window 1.2 *m*/*z*, normalized collision energy 28%, underfill ratio 2%), detected in the Orbitrap (*R* = 15,000 at fixed first mass 100 *m*/*z*). Other settings: charge exclusion—unassigned, 1, >6; peptide match—preferred; exclude isotopes—on; dynamic exclusion—30 s was enabled.

Label-free protein quantification was made by MaxQuant software version 1.5.6.5 using *C. ferrireducens* amino acid FASTA dataset (see above) and a common contaminants database by the Andromeda search engine, with cysteine carbamidomethylation as a fixed modification and protein N-terminal acetylation and methionine oxidations as variable modifications. The false discovery rate (FDR) was set to 0.01 for both proteins and peptides with a minimum length of seven amino acids. Peptide identification was performed with an allowed initial precursor mass deviation up to 20 ppm and an allowed fragment mass deviation of 20 ppm. Downstream bioinformatics analysis was performed with Perseus.1.5.5.1. For Student's *t*-test, missing values were imputed with a width of 0.3 and a downshift of 1.8 over the total matrix. Two sample tests were performed in Perseus with s0 set to 0. Quantification was performed with a minimum ratio count of 1. To quantify proteins in each sample, the iBAQ algorithm, implemented into MaxQuant software, was used (Schwanhäusser et al., 2011). Normalization of each protein's iBAQ value to the sum of all iBAQ values generated a relative iBAQ (riBAQ) values corresponding to the mole percentage of each protein in the sample, taking the whole set of proteins in it as 100% (Shin et al., 2013).

Statistical Analysis of Proteomic Data

Two separate two-way ANOVA tests with multiple comparisons were performed using Graphpad Prism 8 software[1] (i) electron acceptors (column factor) as the main source of variation vs. five groups of proteins (row factor) determining housekeeping, respiratory, and EET processes in *C. ferrireducens* and (ii) electron acceptors (column factor) vs. individual multiheme cytochromes from the group V (row factor). Descriptions of the proteins and their groups are given in the main text below and in the legends to **Figures 5**, **6**. The group statistics (averages, standard deviations, and *n* values) has been calculated for each of five protein groups, or each of individual cytochrome proteins, in order to investigate their contribution into overall variance. Each two-way ANOVA consisted of two steps: (i) ordinary two-way ANOVA tests with *P*-value cut-off < 0.05 (**Supplementary Tables S1**, **S2**) and (ii) ANOVA tests corrected for multiple comparisons controlling FDR at Q = 0.05 (two-stage linear step-up procedure of Benjamini, Krieger, and Yekuteli with individual variances computed for each comparison). Pairwise comparisons of representation of each protein group at each culture conditions were calculated (number of families 1, number of comparisons per families 190, Q = 0.05) (**Supplementary Table S3**), and the same was done for each of individual multiheme cytochromes (number of families 10, number of comparisons per families 6, Q = 0.05) (**Supplementary Table S4**).

Phylogenetic Analysis

Cytochrome *c* protein sequences related to the proteins OmhA, SmhA, SmhB, and SmhC were retrieved from non-redundant NCBI protein database on March 2020, using the BLASTp algorithm (Altschul et al., 1990) with default parameters except the following: model sequences (XM/XP) were excluded, automatic adjustment for short input sequences was restricted, low complexity regions filter was on, max target sequences parameter was adjusted to 10,000, the expectation threshold to 1, and the gap costs to existence 11 and extension 2. The retrieved selections of sequences were manually curated. Considered meaningful were the hits with Ev < 0.001, query coverage ≥ 50%, and identity ≥ 20%, as well as the hits with query coverage between 50 and 20% having Ev ≥ 10^{-5} or a percentage identity ≥ 70% at an Ev ≥ 0.001. These criteria were derived from several comprehensive systematic studies of multiheme cytochromes diversity (Butler et al., 2010; Sharma et al., 2010; Aklujkar et al., 2013; Shi et al., 2014). The resulting selections were further checked for "improper" hits lacking conservative multiheme domains or having no matches with the queries within such domains. These hits were excluded from further analysis. The obtained sets of sequences were then consecutively filtered through 0.95 and 0.80 filters using CD-hit utility[2] to decrease redundancy. For SmhA hits, an additional third 0.60 filter was applied due to the large number of similar blast hits. The selections were then aligned by MAFFT 7 with default parameters using iterative refinement method FFT-NS-i (Katoh et al., 2019). Three final alignments (for SmhA, SmhB, and

[1] www.graphpad.com

[2] http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi

OmhA-SmhC hits) were subjected to a Bayesian inference of phylogeny using the BEAST package (BEAUti v2.6.2, BEAST v2.6.2, TreeAnnotator v2.6.0, FigTree v.1.4.4) (Bouckaert et al., 2019). Tree likelihoods (ignoring ambiguities) were determined for the alignments by creating a Monte-Carlo Markov Chain (10,000,000 generations) in three independent runs. The searches were conducted assuming an equal or a gamma distribution of rates across sites, sampling every 1000th generation and using the WAG empirical amino acid substitution model (Whelan and Goldman, 2001). In each case, the resulting 10,000 trees (omitting the first 400 trees as burn-in) were used to construct unrooted phylogenetic consensus trees. For the analysis of the trees, taxonomic information on the source organisms was retrieved from NCBI Taxonomy database, and heme numbers of the proteins were determined by screening the sequences from the alignments for heme *c*-binding motifs using CLC Genomics Workbench 9.0.

Data Availability

All data for understanding and assessing the conclusions of this research are available in the main text, SI Appendix, DOE's JGI Integrated Microbial Genomes and Microbiomes database[3], ProteomeXchange Consortium[4] via the PRIDE partner repository with identifier PXD018523 (Perez-Riverol et al., 2019).

RESULTS

Growth of *C. ferrireducens* on Ferrihydrite. Cell-to-Mineral Contacts Control Nucleation of Magnetite Crystals

Growth of *C. ferrireducens* with ferrihydrite as the only electron acceptor is accompanied by the formation of biogenic cell-magnetite conglomerates (Gavrilov et al., 2012). Confocal fluorescent microscopy revealed irregular colonization of the mineral surface in two different ways—with small clusters of several cells and with much larger conglomerates of ca. 10 μm in diameter (**Supplementary Figure S1**). SEM showed that the small cell clusters are aggregated with nanosized mineral particles, retaining the morphology of ferrihydrite. The cells in these clusters contact the bulk of the mineral phase with a network of tiny filaments, resembling pili (**Figures 1A–D**). We have previously reported the production of such cell appendages in *C. ferrireducens* in response to the presence of ferrihydrite in the medium (Gavrilov et al., 2012). Analysis of the appendages with scanned conductance (SCM, Bockrath et al., 2002) and conducting probe atomic force (CP-AFM, El-Naggar et al., 2010) microscopy in this study did not detect their electrical conductivity (**Supplementary Figure S2**).

In contrast to the small clusters of "piliated" cells, big cell conglomerates in *C. ferrireducens*' cultures are confined within unusually large (≥10 μm) fine-grained sintered magnetite crystals (**Figure 2**). Such crystals comprise the majority of the mineral phase at the end of logarithmic growth of the culture. No further increase in their number or dimensions was detected by SEM after cell lysis. The majority of *C. ferrireducens* cells, associated with large magnetite crystals, lack extracellular filaments (**Figure 2A**). The large crystals have pronounced growth steps and well developed facets (**Figures 2B,C**) but lack ideal crystallographic form, which is consistent with rapid crystallization that has been abruptly stopped. The cells associated with the large magnetite crystals outnumber the cells bound to the surrounding mineral nanoparticles (**Figures 2B–D** and **Supplementary Figure S1**). Considering this fact, we propose that actively growing microcolonies of *C. ferrireducens* can act as centers of magnetite crystallization. Crystal growth takes place during active cell division and Fe(III) reduction with further Fe(II) mineralization on the cell surfaces. This may lead to the formation of macroscopic microbe-to-mineral conglomerates, ultimately limited by the cell lysis. Interestingly, FIB-milling of the large magnetite crystals revealed cavities matching the shape of the cells. Nanocrystalline filamentous structures, which can pick the cells, were not observed inside the cell-like cavities (**Supplementary Figure S3**), thus indicating the crystallization process to be somehow controlled by the cells entrapped into magnetite.

Electrogenic Activity of *C. ferrireducens*

Considering active colonization of magnetite by *C. ferrireducens*, which implies direct electron transfer from microbial cells to the conductive mineral, we decided to evaluate electrogenic activity of the organism in two high-temperature BESs operated in MFC or potentiostatic mode. Current generation was first observed in duplicate MFC-type reactors (**Figure 3A**) with biofilms of *C. ferrireducens* pregrown at open circuit anodes (**Figures 3B,C**) as homoacetogens using bicarbonate of the culture medium as the main electron acceptor available. After closing the circuit of MFC reactors, their anodes have become available for the cells as the electron acceptors, which led to current production. The higher biomass coverage on the anode in reactor 1 (**Figure 3B**) caused faster electron donor consumption rates as is indicated in **Figure 3A**. Biomass coverage on the graphite felt electrode was lower for reactor 2 (**Figure 3C**), and therefore longer intervals between feedings (**Figure 3A**, *blue lines*). Strong positive current response was observed upon addition of the electron donor glycerol to the medium or complete medium exchange (**Figure 3A**). Interestingly, when ferrihydrite was added to the reactors, an immediate negative current response was observed followed by a rapid recovery back to previous current levels (**Figure 3A**). These data likely indicate increased scavenging of electrons, derived from glycerol oxidation, in the presence of an extra electron acceptor (Fe^{3+} mineral) in the cultures.

In potentiostatic BESs with stainless steel anodes poised at -60 mV vs. SHE, no biofilms were pregrown at open circuit electrodes. Instead, the anodic chamber was inoculated with *C. ferrireducens* culture grown with ferrihydrite and containing magnetite nanoparticles (as those surrounding the cells in **Figures 1C,D**). Initial insoluble Fe content of these BESs comprised 600 μM. Within 9 days of incubation, clearly visible thin transparent biofilms with particulate magnetite inclusions

[3]https://img.jgi.doe.gov/

[4]http://proteomecentral.proteomexchange.org

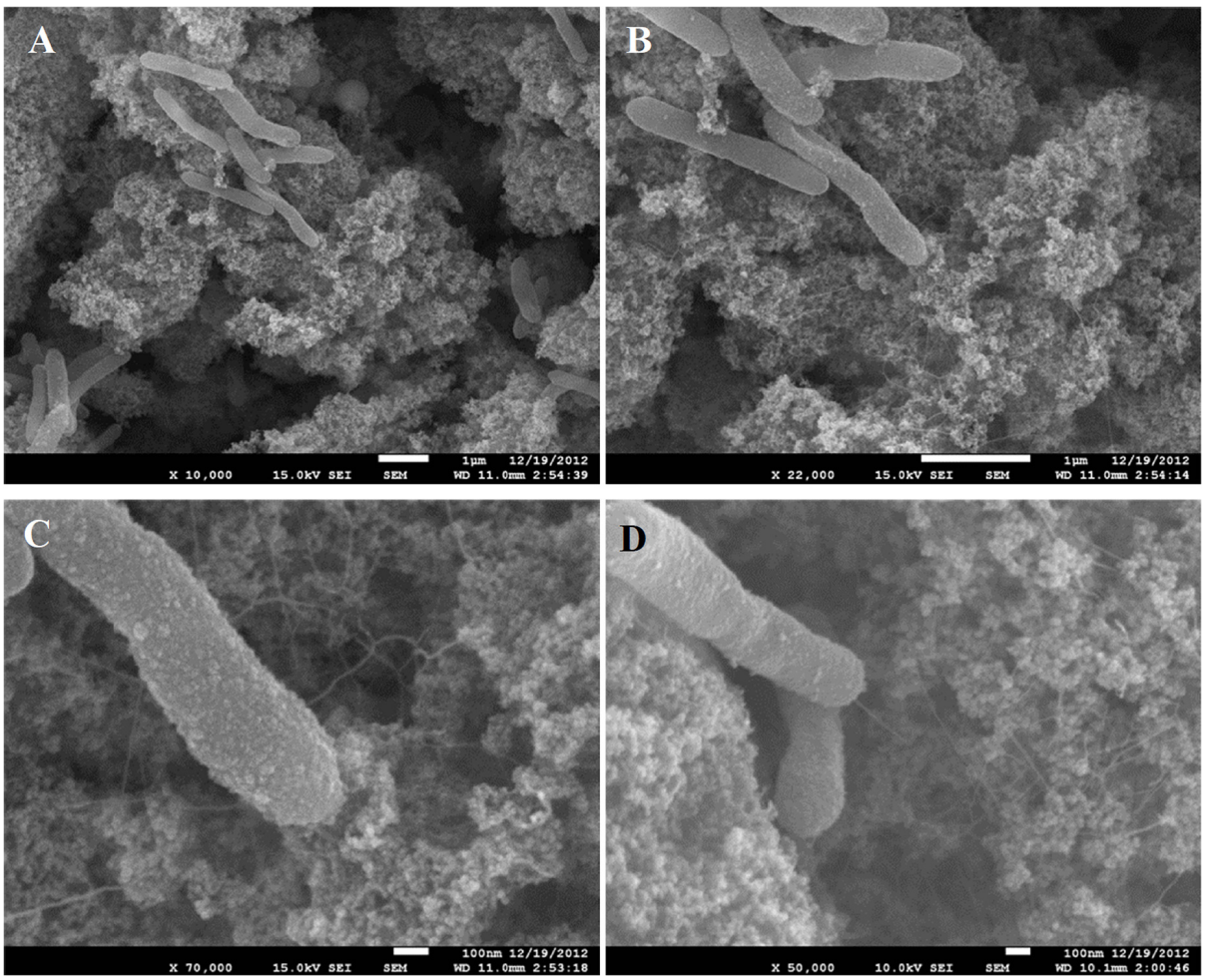


FIGURE 1 | SEM images of small clusters of *C. ferrireducens* cells grown with ferrihydrite as sole electron acceptor. Surrounding material—nanoparticulate magnetite. **(A–C)** consecutive magnification of the same microscopic field. **(D)** Another microscopic field with a cell connected with the bulk of the mineral by a single appendage.

appeared on the anodes (**Supplementary Figures S4A–D**) and have steadily generated 55–68 mA/m^2 current since the second day of incubation (**Supplementary Figure S4E**). The current density appeared to be much higher than observed with biofilms pregrown on graphite anodes (0.04–0.08 mA/m^2, **Figure 3A**).

Genome Analysis

Potential determinants of EET to Fe(III) oxide or anode were annotated in the complete genome sequence of *C. ferrireducens*, revealing a total of 19 genes encoding multiheme *c*-type cytochromes with 4–22 predicted heme-binding motifs per protein sequence. Eighteen multihemes were predicted to be secreted proteins or localized on the cell surface. Thus, any of them could be involved in the terminal step of EET. The genes encoding secreted multihemes Ga0395992_02_129131_130396 and Ga0395992_02_130419_130832 comprise a typical cluster of NrfAH dissimilatory nitrite reductase complex and were excluded from further analysis. Three of the *C. ferrireducens*' multihemes (Ga0395992_01_217646_219709, Ga0395992_01_220048_222057, Ga0395992_01_222075_224618) are encoded in a large cluster containing the complete gene set for type-IV pili assembly (listed in **Table 1** and **Supplementary Table S5**). The structural pilin PilA, encoded in this "pilin-cytochrome" cluster, was reported to lack homology with conductive e-pilins (Holmes et al., 2016). This gene cluster also encodes a "PilA–C" protein, with a C-terminus of similar topology to long type-IVa (poorly conductive) pilins and an N-terminus that is similar to porin-like proteins (Holmes et al., 2016). No genes related to the known determinants of Pcc- or Mtr-type porin-cytochrome complexes were identified in *C. ferrireducens*. Outside the "pilin-cytochrome" and *nrfAH* gene clusters, all the other cytochrome-encoding genes are randomly distributed throughout the *C. ferrireducens* genome. Among those, five proteins share homology with previously characterized parts of EET chains in iron-reducing bacteria *G. sulfurreducens*, *S. oneidensis*, and "*T. potens*" (Shi et al., 2016). Cytochrome Ga0395992_01_182949_184436 is homologous to OmcC outer surface cytochrome of the

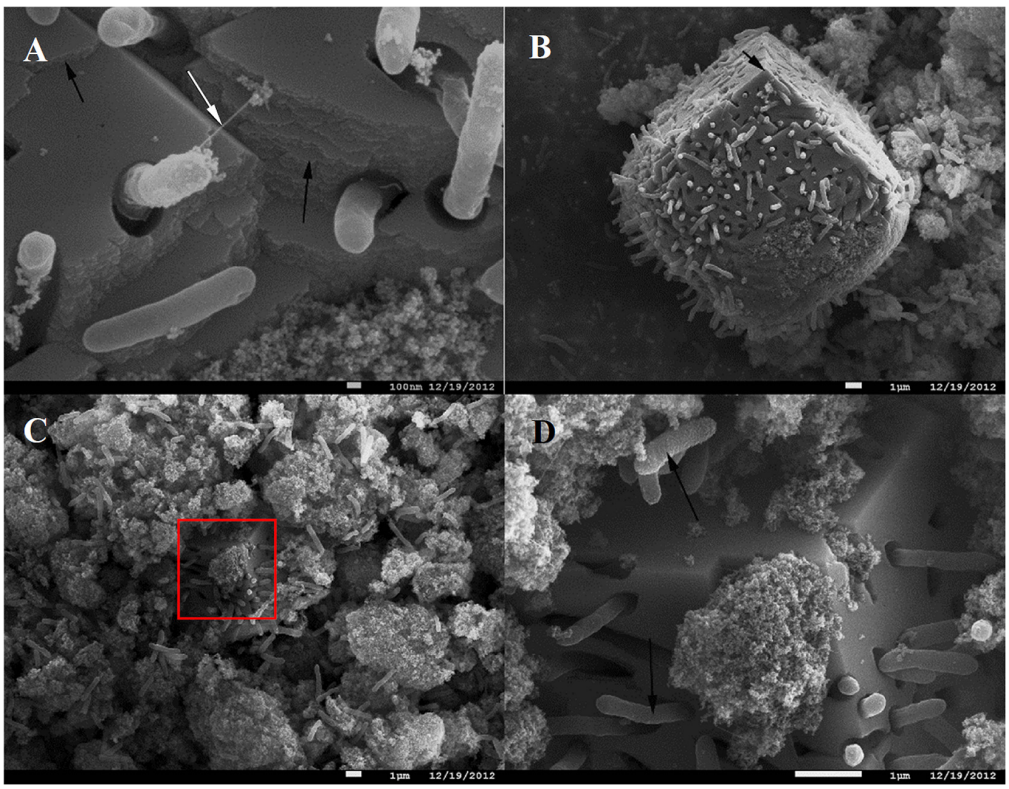


FIGURE 2 | SEM images of *C. ferrireducens* cell colonies associated with macroscopic magnetite crystals in the late logarithmic phase of growth with ferrihydrite. **(A)** Non-piliated and piliated cells on the surface of a large magnetite crystal. **(B)** Overview of a large magnetite crystal. The *white arrow* on **(A)** indicates a singular cell appendage connecting the cell entrapped into magnetite crystal with the mineral surface, almost all the other cells associated with the crystal are non-piliated. *Black arrows* on **(A,B)** indicate incremental growth steps and facets of macroscopic magnetite crystals. **(C)** Different cell aggregates—inside macroscopic magnetite crystals and on the surface of nanoparticulate mineral phase. The red box on **(C)** indicates the area magnified at **(D)**. **(D)** Cells and cell-like cavities on the surface of a large (micron-scale) magnetite crystal. *Black arrows* on **(D)** point out different surfaces of the cells contacting the nanoparticulate mineral phase and the cells entrapped into large magnetite crystals (rough and smooth cell surfaces, respectively).

Pcc-type complex in the Gram-negative *G. sulfurreducens* (Aklujkar et al., 2013), Ga0395992_02_30996_31874 and Ga0395992_02_151260_153788 are homologous to MtrA/MtrD periplasmic cytochromes of the Mtr-type complex in the Gram-negative *S. oneidensis*, Ga0395992_01_647313_648317 is homologous to CymA quinol-oxidizing cytochrome of *S. oneidensis*, while Ga0395992_02_29558_30178 is homologous to periplasmic multiheme DmsE of *S. oneidensis*, as well as to putative soluble secreted cytochromes TherJR_0333, TherJR_1117 of "*T. potens*," and Ga0395992_02_112574_114619 shares equally weak homology (23–30% identities) with three previously reported multihemes (**Supplementary Table S5**)—periplasmic cytochrome MtrA of *S. oneidensis*, OmcB cell surface-associated cytochrome of *G. sulfurreducens* (Butler et al., 2010), and OcwA outer surface cytochrome of the Gram-positive "*T. potens*" (Costa et al., 2019). Sequence analysis of Ga0395992_02_112574_114619 with CW-PRED service indicates its putative localization on the outer surface of the cell. Another *C. ferrireducens* multiheme, Ga0395992_03_307983_310529, shares weak homology with predicted cell wall-linked multiheme TherJR_1122 of "*T. potens*" (Carlson et al., 2012). Several multihemes of *C. ferrireducens*, including putative secreted octaheme Ga0395992_02_137389_139104, appeared to lack homologs among reported determinants of EET. Two multihemes of *C. ferrireducens* (Ga0395992_02_135863_137242 and Ga0395992_01_684739_686151) possess conserved domains of membrane anchor proteins of CISM family enzymes (Rothery et al., 2008), and thus might serve as quinol-oxidizing units transferring electrons from the quinol pool to terminal oxidoreductases. Alternative candidate quinol-oxidizing protein is Ga0395992_01_187284_188528, possessing conserved domain organization and number of hemes similar to those of ActA subunit of the respiratory alternative complex III (Refojo et al., 2013).

Identification of Predominant Multiheme Cytochromes Produced by *C. ferrireducens* During the Growth With Insoluble Electron Acceptors

To determine major EET-related multiheme cytochromes of *C. ferrireducens*, the organism was cultivated at three different growth conditions: with ferrihydrite, soluble ferric citrate,

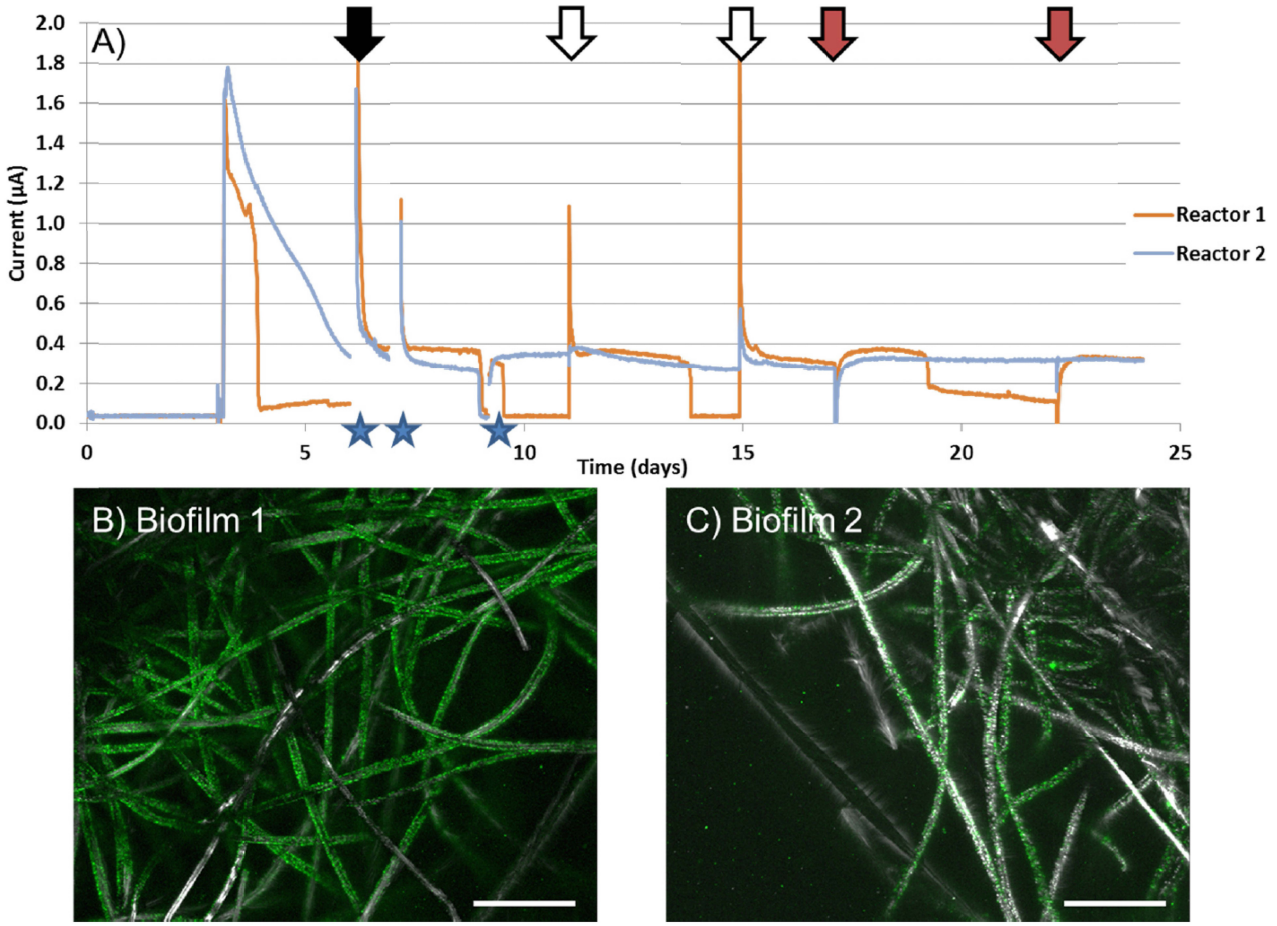


FIGURE 3 | (A) Current vs. time of duplicate MFC reactors with *C. ferrireducens* biofilms pregrown at open circuit anodes with glycerol as the electron donor and bicarbonate as the electron acceptor. **(B)** Confocal image of pregrown *C. ferrireducens* biofilm used for reactor 1 (with more biomass pregrown), stained with SYBR Green fluorescent DNA stain. **(C)** Confocal image of pregrown *C. ferrireducens* biofilm used for reactor 2 (with less biomass pregrown). *Stars* indicate time periods where the circuit was opened to measure anodic and cathodic open-circuit potentials. *Black arrows* indicate full media exchanges. *White arrows* indicate additions of glycerol and yeast extract only. *Red arrows* indicate addition of glycerol, yeast extract, and ferrihydrite simultaneously. Scale bars on **(B,C)** are 200 μm.

or fumarate as electron acceptors. Cell-free extracts of the cells grown under these conditions were analyzed by SDS-PAGE with gels stained for protein and *c*-type hemes. In each heme-stained profile of crude cell extracts, two or three major and several minor heme bands were recovered (**Figure 4**). Molecular masses of heme-stained proteins ranged from 40 to 120 kDa. In-gel tryptic digests followed by LC-MS analysis of the major heme bands revealed, that the most intense bands of ca. 80 kDa, recovered at all the tested culture conditions, contain the cytochromes Ga0395992_02_112574_114619 and Ga0395992_02_137389_139104 with 11 and eight predicted heme-binding motifs, respectively. In the cells grown with ferrihydrite or ferric citrate, the third intensively heme-stained band of ca. 50 kDa was revealed, containing the heptaheme cytochrome Ga0395992_02_433869_435545. In the cells grown with soluble ferric citrate, the fourth heme band of the cytochrome Ga0395992_03_307983_310529 of 110 kDa with 12 predicted heme-binding motifs was revealed. No other multiheme cytochromes were identified in the major heme bands on the gels regardless of the growth condition used. Four multihemes detected in the most intense heme bands were designated according to their predicted subcellular localization (**Table 1** and **Supplementary Table S5**): Ga0395992_02_112574_114619 was designated as OmhA ("Outer MultiHeme"), Ga0395992_02_137389_139104 as SmhA ("Secreted MultiHeme"), Ga0395992_03_307983_310529 as SmhB, and Ga0395992_02_433869_435545 as SmhC.

Proteomic Profiling of *C. ferrireducens* Cells Grown With Different Electron Acceptors

To assess the relevance of identified multihemes for EET and detect any accessory proteins involved in electron transfer to insoluble electron acceptors, the results of a shotgun proteomic analysis of crude extracts of cells, harvested at late logarithmic growth phase, were compared across four different cultivation conditions, each in three biological replicates: ferrihydrite

TABLE 1 | Putative EET-related multiheme cytochromes of *C. ferrireducens* and the summary of their best blast hits among the components of EET pathways in *Shewanella oneidensis* MR-1, *Geobacter sulfurreducens* PCA, and "*Thermincola potens*."

Protein symbol[1]	Locus tag	Predicted number of heme-binding motifs	Predicted[2] localization	Best blast hit	Min E-value	Max% Ident.	Organism	Hit gene name/ ORF number
	Ga0395992_01_182949_184436	22	Secreted	NP_953783.1	1.9E-7	34	*Geobacter*	*omcC omcBC orf2*
				YP_003639882	5.0E-30	28	*sulfurreducens*	
							Thermincola potens	TherJR_1117
Putative Q-oxidase	Ga0395992_01_187284_188528	5	Secreted	–				
Cytochromes encoded in the "pilin-cytochrome" cluster	Ga0395992_01_217646_219709	11	Secreted	NP_717047.1	1.8E-6	27	*Shewanella oneidensis*	*dmsE*
				YP_003639887	2.0E-12	28	*Thermincola potens*	TherJR_1122
	Ga0395992_01_220048_222057	11	Secreted	NP_951727.2	2.7E-6	33	*Geobacter sulfurreducens*	*omcX*
				YP_003639887	1.0E-17	25	*Thermincola potens*	TherJR_1122
	Ga0395992_01_222075_224618	9	Secreted	–				
	Ga0395992_02_29558_30178	6	Transmembrane, outer facing	YP_003639882	3.0E-22	30	*Thermincola potens*	TherJR_1117
	Ga0395992_02_30996_31874	7	Transmembrane, outer facing	NP_717386.1	4.0E-12	30	*Shewanella oneidensis*	*mtrA*
OmhA	Ga0395992_02_112574_114619	11	Secreted	ADG83432.1	4.0E-44	30	*Thermincola potens*	*ocwA*
NrfA	Ga0395992_02_129131_130396	5	Transmembrane, outer facing	–				
NrfH	Ga0395992_02_130419_130832	4	Secreted	ADG83432.1	2.0E-19	29	*Thermincola potens*	*ocwA*
Putative Q-oxidase	Ga0395992_02_135863_137242	20	Secreted	NP_953783.1	2.7E-6	29	*Geobacter sulfurreducens*	*omcC omcBC orf1 and 2*
				NP_953777.1	2.7E-6	29	*Thermincola potens*	TherJR_1117
				YP_003639882	2.0E-29	28		
SmhA	Ga0395992_02_137389_139104	8	Secreted	–				
	Ga0395992_02_147039_148391	14	Transmembrane, outer facing	YP_003639882	6.0E-57	33	*Thermincola potens*	TherJR_1117
	Ga0395992_02_151260_153788	12	Secreted	NP_719884.1	2.5E-6	45	*Shewanella oneidensis*	*mtrA*
	Ga0395992_01_477471_477917	4	Secreted	–				
SmhC	Ga0395992_02_433869_435545	7	Secreted	–				
SmhB	Ga0395992_03_307983_310529	12	Secreted	YP_003639887	7.0E-17	24	*Thermincola potens*	TherJR_1122
	Ga0395992_01_684739_686151	8	Secreted	–				
	Ga0395992_01_647313_648317	10	Transmembrane	YP_003639882	3.0E-81	42	*Thermincola potens*	TherJR_1117

[1]*Refer to the text for details.* [2]*Tested with Phobius, SignalP, TMHMM, TatP 1.0, SecretomeP, PsortB.*

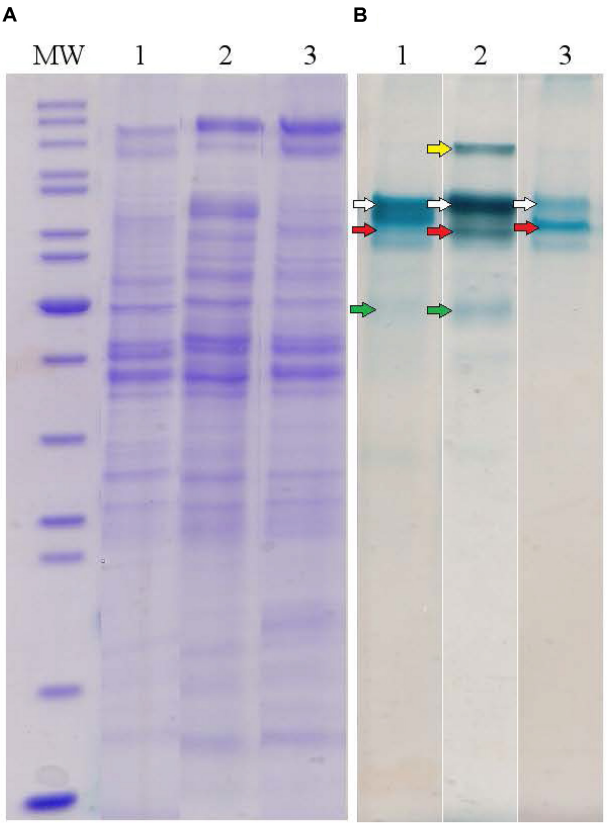


FIGURE 4 | Protein profiles of crude cell extracts of *C. ferrireducens* grown in the presence of ferrihydrite and alternative soluble electron acceptors. **(A)** Lanes1–3 show full-length Coomassie-stained gels for proteins. **(B)** Lanes 1–3 show the same gels, benzidine-stained to visualize *c*-type hemes. Lane numbers indicate different electron acceptors: 1—ferrihydrite; 2—ferric citrate; 3—fumarate. MW—molecular weight markers (top to bottom: 200, 150, 120, 100, 85, 70, 60, 50, 40, 30, 25, 20, 15, 10 kDa); *white arrows* indicate heme bands containing OmhA cytochrome, *red arrows* SmhA, *yellow arrow* SmhB, and *green arrows* SmhC (refer to the text for cytochromes' description). The Coomassie and benzidine stained gels received the same protein loading (13–14 μg per lane).

reduction, electrogenesis, soluble ferric citrate reduction, and fumarate reduction. In each case, the triplicate samples were taken from three different cultures or BES systems, respectively. Tandem MS revealed 1524 different proteins identified with a minimum of two unique peptides. The number of proteins identified per each sample is provided in **Supplementary Table S6**. We summarized mole proportions (riBAQs) of selected marker proteins, which determine housekeeping, respiratory, and EET metabolic processes. Selected proteins were divided into five groups by their predicted functions: group I contains subunits of DNA-directed RNA polymerase, group II—subunits of F_0F_1-type ATP synthase, group III—complexes I and II of the respiratory ETC, group IV involves structural and accessory proteins of type-IV pili assembly encoded in the "pilin-cytochrome" cluster, group V contains all the multiheme cytochromes which were predicted to be secreted or cell surface-associated proteins and were identified in crude cell extracts or proteomic profiles. The full set of marker proteins from all the five groups is highlighted in **Figures 5**, **6** and in **Supplementary Table S6**. Expression profiling (**Figure 5**) revealed similar proportion of RNA polymerase proteins in all the tested cell preparations [individual (ind.) *P*-values in ANOVA pairwise comparisons ranged from 0.43 to 0.81, **Supplementary Table S3**]. Profiles of the ATP-ase and respiratory complexes of the ETC are likely to reflect fluctuations in the activity of oxidative phosphorylation pathways when respiring different electron acceptors. In particular, the highest and lowest molar proportions of ATP-ase and ETC complexes are observed in the cells grown with the most and the least energetically favorable electron acceptors of *C. ferrireducens*, ferrihydrite, and fumarate, respectively (Thauer et al., 1977; Gavrilov et al., 2012). Expression levels of ATP-ase and ETC complexes are, respectively, 45 and 69% higher in ferrihydrite-grown cells comparing to fumarate-grown ones. However, only ATP-ase overexpression is statistically supported (ind. *P*-value 0.0017, **Supplementary Table S3**) due to high expression dissimilarity of individual proteins from the ETC complex (**Supplementary Table S6**). The highest proportion of multiheme cytochromes is clearly observed in ferrihydrite-grown cells (**Figure 5**), 93–304% difference in molar proportions of this group proteins was quantified when comparing ferrihydrite respiration conditions with the others tested (ind. *P*-values < 0.0004, **Supplementary Table S3**).

Profiling of individual multiheme proteins uncovered the same four predominant cytochromes that were identified in SDS-PAGE gels (OmhA, SmhA, SmhB, SmhC, **Figure 4**). These cytochromes are constitutively expressed in *C. ferrireducens* cells but their expression level differs significantly depending on the electron acceptor supplied (**Figure 6**), that is supported by ordinary two-way ANOVA tests retrieving $F(3,80)$ factor of 69.17 at a *P*-value < 0.0001 for electron acceptors as a source of variation (**Supplementary Table S2**). OmhA multiheme revealed the highest molar proportion among cytochromes at all the conditions tested with maximal share in ferrihydrite-grown cells (**Figure 6**), which is 237–400% higher than in other cells (ind. *P*-values < 0.15 for all the pairwise comparisons, **Supplementary Table S4**). SmhA cytochrome appeared to be less influenced by the electron acceptor utilized. Statistically supported upregulation was only observed in ferrihydrite-reducing and electrogenic cells comparing to ferric citrate-reducing ones (227 and 298% difference at ind. *P*-values 0.0099 and 0.0009, respectively). The cytochrome SmhC is clearly upregulated in the cells grown with Fe(III) compounds, both soluble and insoluble (from 846% difference in ferrihydrite-reducing vs. electrogenic cells to 6955% difference in citrate-reducing vs. fumarate-reducing cells at ind. *P*-values < 0.0001 for all pairwise comparisons, **Supplementary Table S4**). SmhB cytochrome is rather specific for the cells reducing soluble ferric citrate (**Figure 6**), its relative abundance increased by 2435% when comparing these cells with electrogenic ones, and by 25,000 and 57,000% when comparing with fumarate- and ferrihydrite-reducing cells, respectively, at ind. *P*-values < 0.0001 in all the cases (**Supplementary Table S4**). Among minor multihemes with molar proportion below 0.1 (**Figure 6**, *inset*), of note is constitutive production of putative

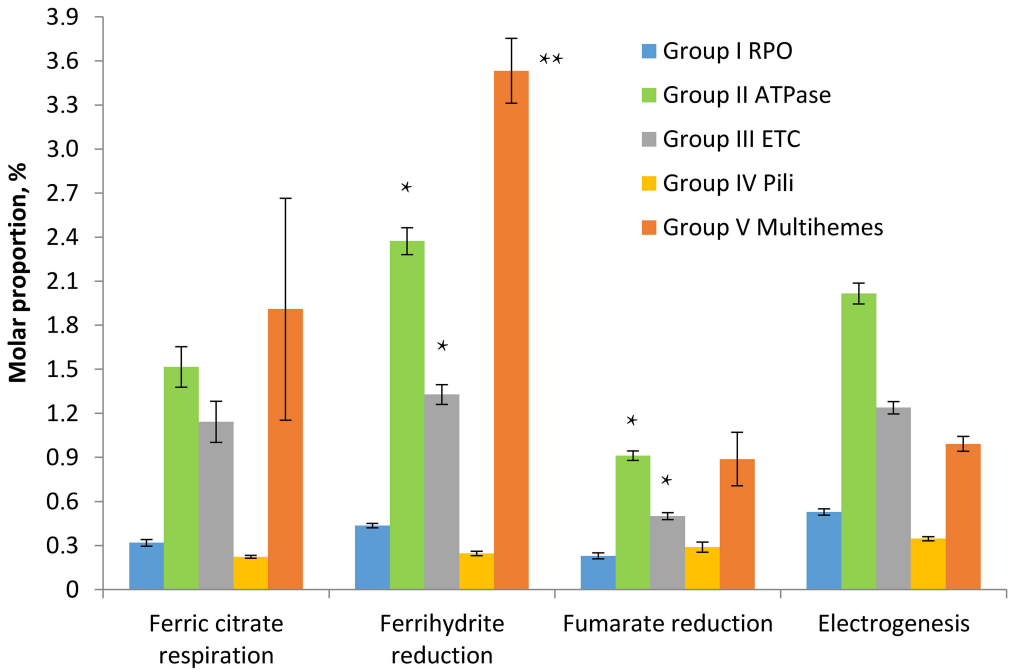


FIGURE 5 | Molar proportions (riBAQs) of selected protein groups determining housekeeping, respiratory, and EET processes in *C. ferrireducens* cells grown with different electron acceptors. For each protein group, summarized molar proportion (riBAQ) of all its proteins is presented. Whole set of proteins detected in a sample is taken as 100%. Error bars indicate standard deviations in summarized riBAQs of each group. Asterisks indicate the most significant, statistically supported differences in molar proportions of protein groups (refer to the text for detail). *Electrogenic* growth conditions imply utilization of a stainless steel anode as the electron acceptor. The cells at all the growth conditions were harvested at the late logarithmic growth phase. Target proteins are grouped as follows: *Group I RPO*—four proteins of α, β, β', and ω subunits of DNA-directed RNA polymerase, *Group II ATPase*—six proteins of α, β, γ, δ, ε, and B subunits of F_0F_1-type ATP synthase, *Group III ETC*—totally 17 proteins of A to F and H to N subunits of proton-translocating type I NADH-dehydrogenase together with four subunits of membrane-bound succinate dehydrogenase/fumarate reductase, *Group IV Pili*—10 proteins of pili assembly encoded in the "pilin-cytochrome" cluster ("PilA-C," PilBT, PilVMNO proteins, and three other proteins with type-IV pilin N-terminal methylation sites), *Group V Multihemes*—10 multiheme cytochromes which are predicted to be secreted or cell surface-associated proteins (refer to the legend of **Figure 6** for detail). The full set of the proteins included in each group appears in **Supplementary Table S6**.

quinol-oxidizing cytochromes Ga0395992_02_135863_137242 and Ga0395992_01_187284_188528, but not the CymA homolog Ga0395992_01_647313_648317. Also, a pair of adjacently encoded cytochromes Ga0395992_02_29558_30178 and Ga0395992_02_30996_31874, homologous to secreted periplasmic cytochromes of *Shewanella* (**Supplementary Table S5**), was expressed constitutively at a minor level. Surprisingly, the cytochromes Ga0395992_01_217646_219709 and Ga0395992_01_222075_224618, encoded in the "pilin-cytochrome" cluster, were only detected in electrogenic and ferric citrate-respiring cells and their expression profile across different electron acceptors (**Figure 6**, *inset*) did not correlate with that of the pilin assembly proteins (**Figure 5**).

OmhA Cytochrome Distribution in the Fractions of Ferrihydrite-Reducing Cultures

Magnetite produced by *C. ferrireducens* culture was separated from the cells and further washed with SDS. The wash contained only eight proteins identified in tryptic digests by LC-MS/MS (**Supplementary Table S7**). Among those proteins the only cytochrome was OmhA. It was also the only cytochrome identified in the fraction of proteins, sheared from the cell surface by non-destructive shaking (**Supplementary Table S7**). This result correlates with predicted localization of OmhA on the cell surface and with our previous finding that *C. ferrireducens* loses ferrihydrite-reducing activity after shearing off the cell surface proteins (Gavrilov et al., 2012). SDS-PAGE-MS/MS analysis revealed that OmhA is recovered in major heme bands in both soluble and insoluble protein fractions of crude extracts of the cells grown with ferrihydrite (**Supplementary Figure S5**). OmhA detection in the soluble protein fraction and easy shearing of the cytochrome from the cells probably result from its weak binding to the cell surface. From the other hand, the abundance of OmhA cytochrome in the fraction of insoluble proteins (presumably, membrane-associated) and on the surface of cells and biogenic magnetite they produced supports our proposal for the key role of this multiheme protein in EET chain of *C. ferrireducens*.

Phylogeny of the Major EET-Related Multiheme Cytochromes

Phylogenetic analysis of OmhA, SmhA, SmhB, and SmhC protein sequences revealed their relationship with diverse groups of cytochromes retrieved from both monoderm and

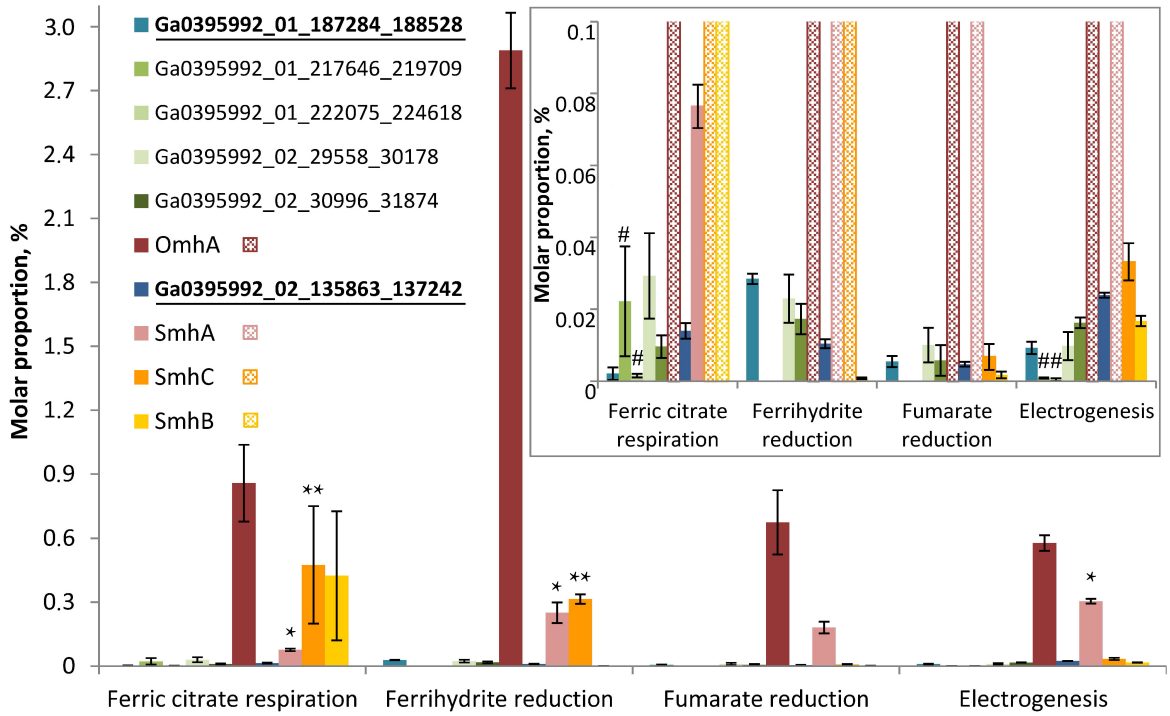


FIGURE 6 | Molar proportions (riBAQs) of individual multiheme cytochrome proteins. The data are shown for all the putative EET-related cytochromes of *C. ferrireducens*, which were identified in protein profiles: OmhA, SmhA, SmhB, SmhC cytochromes, Ga0395992_01_217646_219709 and Ga0395992_01_222075_224618 encoded in the "pilin-cytochrome" cluster, as well as cytochromes Ga0395992_02_29558_30178, Ga0395992_02_30996_31874, and putative quinol-oxidizing cytochromes Ga0395992_01_187284_188528 and Ga0395992_01_684739_686151. *Inset*—molar proportions of minor cytochromes (with riBAQs below 0.1), predominating cytochromes are cross-hatched in the inset. Putative Q-oxidases (Ga0395992_02_135863_137242 and Ga0395992_01_187284_188528) are marked bold and underlined. Asterisks indicate significant, statistically supported differences, which are not visually obvious on the figure. Singular asterisks indicate the differences in molar proportions of SmhA cytochrome in the cells grown with ferric citrate, ferrihydrite, and on anode; double asterisks indicate the differences in SmhC representation in ferric citrate-reducing or ferrihydrite-reducing cells (refer to the text for detail). Hash symbols indicate the cytochromes Ga0395992_01_217646_219709 and Ga0395992_01_222075_224618, encoded in the "pilin-cytochrome" cluster and expressed only in electrogenic and ferric citrate-respiring cells. Clear differences in molar proportions (such as for OmhA in ferrihydrite-reducing and other cells) are not specially marked.

diderm prokaryotes. Since the majority of these multihemes have no functional assignment yet, this blurs evolutionary traits of *C. ferrireducens*' proteins. Their distant relatedness with previously described EET-driving multihemes is clear (**Figure 7** and **Supplementary Figure S6**) as well as the more recent evolutionary history of the SmhA and SmhB cytochromes. The octaheme SmhA falls into a clade sharing the root with a large clade of octaheme tetrathionate reductases (Otr) from Gram-negative bacteria (**Supplementary Figure S6**). Dodecaheme SmhB and the 11-heme cytochrome Ga0395992_01_217646_219709, both of which are strongly upregulated under ferric citrate respiration (10^3–10^4% differences when comparing with other culture conditions, ind. *P*-values < 0.0001, **Supplementary Table S4** and **Figure 6**), belong to a clade, containing two other multihemes of *C. ferrireducens* and branching from a group of bacillary hexahemes (**Figure 7B**). More obscure is the phylogeny of OmhA and SmhC cytochromes, both upregulated upon ferrihydrite respiration (**Figure 6**). Those appear to be related to each other and belong to different subclades of proteins from Gram-positive bacteria containing 11 or 6–7 hemes, respectively (**Figure 7A**). In contrast, their distant relative, OcwA cytochrome of "*T. potens*," is the only protein from a monoderm organism in a subclade of nonahemes retrieved from Gram-negative bacteria (**Figure 7A**). However, three distinct clades enclosing the cytochromes OmhA, SmhC, and OcwA, respectively, share common roots with diverse cytochromes possessing 5–17 heme-binding sites and retrieved from various organisms, mainly uncultured.

DISCUSSION

Direct electron exchange between microorganisms and redox-active minerals plays a pivotal role in the biogeochemical cycles of metals. Understanding mechanisms of electron flow, especially those evolved in thermophilic Fe(III)-reducing prokaryotes closely related to ancestral forms of life, is essential for gaining insights into the processes that affected Earth's surface mineralogy in the Eoarchean Era and continue to influence it at present (Hazen et al., 2008). Our observations of magnetite crystallizing activity and magnetite-facilitated exoelectrogenesis in the

cultures of the thermophile *C. ferrireducens* support this notion. Active colonization of ferrihydrite by *C. ferrireducens* (**Figure 1** and **Supplementary Figure S1**), ferrihydrite transformation into unusually structured large magnetite crystals (**Figure 2** and **Supplementary Figure S3**) by growing cultures, and previously reported correlation between kinetics of microbial growth and Fe(III) reduction (Gavrilov et al., 2012) suggest biocontrolled formation of the novel mineral phase. This control could be realized via electron exchange between the mineral and redox enzymes, localized on the cell surface.

We previously reported outer surface *c*-type cytochromes to be important for ferrihydrite reduction in *C. ferrireducens*



FIGURE 7 | Continued

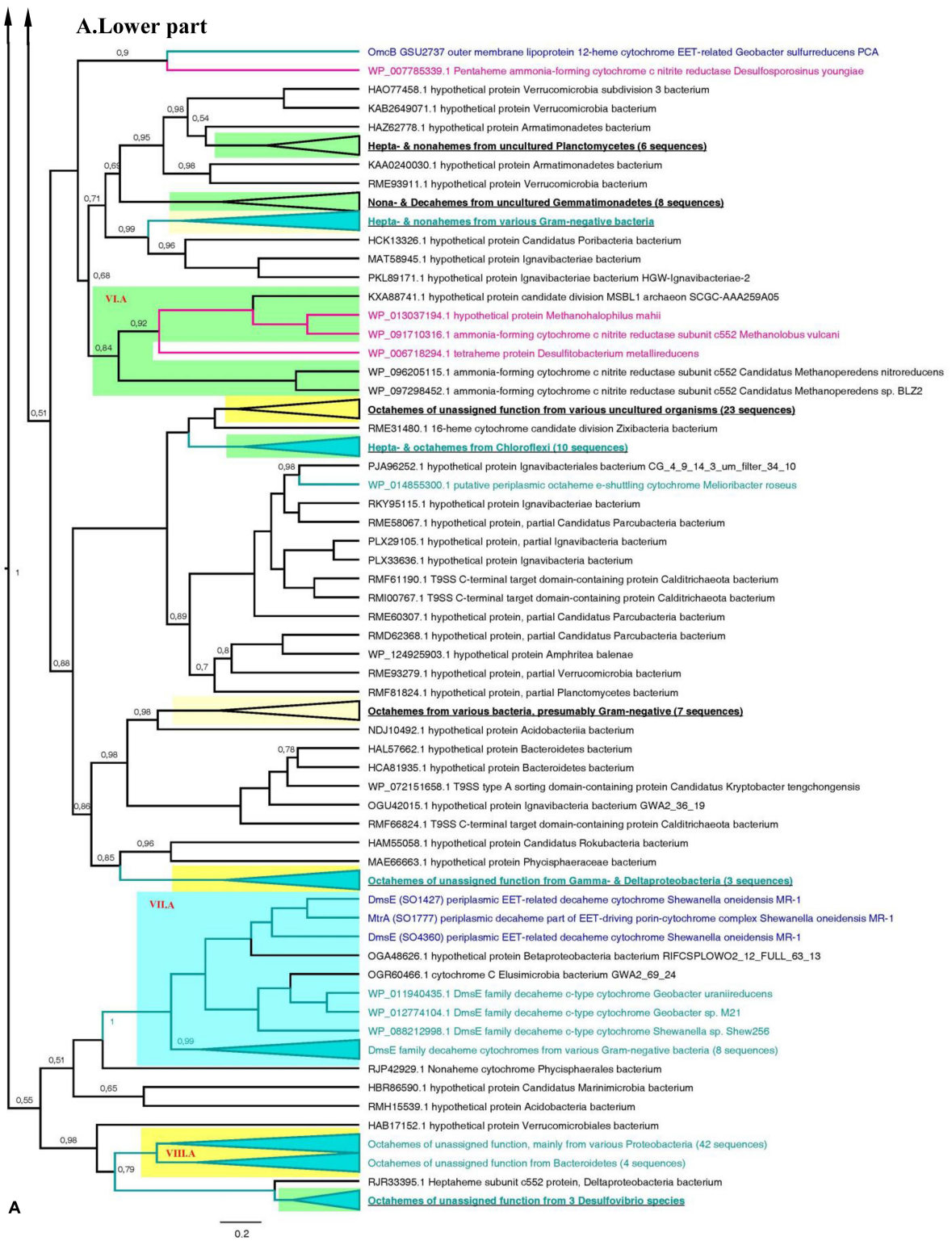


FIGURE 7 | Continued



FIGURE 7 | Consensus trees constructed after Bayesian inference of phylogeny from the MAFFT alignment of OmhA and SmhC **(A)**, and SmhB **(B)** cytochromes of *C. ferrireducens* and their best blast hits (see section "Materials and Methods" for detail). Unrooted 50% majority rule consensus phylograms are displayed as rectangular trees, for which posterior probability values are shown; posterior values of 1 are omitted for clarity. Mean branch lengths are characterized by scale bars indicating the evolutionary distance between the proteins (changes per amino acid position). The branches are annotated with labels indicating the protein sequence accession number, the protein name as retrieved from the database and the source organism. Branches and labels of the proteins are colored *pink* for the proteins retrieved from monoderm cultured organisms, *cyan* for the proteins from diderm cultured organisms, and *black* for the proteins from uncultured organisms. The target sequences of OmhA, SmhC, SmhB, and homologous cytochromes of *C. ferrireducens* are labeled *red* and their homologs from previously described EET-related cytochromes of other organisms are labeled *dark blue*. Clusters of sequences are categorized and highlighted where appropriate as follows. *Blue cluster* combines sequences of DmsE family decaheme cytochromes. *Light and deep yellow clusters* combine related sequences by the number of predicted heme-binding sites in them. *Green clusters* contain related sequences of various heme numbers from distinct taxonomic or physiological groups of source organisms. Highlighted clusters are labeled whether by collapsed subtrees with annotations given in bold underlined font or by red Roman numerals denoting the following. Cluster I.A—OmhA and related 11-heme proteins from *Carboxydothermus* spp.; cluster II.A—various multihemes from Acidobacteria; cluster III.A—octahemes from *Epsilonproteobacteria*; cluster IV.A—OcwA from "*T. potens*" and related nonaheme proteins; cluster V.A—SmhC and related hexa- and heptaheme proteins; cluster VI.A—9- to 16-heme proteins from monoderm prokaryotes, methanogenic archaea; cluster VII.A—see the *blue cluster* above; cluster VIII.A—octahemes of unassigned function from various Proteobacteria and Bacteroidetes; cluster I.B—SmhB and its homologs from *Carboxydothermus* species; cluster II.B—hexaheme proteins with unassigned function from various Firmicutes.

(Gavrilov et al., 2012). In the current work, we have identified candidate genes encoding such cytochromes and revealed the multiheme proteins with strongly upregulated expression in ferrihydrite-reducing cells. The most abundant of these cytochromes (OmhA) is predicted to be localized on the cell surface in a strategic position to provide reductive foci for direct interaction with the outermost Fe(III) atoms of ferrihydrite.

Detection of the OmhA protein in the cell-to-mineral interface, i.e., among the proteins sheared from surfaces of cells or biogenic magnetite (**Supplementary Table S7** and **Supplementary Figure S5**), indicate weak binding of the cytochrome to the cell envelope. We propose that this allows it to diffuse from the cell and rendezvous with remote mineral nanoparticles, donating electrons to them (as predicted for external cytochromes MtrC and OmcA of *S. oneidensis*) (Subramanian et al., 2018). Accordingly, the OmhA cytochrome could function either as a direct electron conduit, connecting the cellular energy transducing system with the extracellular mineral electron acceptor, or as an electron shuttle inside restricted volumes of the mineral cavities near the cell surface (shown in **Figure 2A**), where the losses of secreted compounds are limited, and the protein could transfer electrons to remote mineral particles inaccessible for cell surface structures.

Additionally, the EET pathway in *C. ferrireducens* could indirectly involve pili-like filamentous appendages forming complex networks, which connect the cells with Fe-containing nanoparticles (**Figure 1C**). These proteinaceous structures are non-conductive (**Supplementary Figure S2**) and rather serve for cell-to-mineral anchoring or for aligning the mineral nanoparticles together along the filaments in microenvironments surrounding the cells. This function could be important during initial steps of colonial growth and nucleation of magnetite crystals within the colonies. During further "big scale" crystallization, the redox mineral becomes more readily accessible to the cell surface cytochromes.

Micron-scale magnetite crystals can serve as more effective mediators of direct interspecies electron transfer than nanoparticles in natural anoxic sedimentary environments. The electrons could be tunneled from a single reductive focus on the cell surface of *C. ferrireducens* to the bulk of the conductive mineral and further, to several symbiotic organisms, which could utilize the same magnetite crystal as the electron donor, as methanogens (Shi et al., 2016; Zhuang et al., 2018) or anoxigenic phototrophs (Byrne et al., 2015) do. Thus, enhancing the formation of large mineral structures is likely to be energetically advantageous for Fe(III)-respiring microbial colonies. Exoelectrogenic activity of *C. ferrireducens* (**Figure 3**) would allow the organism to discharge electrons to the conductive biogenic mineral. The proposed efficiency of electron transfer out of *C. ferrireducens* cells through magnetite crystals is supported by a stabilizing (**Figure 3A**) effect of ferrihydrite particles on electrogenic activity of the organism. Previously, several reports indicated such an effect of solid-phase Fe(III/II) nanoparticles in mesophilic MFCs (Kato et al., 2010, 2012).

Interestingly, among the wide diversity of cultured prokaryotes capable of EET, only Gram-positive Fe(III) reducing clostridia of the genus *Thermoanaerobacter* have been reported to induce formation of micron-scale magnetite crystals (Zhang et al., 1998; Li, 2012). In contrast, nanoparticulate biogenic magnetite forms are well characterized in Fe(III) reducers of various phylogeny and physiology, and the microorganisms are regarded to have little, if any, control over nanoparticles biomineralization (Jimenez-Lopez et al., 2010). The first report on large biogenic magnetite structures described 15–40 μm octahedron and cubic crystals recovered from a thermophilic enrichment dominated by Firmicutes (Slobodkin et al., 1995). Internal structure of this mineral has not been studied, and no imprints or invasion of bacterial cells have been observed so far in this (Slobodkin et al., 1995) or any other (Zhang et al., 1998; Jimenez-Lopez et al., 2010; Li, 2012) biogenic magnetites. Our results provide the first observations of these fascinating minerals. Considering the abundance of Gram-positive bacteria among thermophiles capable of EET (Wrighton et al., 2008), their ability to enhance magnetite crystallization and benefit from this process may increase the overall efficiency of energy conservation from metal reduction in thermophilic microbial communities compared to their mesophilic counterparts.

Predictably, the most probable determinants of EET we have identified in *C. ferrireducens* by genomic and proteomic approaches (**Figure 6** and **Supplementary Table S5**) appear to be multiheme *c*-type cytochromes. This finding supports previous proposals on global distribution of multiheme-related EET pathways throughout the prokaryotic world (Shi et al., 2016) and argues against an alternative concept on flavin-based cytochrome-independent EET in Gram-positive mesophiles (Light et al., 2018; Pankratova et al., 2019). Indeed, in our experiments at all the tested culture conditions, the total proportion of secreted or cell-surface associated multiheme cytochromes was comparable with that of ATP-ase or ETC complexes, pointing out the equivalent significance of the multihemes and basic components of oxidative phosphorylation pathways for the metabolism of *C. ferrireducens* during its growth with Fe(III) compounds, anodes, and even soluble electron acceptors (**Figure 5**). Low expression of multihemes in electrogenic cells comparing to those grown with ferrihydrite (-78% difference, ind. P-value < 0.0001) probably reflects their biofilm lifestyle (**Figures 3B,C** and **Supplementary Figures S4A–C**), which imposes spatial constraints on the distribution of terminal oxidoreductases over the cell surface. The cells in the biofilm, immobilized on the anode surface, contact this electron acceptor with a restricted area of their envelope. Only the cytochrome molecules located in this restricted area remain useful for anode respiration. In our experiments, the highest current density generated by anode-respiring *C. ferrireducens* cells was comparatively low (68 mA/m^2), indicative of one layer ($\leq$2 μm) thick electrogenic biofilm, suggesting EET via direct contact of cell wall-associated cytochromes with the anode (Torres et al., 2010). In contrast, ferrihydrite-reducing cells, which are surrounded by nanoparticles or larger crystal structures of their electron

acceptor (**Figures 1**, **2**), could employ larger quantities of multiheme molecules, distributed all over the cell surface, to get advantage of using all the available cell-to-mineral contact points for electron transfer to Fe(III). Surprisingly, the cytochromes encoded in the proposed "pilin-cytochrome" cluster had negligible expression levels in all the tested cultures. This correlates with the absence of conductivity in pili-like filaments inferred both from PilA sequence analysis and CP-AFM measurements (**Supplementary Figure S2**). However, the observed production of filamentous networks in magnetite-forming *C. ferrireducens* cultures (**Figures 1C,D**) necessitates further investigation of the nature and physiological role of these extracellular filaments.

Multiheme cytochromes profiles, and particularly, constitutive production and the strongest positive expression response of OmhA to ferrihydrite (237–400% difference comparing to other culture conditions, ind. *P*-values < 0.0001, **Supplementary Table S4** and **Figure 6**), allow us to conclude that this cytochrome is the key protein for EET to ferrihydrite and is likely to be the major multiheme protein, determining the reduction of other electron acceptors tested, including stainless steel anodes. At the same time, differential expression of putative secreted SmhB and SmhC cytochromes in response to the electron acceptor utilized indicates modularity of the respiratory pathway in *C. ferrireducens*, which implies constitutive activity of core electron transferring multihemes, such as OmhA, SmhA, and putative quinol-oxidizing cytochromes Ga0395992_01_187284_188528 and Ga0395992_02_135863_137242, supplemented with the activity of specific multihemes targeting electrons to different electron acceptors. The prevalence of OmhA and SmhA multihemes among other cytochromes at three of the four culture conditions tested supports probable involvement of these proteins in several EET pathways. Indeed, as illustrated by **Figure 6**, OmhA had the highest molar proportion among multiheme cytochromes at all the culture conditions, ranging from 0.577% in electrogenic cells to 2.89% in ferrihydrite-respiring cells. SmhA had the second biggest molar proportion among cytochrome proteins in electrogenic and fumarate-respiring cells (0.304 and 0.180, respectively) and was almost equally represented with SmhC cytochrome in ferrihydrite-reducing cells (0.250 and 0.314% for SmhA and SmhC, respectively).

High expression level of OmhA and SmhA oxidoreductases could be sustained constitutively by a regulome of *C. ferrireducens* in order to compensate for electron acceptors that fluctuate rapidly in their availability in the natural environment of the organism. Such a metabolic strategy could represent a widespread response of microorganisms to sharp changes in the availability of electron acceptors. In particular, constitutive expression has been reported for different terminal oxidoreductases of a thermophilic iron-reducer *Melioribacter roseus* from a deep subsurface aquifer (Gavrilov et al., 2017), as well as for several multihemes involved in electron transfer to different electron acceptors in a mesophilic iron-reducing soil bacterium *Geobacter soli* (Cai et al., 2018). A modular principle of the respiratory system organization, which implies differential activity of various multihemes, has been previously proposed for *S. oneidensis* (Coursolle and Gralnick, 2010) and *Geobacter* species (Butler et al., 2010; Aklujkar et al., 2013; Cai et al., 2018). However, the key proteins of the EET chain in *C. ferrireducens* differ significantly from their previously described counterparts of *Geobacter*, *Shewanella*, and *Thermincola* species (**Figure 7** and **Supplementary Figure S6**). Even putative terminal reductases OcwA from "*T. potens*" (Costa et al., 2019) and OmhA from *C. ferrireducens*, sharing considerable sequence similarity (**Table 1** and **Supplementary Table S5**), fall into different clades of multihemes which diverged early from their common ancestor (**Figure 7A**). Moreover, it remains unclear if this ancestor protein belonged to a monoderm or diderm prokaryote.

Our phylogenetic reconstructions did not reveal any distinct evolutionary traits which have been recently proposed for EET-related cytochromes of thermophilic monoderm prokaryotes (Lusk, 2019). Instead, our data correlate with previous statements that gene truncation or duplication followed by sequence divergence and active horizontal gene transfer is the main mechanism of multiheme cytochromes evolution which is rather independent of the cell envelope structure of an organism and is mainly driven by its metabolic needs (Sharma et al., 2010; Kern et al., 2011). In the case of *C. ferrireducens*, an ancestor octaheme protein of SmhA cytochrome is likely to be acquired by horizontal gene transfer from a diderm bacterium (**Supplementary Figure S6**), while 12-heme SmhB cytochrome seems to originate from an ancestral monoderm bacterium by gene duplication and further vertical evolution (**Figure 7B**). Evolutionary history of OmhA and SmhC proteins is more complicated and probably involves multiple events driven by the need of *Carboxydothermus* species and their ancestors to effectively utilize high-potential insoluble electron acceptors, such as Fe(III) minerals.

CONCLUSION

Newly identified secreted multiheme cytochromes OmhA and SmhA in *C. ferrireducens* have been shown to be involved in ferrihydrite to magnetite transformation and exoelectrogenesis, while a distinct cytochrome SmhB is likely involved in the reduction of soluble Fe(III). All these cytochromes are not associated with porin–cytochrome complexes or pilin-cytochrome assemblies, are phylogenetically distinct from the known external EET-driving cytochromes and thus represent novel extracellular electron transferring proteins. Further structural and functional characterization is planned to trace and reconstruct the pathways for cell-to-mineral electron flow in thermal environments inhabited by diverse thermophiles.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

SG and EB-O convened the research. SG, DZ, and IE designed and performed cultivation experiments and mineral phase analysis. VP, TT, and ND designed and performed protein separation and identification experiments. JL and DK designed and performed the experiments on cytochromes identification in microbe-to-magnetite interphase. ME-N, KL, and SP designed and performed SEM, FIB, and conductive AFM studies. MZ analyzed proteomic data. OB designed and performed MFC experiments. SG and FR analyzed genomic data. SG and IE performed phylogeny reconstructions. SG, DZ, TT, JL, ME-N, FR, OB, and EB-O analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Russian Science Foundation project No. 17-74-30025 (cultivation experiments, electrochemical and mineralogical studies) and No. 20-14-00314 (fractionation and target proteins identification). SG, DZ, IE, ND, TT, VP, and EB-O also acknowledge support from the Russian Ministry of Science and Higher Education. FTR was supported by grants from the NASA Exobiology Program and the C-DEBI consortium funded by the US National Science Foundation. ME-N was supported by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy through grant DE-FG02-13ER16415.

ACKNOWLEDGMENTS

FTR acknowledges the support from the US Department of Energy Joint Genome Institute and kind collaboration from Barney Whitman and Tanja Woyke who enabled the sequencing of the *C. ferrireducens* genome. Potentiostatic biofilms (**Supplementary Figures S4A,B**) were imaged using the equipment of the Core Centrum of Koltsov Institute of Developmental Biology RAS. We are thankful to Alexandr D. Finoshin and Dr. Yulia V. Lupina for this imaging.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.597818/full#supplementary-material

REFERENCES

Aklujkar, M., Coppi, M. V., Leang, C., Kim, B. C., Chavan, M. A., Perpetua, L. A., et al. (2013). Proteins involved in electron transfer to Fe(III) and Mn(IV) oxides by *Geobacter sulfurreducens* and *Geobacter uraniireducens*. *Microbiology* 159, 515–535. doi: 10.1099/mic.0.064089-0

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.

Bockrath, M., Markovic, N., Shepard, A., Tinkham, M., Gurevich, L., Kouwenhoven, L. P., et al. (2002). Scanned conductance microscopy of carbon nanotubes and λ-DNA. *Nano Lett.* 2, 187–190. doi: 10.1021/nl0100724

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650

Bretschger, O., Obraztsova, A., Sturm, C. A., In, S. C., Gorby, Y. A., Reed, S. B., et al. (2007). Current production and metal oxide reduction by *Shewanella oneidensis* MR-1 wild type and mutants. *Appl. Environ. Microbiol.* 73, 7003–7012. doi: 10.1128/aem.01087-07

Butler, J. E., Young, N. D., and Lovley, D. R. (2010). Evolution of electron transfer out of the cell: comparative genomics of six *Geobacter* genomes. *BMC Genomics* 11:40. doi: 10.1186/1471-2164-11-40

Byrne, J. M., Klueglein, N., Pearce, C., Rosso, K. M., Appel, E., and Kappler, A. (2015). Redox cycling of Fe(II) and Fe(III) in magnetite by Fe-metabolizing bacteria. *Science* 347, 1473–1476. doi: 10.1126/science.aaa4834

Cai, X., Huang, L., Yang, G., Yu, Z., Wen, J., and Zhou, S. (2018). Transcriptomic, proteomic, and bioelectrochemical characterization of an exoelectrogen *Geobacter soli* grown with different electron acceptors. *Front. Microbiol.* 9:1075. doi: 10.3389/fmicb.2018.01075

Carlson, H. K., Iavarone, A. T., Gorur, A., Yeo, B. S., Tran, R., Melnyk, R. A., et al. (2012). Surface multiheme c-type cytochromes from *Thermincola potens* and implications for respiratory metal reduction by Gram-positive bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1702–1707. doi: 10.1073/pnas.1112905109

Costa, N. L., Carlson, H. K., Coates, J. D., Louro, R. O., and Paquete, C. M. (2015). Heterologous expression and purification of a multiheme cytochrome from a Gram-positive bacterium capable of performing extracellular respiration. *Protein Expr. Purif.* 111, 48–52. doi: 10.1016/j.pep.2015.03.007

Costa, N. L., Hermann, B., Fourmond, V., Faustino, M. M., Teixeira, M., Einsle, O., et al. (2019). How thermophilic Gram-positive organisms perform extracellular electron transfer: characterization of the cell surface terminal reductase OcwA. *mBio* 10:e01210-19. doi: 10.1128/mBio.01210-19

Coursolle, D., and Gralnick, J. A. (2010). Modularity of the Mtr respiratory pathway of *Shewanella oneidensis* strain MR-1. *Mol. Microbiol.* 77, 995–1008.

Dopson, M., Ni, G., and Sleutels, T. H. J. A. (2016). Possibilities for extremophilic microorganisms in microbial electrochemical systems. *FEMS Microbiol. Rev.* 40, 164–181. doi: 10.1093/femsre/fuv044

El-Naggar, M. Y., Wanger, G., Leung, K. M., Yuzvinsky, T. D., Southam, G., Yang, J., et al. (2010). Electrical transport along bacterial nanowires from *Shewanella oneidensis* MR-1. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18127–18131. doi: 10.1073/pnas.1004880107

Filman, D., Marino, S. F., Ward, J. E., Yang, L., Mester, Z., Bullit, E., et al. (2019). Cryo-EM reveals the structural basis of long-range electron transport in a cytochrome-based bacterial nanowire. *Commun. Biol.* 2:219. doi: 10.1038/s42003-019-0448-9

Francis, R. T., and Becker, R. R. (1984). Specific indication of hemoproteins in polyacrylamide gels using a double-staining process. *Anal. Biochem.* 136, 509–514. doi: 10.1016/0003-2697(84)90253-7

Gavrilov, S., Podosokorskaya, O., Alexeev, D., Merkel, A., Khomyakova, M., Muntyan, M., et al. (2017). Respiratory pathways reconstructed by multi-omics analysis in *Melioribacter roseus*, residing in a deep thermal aquifer of the West-Siberian Megabasin. *Front. Microbiol.* 8:1228. doi: 10.3389/fmicb.2017.01228

Gavrilov, S. N., Lloyd, J. R., Kostrikina, N. A., and Slobodkin, A. I. (2012). Fe(III) oxide reduction by a Gram-positive thermophile: physiological mechanisms for dissimilatory reduction of poorly crystalline Fe(III) oxide by a thermophilic Gram-positive bacterium *Carboxydothermus ferrireducens*. *Geomicrobiol. J.* 29, 804–819. doi: 10.1080/01490451.2011.635755

Hazen, R. M., Papineau, D., Downs, R. T., Ferry, J. M., Mccoy, T. J., Sverjensky, D. A., et al. (2008). Mineral evolution. *Am. Mineral.* 93, 1693–1720.

Holmes, D. E., Dang, Y., Walker, D. J. F., and Lovley, D. R. (2016). The electrically conductive pili of *Geobacter* species are a recently evolved feature for

extracellular electron transfer. *Microb. Genom.* 2:e000072. doi: 10.1099/mgen.0.000072

Jimenez-Lopez, C., Romanek, C. S., and Bazylinski, D. A. (2010). Magnetite as a prokaryotic biomarker: a review. *J. Geophys. Res. Biogeosci.* 115:G00G03. doi: 10.1029/2009JG001152

Kashefi, K., Tor, J. M., Holmes, D. E., Gaw Van Praagh, C. V., Reysenbach, A. L., and Lovley, D. R. (2002). *Geoglobus ahangari* gen. nov., sp. nov., a novel hyperthermophilic archaeon capable of oxidizing organic acids and growing autotrophically on hydrogen with Fe(III) serving as the sole electron acceptor. *Int. J. Syst. Evol. Microbiol.* 52, 719–728. doi: 10.1099/ijs.0.01953-0

Kato, S., Hashimoto, K., and Watanabe, K. (2012). Microbial interspecies electron transfer via electric currents through conductive minerals. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10042–10046. doi: 10.1073/pnas.1117592109

Kato, S., Nakamura, R., Kai, F., Watanabe, K., and Hashimoto, K. (2010). Respiratory interactions of soil bacteria with (semi) conductive iron-oxide minerals. *Environ. Microbiol.* 12, 3114–3123. doi: 10.1111/j.1462-2920.2010.02284.x

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Kavanagh, P., Botting, C. H., Jana, P. S., Leech, D., and Abram, F. (2016). Comparative proteomics implicates a role for multiple secretion systems in electrode-respiring *Geobacter sulfurreducens* biofilms. *J. Proteome Res.* 15, 4135–4145. doi: 10.1021/acs.jproteome.5b01019

Kern, M., Klotz, M. G., and Simon, J. (2011). The *Wolinella succinogenes* mcc gene cluster encodes an unconventional respiratory sulphite reduction system. *Mol. Microbiol.* 82, 1515–1530. doi: 10.1111/j.1365-2958.2011.07906.x

Krukenberg, V., et al. (2018). Gene expression and ultrastructure of meso- and thermophilic methanotrophic consortia. *Environ. Microbiol.* 20, 1651–1666. doi: 10.1111/1462-2920.14077

Li, Y.-L. (2012). Hexagonal platelet-like magnetite as a biosignature of thermophilic iron-reducing bacteria and its applications to the exploration of the modern deep, hot biosphere and the emergence of iron-reducing bacteria in early Precambrian oceans. *Astrobiology* 12, 1100–1108. doi: 10.1089/ast.2012.0847

Light, S. H., Su, L., Rivera-Lugo, R., Cornejo, J. A., Louie, A., Iavarone, A. T., et al. (2018). A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. *Nature* 562, 140–157. doi: 10.1038/s41586-018-0498-z

Lovley, D. R., and Malvankar, N. S. (2015). Seeing is believing: novel imaging techniques help clarify microbial nanowire structure and function. *Environ. Microbiol.* 17, 2209–2215. doi: 10.1111/1462-2920.12708

Lovley, D. R., and Phillips, E. J. P. (1988). Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl. Environ. Microbiol.* 54, 1472–1480. doi: 10.1128/aem.54.6.1472-1480.1988

Lusk, B. G. (2019). Thermophiles; or, the modern Prometheus: the importance of extreme microorganisms for understanding and applying extracellular electron transfer. *Front. Microbiol.* 10:818. doi: 10.3389/fmicb.2019.00818

Manzella, M. P., Holmes, D. E., Rocheleau, J. M., Chung, A., Reguera, G., and Kashefi, K. (2015). The complete genome sequence and emendation of the hyperthermophilic, obligate iron-reducing archaeon "*Geoglobus ahangari*" strain 234T. *Stand. Genomic Sci.* 10:77. doi: 10.1186/s40793-015-0035-8

Mardanov, A. V., Slododkina, G. B., Slobodkin, A. I., Beletsky, A. V., Gavrilov, S. N., Kublanov, I. V., et al. (2015). The *Geoglobus acetivorans* genome: Fe(III) reduction, acetate utilization, autotrophic growth, and degradation of aromatic compounds in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 81, 1003–1012. doi: 10.1128/aem.02705-14

Mehta, T., Coppi, M. V., Childers, S. E., and Lovley, D. R. (2005). Outer membrane c-type cytochromes required for Fe(III) and Mn(IV) oxide reduction in *Geobacter sulfurreducens*. *Appl. Environ. Microbiol.* 71, 8634–8641. doi: 10.1128/aem.71.12.8634-8641.2005

Nealson, K. H. (2017). Bioelectricity (electromicrobiology) and sustainability. *Microb. Biotechnol.* 10, 1114–1119. doi: 10.1111/1751-7915.12834

Nevin, K. P., Kim, B., Glaven, R. H., Johnson, J. P., Woodard, T. L., Didonato, R. J., et al. (2009). Anode biofilm transcriptomics reveals outer surface components essential for high density current production in *Geobacter sulfurreducens* fuel cells. *PLoS One* 4:e5628. doi: 10.1371/journal.pone.0005628

Pankratova, G., Hederstedt, L., and Gorton, L. (2019). Extracellular electron transfer features of Gram-positive bacteria. *Anal. Chim. Acta* 1076, 32–47. doi: 10.1016/j.aca.2019.05.007

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. doi: 10.1093/nar/gky1106

Refojo, P. N., Ribeiro, M. A., Calisto, F., Teixeira, M., and Pereira, M. M. (2013). Structural composition of alternative complex III: variations on the same theme. *Biochim. Biophys. Acta Bioenerg.* 1827, 1378–1382. doi: 10.1016/j.bbabio.2013.01.001

Rothery, R. A., Workun, G. J., and Weiner, J. H. (2008). The prokaryotic complex iron–sulfur molybdoenzyme family. *Biochim. Biophys. Acta Biomembr.* 1778, 1897–1929. doi: 10.1016/j.bbamem.2007.09.002

Schicklberger, M., Sturm, G., and Gescher, J. (2013). Genomic plasticity enables a secondary electron transport pathway in *Shewanella oneidensis*. *Appl. Environ. Microbiol.* 79, 1150–1159. doi: 10.1128/aem.03556-12

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Sharma, S., Cavallaro, G., and Rosato, A. (2010). A systematic investigation of multiheme c-type cytochromes in prokaryotes. *J. Biol. Inorg. Chem.* 15, 559–571. doi: 10.1007/s00775-010-0623-4

Shi, L., Dong, H., Reguera, G., Beyenal, H., Lu, A., Liu, J., et al. (2016). Extracellular electron transfer mechanisms between microorganisms and minerals. *Nat. Rev. Microbiol.* 14, 651–652. doi: 10.1038/nrmicro.2016.93

Shi, L., Fredrickson, J. K., and Zachara, J. M. (2014). Genomic analyses of bacterial porin-cytochrome gene clusters. *Front. Microbiol.* 5:657. doi: 10.3389/fmicb.2014.00657

Shin, J. B., Krey, J. F., Hassan, A., Metlagel, Z., Tauscher, A. N., Pagana, J. M., et al. (2013). Molecular architecture of the chick vestibular hair bundle. *Nat. Neurosci.* 16:365. doi: 10.1038/nn.3312

Slobodkin, A. I., Eroshchev-Shak, V. A., Kostrikina, N. A., Lavrushin, V. Y., Dainyak, L. G., and Zavarzin, G. A. (1995). Magnetite formation by thermophilic anaerobic microorganisms. *Dokl. Akad. Nauk [In Russian].* 345, 694–697.

Slobodkin, A. I., Reysenbach, A., Strutz, N., Dreier, M., and Wiegel, J. (1997). Thermophilic anaerobic dissimilatory Fe(III)-reducing bacterium from a continental hot spring. *Int. J. Syst. Bacteriol.* 47, 541–547. doi: 10.1099/00207713-47-2-541

Slobodkin, A. I., Sokolova, T. G., Lysenko, A. M., and Wiegel, J. (2006). Reclassification of *Thermoterrabacterium ferrireducens* as *Carboxydothermus ferrireducens* comb. nov., and emended description of the genus *Carboxydothermus*. *Int. J. Syst. Evol. Microbiol.* 56, 2349–2351. doi: 10.1099/ijs.0.64503-0

Smith, J. A., Aklujkar, M., Risso, C., Leang, C., Giloteaux, L., and Holmes, D. E. (2015). Mechanisms involved in Fe(III) respiration by the hyperthermophilic archaeon *Ferroglobus placidus*. *Appl. Environ. Microbiol.* 81, 2735–2744. doi: 10.1128/aem.04038-14

Subramanian, P., Pirbadian, S., El-Naggar, M. Y., and Jensen, G. J. (2018). Ultrastructure of *Shewanella oneidensis* MR-1 nanowires revealed by electron cryotomography. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3246–E3255.

Thauer, R. K., Jungermann, K., and Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* 41, 100–180. doi: 10.1128/mmbr.41.1.100-180.1977

Torres, C. I., Marcus, A. K., Lee, H.-S., Parameswaran, P., Krajmalnik-Brown, R., and Rittmann, B. E. (2010). A kinetic perspective on extracellular electron transfer by anode-respiring bacteria. *FEMS Microbiol. Rev.* 34, 3–17. doi: 10.1111/j.1574-6976.2009.00191.x

Toshchakov, S. V., Lebedinsky, A. V., Sokolova, T. G., Zavarzina, D. G., Korzhenkov, A. A., Teplyuk, A. V., et al. (2018). Genomic insights into energy metabolism of *Carboxydocella thermautotrophica* coupling hydrogenogenic CO oxidation with the reduction of Fe(III) minerals. *Front. Microbiol.* 9:1759. doi: 10.3389/fmicb.2018.01759

Walker, D. J. F., Adhikari, R. Y., Holmes, D. E., Ward, J. E., Woodard, T. L., Nevin, K. P., et al. (2018). Electrically conductive pili from pilin genes of phylogenetically diverse microorganisms. *ISME J.* 12, 48–58. doi: 10.1038/ismej.2017.141

Walker, D. J. F., Martz, E., Holmes, D. E., Zhou, Z., Nonnenmann, S. S., and Lovley, D. R. (2019). The archaellum of *Methanospirillum hungatei* is electrically conductive. *mBio* 10:e00579-19. doi: 10.1128/mBio.00579-19

Wang, F., Gu, Y., O'Brien, J. P., Yi, S. M., Yalcin, S. E., Srikanth, V., et al. (2019). Structure of microbial Nanowires reveals stacked Hemes that transport electrons over micrometers. *Cell* 177, 361–369. doi: 10.1016/j.cell.2019.03.029

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

White, G. F., Edwards, M. J., Gomez-Perez, L., Richardson, D. J., Butt, J. N., and Clarke, T. A. (2016). "Mechanisms of bacterial extracellular electron exchange," in *Advances in Microbial Physiology. Advances in Bacterial Electron Transport Systems and Their Regulation*, ed. R. K. Poole (Cambridge, MA: Academic Press), 87–138. doi: 10.1016/bs.ampbs.2016.02.002

Wrighton, K. C., Agbo, P., Warnecke, F., Weber, K. A., Brodie, E. L., DeSantis, T. Z., et al. (2008). A novel ecological role of the Firmicutes identified in thermophilic microbial fuel cells. *ISME J.* 2, 1146–1156. doi: 10.1038/ismej.2008.48

Xiong, Y., Shi, L., Chen, B., Mayer, M. U., Lower, B. H., Londer, Y., et al. (2006). High-affinity binding and direct electron transfer to solid metals by the *Shewanella oneidensis* MR-1 outer membrane c-type cytochrome OmcA. *J. Am. Chem. Soc.* 128, 13978–13979. doi: 10.1021/ja063526d

Zhang, C., Vali, H., Romanek, C. S., Phelps, T. J., and Liu, S. V. (1998). Formation of single-domain magnetite by a thermophilic bacterium. *Am. Mineral.* 83, 1409–1418. doi: 10.2138/am-1998-11-1230

Zhang, H., Tang, X., Munske, G. R., Zakharova, N., Zheng, C., Wolff, M. A., et al. (2008). In vivo identification of outer membrane protein OmcA-MtrC interaction network in *Shewanella oneidensis* MR-1 cells using novel chemical cross-linkers. *J. Proteome Res.* 7, 1712–1720. doi: 10.1021/pr7007658

Zhuang, L., Ma, J., Yu, Z., Wang, Y., and Tang, J. (2018). Magnetite accelerates syntrophic acetate oxidation in methanogenic systems with high ammonia concentrations. *Microb. Biotechnol.* 11, 710–720. doi: 10.1111/1751-7915.13286

Conflict of Interest: 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 © 2021 Gavrilov, Zavarzina, Elizarov, Tikhonova, Dergousova, Popov, Lloyd, Knight, El-Naggar, Pirbadian, Leung, Robb, Zakhartsev, Bretschger and Bonch-Osmolovskaya. 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: 12 January 2021
doi: 10.3389/fmicb.2020.609996

The Transcriptomic and Phenotypic Response of the Melanized Yeast *Exophiala dermatitidis* to Ionizing Particle Exposure

*Zachary Schultzhaus[1], Amy Chen[2], Igor Shuryak[3] and Zheng Wang[1]**

[1] Center for Biomolecular Science and Engineering, United States Naval Research Laboratory, Washington, DC, United States, [2] Virginia Tech Carilion School of Medicine, Roanoke, VA, United States, [3] Center for Radiological Research, Columbia University Irving Medical Center, New York, NY, United States

Fungi can tolerate extremely high doses of ionizing radiation compared with most other eukaryotes, a phenomenon encompassing both the recovery from acute exposure and the growth of melanized fungi in chronically contaminated environments such as nuclear disaster sites. This observation has led to the use of fungi in radiobiology studies, with the goal of finding novel resistance mechanisms. However, it is still not entirely clear what underlies this phenomenon, as genetic studies have not pinpointed unique responses to ionizing radiation in the most resistant fungi. Additionally, little work has been done examining how fungi (other than budding yeast) respond to irradiation by ionizing particles (e.g., protons, α-particles), although particle irradiation may cause distinct cellular damage, and is more relevant for human risks. To address this paucity of data, in this study we have characterized the phenotypic and transcriptomic response of the highly radioresistant yeast *Exophiala dermatitidis* to irradiation by three separate ionizing radiation sources: protons, deuterons, and α-particles. The experiment was performed with both melanized and non-melanized strains of *E. dermatitidis*, to determine the effect of this pigment on the response. No significant difference in survival was observed between these strains under any condition, suggesting that melanin does not impart protection to acute irradiation to these particles. The transcriptomic response during recovery to particle exposure was similar to that observed after γ-irradiation, with DNA repair and replication genes upregulated, and genes involved in translation and ribosomal biogenesis being heavily repressed, indicating an attenuation of cell growth. However, a comparison of global gene expression showed clear clustering of particle and γ-radiation groups. The response elicited by particle irradiation was, in total, more complex. Compared to the γ-associated response, particle irradiation resulted in greater changes in gene expression, a more diverse set of differentially expressed genes, and a significant induction of gene categories such as autophagy and protein catabolism. Additionally, analysis of individual particle responses resulted in identification of the first unique expression signatures and individual genes for each particle type that could be used as radionuclide discrimination markers.

Keywords: melanin, ionizing particle, transcriptomic, radiobiology, microbiology, fungi, yeast

OPEN ACCESS

Edited by:
Brian P. Hedlund, University of Nevada, Las Vegas, United States

Reviewed by:
Emma Camacho, Johns Hopkins University, United States
Ana Plemenitas, University of Ljubljana, Slovenia

***Correspondence:**
Zheng Wang
zheng.wang@nrl.navy.mil

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

Received: *24 September 2020*
Accepted: *14 December 2020*
Published: *12 January 2021*

Citation:
Schultzhaus Z, Chen A, Shuryak I and Wang Z (2021) The Transcriptomic and Phenotypic Response of the Melanized Yeast Exophiala dermatitidis to Ionizing Particle Exposure. Front. Microbiol. 11:609996. doi: 10.3389/fmicb.2020.609996

INTRODUCTION

The impact of ionizing radiation (IR) on biological systems has been a focus of research for decades, and it remains a relevant and fruitful area of research for several reasons. First, the effects of IR are similar in nearly every type of cell: reactive oxygen species (ROS) produced through the radiolysis of water result in protein oxidation and DNA damage, particularly double strand breaks (DSBs), which are fatal to cells if misrepaired (Ward, 1988; Riley, 1994; Battista et al., 1999; United Nations Scientific Committee on the Effects of Atomic Radiation, 2000). Distantly related organisms, therefore, essentially have to solve the same problem when recovering from IR exposure, imparting value on comparative studies. This damage is also similar to that encountered in cancer and age-related diseases, giving radiobiological studies broad application. Second, IR exposure is an ever-present threat, with risks in the clinic (e.g., for diagnostics and cancer treatment), in combat zones (e.g., nuclear warfare), sites of nuclear disasters (e.g., Chernobyl, Fukushima), and for extended missions in deep space. Finally, despite the massive radiobiological literature that has accumulated over the years, treatments for IR exposure are elusive, with virtually no prophylactic compounds available and little recourse for acute exposure (DiCarlo et al., 2017).

Research continues, therefore, to detail the shared and unique IR resistance strategies throughout life. Two broad veins of research have emerged from this work. The first focuses on the importance of DNA repair enzymes. In bacteria, this led to the discovery of the *rec* genes which facilitate homologous recombination (HR; McEntee et al., 1979; Roberts et al., 1979; Sargentini and Smith, 1986; Clark, 1991; Gruenig et al., 2011). A RecA homolog is known in eukaryotes [Rad51 (Shinohara et al., 1992)], though further studies in budding yeast and other models have described a much more complex DNA repair repertoire, including non-homologous end-joining (NHEJ) pathways that are much less common in prokaryotes (Critchlow and Jackson, 1998). In a parallel vein of research, intracellular Mn^{2+} complexes and other small metabolites have been shown to protect proteins from IR-induced oxidation and correlate with IR resistance in many organisms (Daly et al., 2004; Sharma et al., 2017; Doble and Miklos, 2018).

More recently, the observation that dozens of mold species could be isolated in and around the damaged nuclear reactor at Chernobyl (Zhdanova et al., 2000) has sparked interest in fungi outside of *Saccharomyces cerevisiae* for radiobiological studies. These organisms are indeed good models for this work – they have small, easily manipulated genomes and can be grown and irradiated easily (Loftus et al., 2005; Dadachova et al., 2007; Zakrzewska et al., 2011; Chen et al., 2014; Jung et al., 2016). Significant findings from studies in these organisms include the stimulation of growth of some fungi by low dose IR (Dadachova et al., 2007, 2008; Dadachova and Casadevall, 2008); the possibility that melanin in the cell walls of certain fungi may protect against IR damage (Dadachova et al., 2008; Pacelli et al., 2017a); the high IR resistance of organisms such as *Ustilago maydis* (Holliday, 1975; Lee and Yarranton, 1982; Holloman et al., 2007; Milisavljevic et al., 2018), *Cryptococcus neoformans* (Shuryak et al., 2014; Jung et al., 2016; Pacelli et al., 2017a; Schultzhaus et al., 2019), *Cryomyces antarcticus* (Pacelli et al., 2017b), and *Exophiala dermatitidis* (Romsdahl et al., 2020; Schultzhaus Z. et al., 2020; Schultzhaus Z. S. et al., 2020); and the discovery of novel IR resistance-associated proteins through transcriptomics and targeted mutagenesis (Jung et al., 2016; Schultzhaus et al., 2019; Schultzhaus Z. et al., 2020).

The vast majority of studies looking at the response to IR in fungi, however, have been done using either X-rays or γ-rays. These types of IR are not entirely indicative of most situations that are encountered in the clinic or in the field. Radiation treatments, nuclear accidents, and environments with high background IR (e.g., outer space), almost never contain a spectrum of pure γ- or X-rays. Rather, they involve ionizing particles of higher Relative Biological Effectiveness (RBE) and Linear Energy Transfer (LET) (Cucinotta and Durante, 2006). RBE refers to the ability of a type of IR to damage cells relative to a similar dose of γ-rays, while LET refers to energy transfer to a medium an ionizing particle produces per unit of track length. The distinction in the biological effects between high and low LET IR sources may be great, as charged particles are known to cause distinct types of DNA damage (e.g., more complex and clustered lesions), and could interact with other cellular structures (e.g., cell membranes, proteins, organelles, and melanin) to a greater extent than IR from photons (Ward, 1988; Hada and Georgakilas, 2008). For example, carbon ion (high LET) induced specific mutations in the *S. cerevisiae* and the filamentous fungus *Neurospora crassa* when compared with lower LET irradiation sources [X-rays or γ-rays (Ma et al., 2018; Matuo et al., 2019)], and irradiation with high or low LET sources produce different cell killing rates based on strain melanization status in *C. neoformans* (Pacelli et al., 2017a).

We are interested in the factors that mediate the response to IR exposure in fungi, and have previously characterized how melanin affects the response of *E. dermatitidis* to γ-IR, because it is an intrinsically melanized fungus that is highly radioresistant (Robertson et al., 2012; Romsdahl et al., 2020; Schultzhaus Z. et al., 2020). Acute and chronic γ-IR exposures induced dramatic transcriptomic responses in *E. dermatitidis* that allowed us to identify previously uncharacterized proteins important for its resistance, but nothing is currently known about how these results relate to the effects that ionizing particles have on this organism. Therefore, in this study we have characterized the transcriptomic and phenotypic response of melanized and non-melanized *E. dermatitidis* strains to ionizing particles of various LET (protons, deuterons, and α-particles) to understand the effects these particles have on this organism and how these differ from the effects produced by γ-IR.

MATERIALS AND METHODS

Growth of Fungal Cultures

Exophiala dermatitidis cultures (both melanized and non-melanized) were grown in yeast peptone dextrose (YPD) liquid

or solid (adding 1.5% agar) medium at 30°C throughout the experiment. To prepare an individual biological replicate for irradiation, a single colony of the appropriate strain, stored on YPD medium, was inoculated into 2 mL YPD and grown in a shaking incubator set at 30°C and shaking at 200 RPM, for 2 days, after which the resulting culture was diluted 1:1000 in 40 mL of YPD and grown for an additional 2 days. At this point, the cell concentration of the resulting culture was measured using a Cellometer (Nexcelom Bioscience, MA, United States) set to measure clustered yeast cells. Cells were then centrifuged at 4000 × *g* for 5 m and re-suspended in YPD to a final concentration of 1×10^9 cells/mL for radiation exposure. Three biological replicates for each strain were grown for irradiation in this experiment.

Irradiation of *E. dermatitidis*

Ionizing particles, unlike other forms of IR such as γ-radiation, have limited penetration into materials, depending on the particle size and the composition of the materials (Fano, 1963). α-particles, for example, could potentially be blocked by only a few layers of *E. dermatitidis* cells. Therefore, to ensure complete and homogeneous cell exposure, cells were irradiated in chambers composed of a 10 μm thick Mylar sheet stretched across the bottom of a metal ring. A 50 μL drop of concentrated cell suspension (1×10^9 cells/mL) was placed on the center of the sheet, and was carefully flattened under an 18 mm × 18 mm glass coverslip. Particle beams entered the chamber from below, passing first through the Mylar sheet and then through the cell layer.

Due to the length of sample processing and technical limitations of switching between particle sources, proton and α-particle irradiations were performed on 1 day, and their control samples were pooled, while growth and processing of cells for the deuteron irradiation was shifted by 1 day, and had its own set of controls. Total doses used were 125 Gy, 250 Gy, and 500 Gy for α-particle irradiations, 420 Gy, 840 Gy, and 1680 Gy for proton irradiations, and 250 Gy, 500 Gy, and 1000 Gy for deuteron irradiations. This corresponded to exposures to 500 Gy, 1000 Gy, and 2000 Gy irradiation by γ-rays, doses which we used previously to compare the IR resistance of melanized and non-melanized *E. dermatitidis* strains (Schultzhaus Z. et al., 2020). After exposure, cells were collected by carefully lifting each coverslip and removing the cells adhered to it, and to the Mylar sheet, through addition of 3 mL of sterile ddH_2O and vigorous pipetting. To obtain enough cells for RNA extraction and plating, cells from three different dishes were pooled. The resulting cell suspensions, each totaling approximately 9 mL, were added to 50 mL tubes.

To determine the cell survival, 100 μL was collected from each cell suspension and diluted in an additional 900 μL of ddH_2O, followed by counting on a Cellometer. Each suspension was then diluted further to appx. 5×10^3 cells/mL, and 100 μL of this suspension was plated on each of two solid YPD plates, to obtain a total of six plates per condition and strain. CFUs were then counted after a 5-day incubation at 30°C, and from these measurements, survival after each exposure was determined.

Modeling of Survival Data

The standard linear quadratic (LQ) model of radiation-induced cell death is commonly used in radiation biology and oncology. This formalism predicts the clonogenic surviving fraction (*S*) of cells exposed to an acute dose *D* by the following equation, where *PE* is the cell plating efficiency without radiation, α is the linear dose response component and α/β is the "alpha/beta ratio" which quantifies the dose response "curvature":

$$S = PE \times \exp[-\alpha \times D - \alpha/(\frac{\alpha}{\beta}) \times D^2] \quad (1)$$

In this model, survival at relatively low doses ($D < \alpha/\beta$) is dominated by parameter α, whereas survival at high doses ($D > \alpha/\beta$) is dominated by parameter β. *E. dermatitidis* data over the studied dose range of 0–1.6 kGy did not exhibit any visible "downward curvature" on a logarithmic scale, suggesting that the contribution of the α/β parameter is negligible in this dose range. Indeed, preliminary fits of the model to these data showed that the quadratic term in Eq. (1) is not statistically distinguishable from zero. This finding allowed us to simplify the model by eliminating the parameter α/β and retaining only the $\alpha \times D$ term.

Importantly, not all cells in an irradiated cell population may have the same radiosensitivity, for example due to differences between cells in cell cycle stage, age, nutrient status, epigenetics, oxygenation, *etc*. Models that take such heterogeneous radiosensitivity into account have been used to describe the survival of tumor cells during radiotherapy (e.g., Carlone et al., 2004; Shuryak et al., 2015). One simple example of such a model, which allows analytic calculations, employs the exponential distribution to describe the probability distribution of the linear dose response parameter among the cell population. This can be done as follows, where $P(u)$ is the probability that the linear dose response parameter is equal to u, where the mean value of this parameter is α:

$$P(u) = (\frac{1}{\alpha}) \times \exp[-\frac{u}{\alpha}] \quad (2)$$

The mean of Eq. (2) is α, and the variance is α^2.

Equations 1 and 2 can be combined to generate the following equation for the cell surviving fraction (*SF*) as function of radiation dose *D*:

$$SF = \int_0^\infty PE \times \exp[-u \times D] \times P(u) \times du = PE/(1 + \alpha \times D) \quad (3)$$

This survival function (Eq. 3) was fitted to the data for *E. dermatitidis* WT and PKS strains using maximum likelihood. The log likelihood (*LL*) function was generated from the binomial distribution, using the observed number of colonies and the estimated number of plated cells for each experimental result as the inputs ("number of successes" and "number of trials," respectively, in commonly used binomial terminology). Best-fit values for parameters *PE* and α were found by maximizing the *LL* using the sequential quadratic programming (SQP) algorithm in Maple 2019 software. Data for the WT and PKS strains were analyzed separately, but for each strain the data for different particle types were combined because of their similarity.

For each strain, fitting was performed 1000 times with randomly chosen initial parameter values, and the best result was recorded. This procedure was intended to maximize the probability of finding the global optimum rather than a local one. Model parameter uncertainties (95% confidence intervals, CIs) were estimated by generating 3000 synthetic data sets where the observed number of colonies and estimated number of plated cells for each experimental result were replaced by Poisson-distributed random numbers with the mean set to the corresponding value in the real data set. The model was fitted to each synthetic data set, and parameter values recorded. The 2.5th and 97.5th percentiles of the distribution of each parameter across synthetic data sets were used as estimates of 95% CI for this parameter.

RNA Extraction and Sequencing

Additionally, control cultures, and cultures irradiated with the second highest doses (250 Gy α-particles, 840 Gy protons, and 500 Gy deuterons) were processed further for RNA extraction: 9 mL of 2 × concentrated YPD medium was added to each of the cell suspensions collected from irradiated dishes, and these tubes were incubated for 1 h at 30°C and shaking at 200 RPM to allow for recovery. After this final incubation, cell suspensions were collected by centrifugation (4000 × *g*, 5 m) and re-suspended in 1 mL of RNAlater (Thermo Fisher, MA, United States). RNAlater solutions were split into two tubes (500 μL each) and stored at RT for 2 days before pelleting through centrifugation (15000 × *g*, 1 m), removal of the supernatant, and storing them at −80°C until RNA extraction.

RNA was extracted from individual cell pellets with the RiboPure RNA Purification Kit for yeast (Thermo Fisher), and RNA quality was verified using a 2100 Bioanalyzer (Agilent, CA, United States) to confirm that RNA integrity numbers (RIN) were greater than 7.5. Two samples from each condition were then sent on dry ice and subjected to sequencing on an Illumina HiSeq4000 Sequencer by the Yale Center for Genome Analysis (YCGA, Yale University School of Medicine, West Haven, CT, United States). Verification with qRT-PCR and validation that two replicates was sufficient to identify many transcriptomic changes in response to IR were completed in a previous publication (Schultzhaus Z. et al., 2020). Data from this RNA-seq experiment is publicly available through the Gene Expression Omnibus using accession number GSE152116.

RNA-Seq Data Analysis

To analyze gene expression changes in response to particle irradiation, raw FASTA files produced by the sequencing runs were downloaded from the YCGA. These files were subjected to alignment and quantification using Salmon 2.0 software (Patro et al., 2017), with the transcriptome and genome of *E. dermatitidis*, both assembled and characterized previously (Robertson et al., 2012; Chen et al., 2014; Schultzhaus Z. et al., 2020), used as a reference. Transcript abundance and quantification information provided through this analysis was analyzed using the DEseq2 package (Love et al., 2014) in RStudio (RStudio Team, 2015). Deuteron had an independent set of controls due to a different time of irradiation, as mentioned above, but was otherwise treated the same and it exhibited similar levels of correlation between biological replicates.

Further analysis to identify patterns within the expression of the RNAseq data were performed using several software packages. Venn diagrams were produced using Venny 2.1[1] (Oliveros, 2007). Statistical analysis of enrichment of genes was performed using FungiFun 2[2] (Priebe et al., 2011, 2014). Search settings in this case were adjusted to identify Gene Ontology terms under the category of Biological Process, with an FDR of <0.01.

EuKaryotic Orthologous Group (KOG) analysis of highly regulated genes from both particle and γ-regulated and was performed using information provided by previous publications and available at MycoCosm (Chen et al., 2014; Grigoriev et al., 2014). Genes were placed into the broadest KOG categories for presentation of the least convoluted depiction of the datasets.

Network analysis was performed to identify connected and similarly regulated genes using STRING software[3] (Snel et al., 2000; Mering et al., 2003), with settings programmed to show only connections with the highest significance (confidence > 0.9). Data on the nodes identified through this search were downloaded, and the results presented in the manuscript were compiled to include the genes that had one or more connection to another upregulated gene with a confidence value of 0.999.

Motifs that were enriched among highly regulated genes were discovered using the MEME-suite[4] (Bailey et al., 2009, Bailey et al., 2015) using the Multiple Em for Motif Elicitation (MEME) software for identification. Promoters for all genes regulated >5-fold and <5-fold were used as input. Shuffled input sequences, random promoter sequences, and random ORF sequences were used as controls to verify the identification of the enriched motif. The enrichment of the identified motif was then verified using Analysis of Motif Enrichment (AME), and the complete set of promoters was searched for this motif using the Find Individual Motif Occurrences (FIMO) program.

RESULTS

Susceptibility of *E. dermatitidis* to Proton, Deuteron, and α-Particle Irradiation

We initiated this study by characterizing the resistance of *E. dermatitidis* melanized (WT) and non-melanized (Δ*pks*) strains to three ionizing particles with different RBE, LET, and mass [in ascending order: protons (p), deuterons (D), and α-particles (α) (Hall and Giaccia, 2006)]. To enable comparisons with γ-IR, we considered several factors (see **Figure 1**) (see section "Materials and Methods"), including: the lower penetration of these particles into materials; the RBE of each particle type in comparison with γ-IR (Fano, 1963; Nikjoo et al., 1999); and the shape of previously constructed survival

[1]https://bioinfogp.cnb.csic.es/tools/venny/

[2]https://elbe.hki-jena.de/fungifun/

[3]https://string-db.org/

[4]https://meme-suite.org

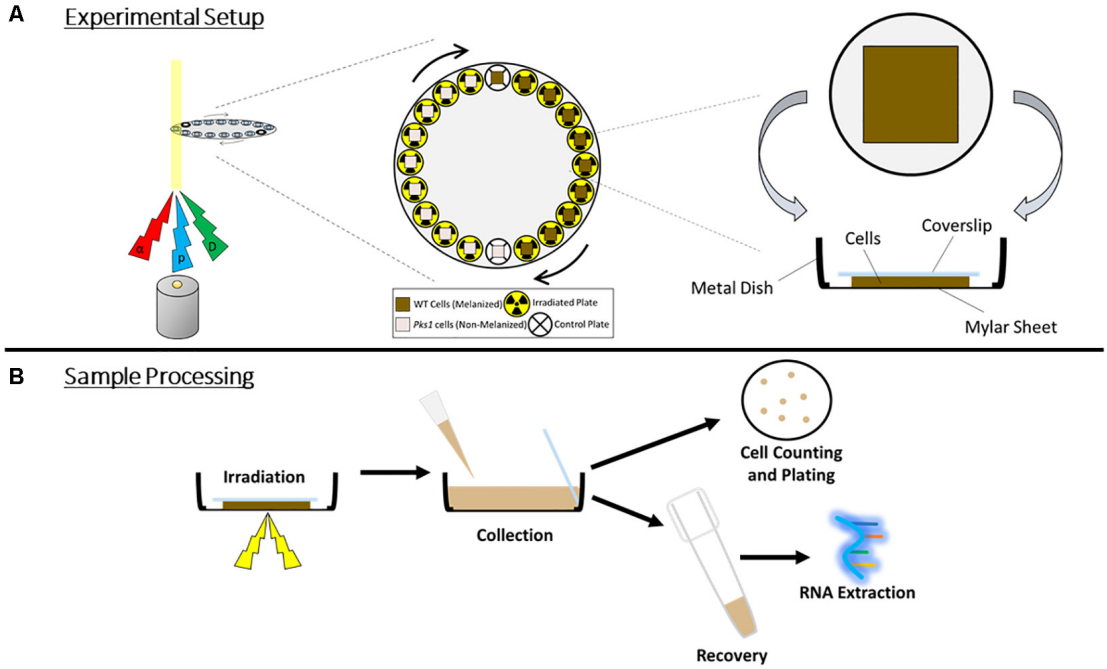


FIGURE 1 | (A) Irradiation setup. Particles (α, alpha; p, proton; D, deuteron) were produced and passed through cell sample chambers placed along the circumference of a disk, which rotated automatically to provide each sample with an even exposure. For control samples, particle source was momentarily stopped as the cells passed through the beam. **(B)** Overview of sample processing during and post-irradiation, demonstrating the orientation of the cells and the coverslip within the chamber, and the harvesting of cells for survival measurements and RNA extraction.

curves exhibited by *E. dermatitidis* after irradiation. At doses much greater than 2000 Gray (Gy, J/kg), for example, survival decreases rapidly, and differences between strains become harder to interpret due to the low percentage of recovering colonies produced (Schultzhaus Z. et al., 2020). Effects (such as protection imparted by melanin) are therefore best detected at moderate exposures. Taking this into account, we chose to irradiate each strain with doses equivalent to 500, 1000, and 2000 Gy of γ-IR, similar to the doses we used to compare melanized and non-melanized strains in a previous publication (Schultzhaus Z. et al., 2020).

After irradiation, a known quantity of cells from each sample was inoculated on separate plates. Survival was then modeled with the resulting measurements. The best-fit parameter values for the WT strain were: α = 6.38 (95% CI: 0.26, 16.51) kGy^{-1}, *PE* (Plating Efficiency) = 0.60 (0.21, 0.99). Corresponding values for the Δ*pks* strain were: α = 2.10 (0.01, 10.04) kGy^{-1}, *PE* = 0.32 (0.16, 0.80) (see the section "Materials and Methods" for explanation of modeling and **Supplementary Tables 1, 2** for plating and modeling data). The fitted dose response curves and data points are shown in **Figure 2**. Notably, these differences between strains in terms of parameter values did not reach statistical significance (95% CIs overlapped). However, these results suggest some evidence of heterogeneous radiosensitivity in the irradiated fungal population. The Δ*pks* strain appears to have a smaller α parameter (see the section "Materials and Methods" for description of this parameter) than the WT strain, suggesting that Δ*pks* has lower radiosensitivity. PE, though, was also smaller for Δ*pks* than WT, which may suggest that only some selected "healthy" sub-population of Δ*pks* cells was able to grow even without radiation, and possibly this selection was correlated with improved radioresistance as well. Indeed, Δ*pks* cells have been shown to clump together, possibly due to a decreased ability to completely cleave apart after cell division, which can affect plating efficiency (Feng et al., 2001; Schultzhaus Z. et al., 2020). In all, though, there was no clear protection, or difference at all, in WT cells compared to those without melanin.

The Transcriptomic Response of *E. dermatitidis* to Particle Irradiation

The effect of exposure on cell survival was roughly equivalent between these particles and γ-IR for similar doses (Schultzhaus Z. et al., 2020), but with knowledge of their varying LET characteristics, we hypothesized that differences between these conditions could be accessed by looking at the transcriptomic response for each source. Therefore, next, we performed RNA-seq on cultures recovering from the second highest dose (equivalent to 1000 Gy of γ-IR) for each particle, corresponding to 30–50% cell survival depending on the type of IR. Before collecting RNA we also incubated cells in fresh medium for 1 h to allow for them to mount a transcriptomic response (see "Materials and Methods").

Each particle irradiation induced a strong and consistent gene expression pattern in both strains (**Supplementary Tables 3, 4**). For example, R^2 values from comparative analysis of transcripts

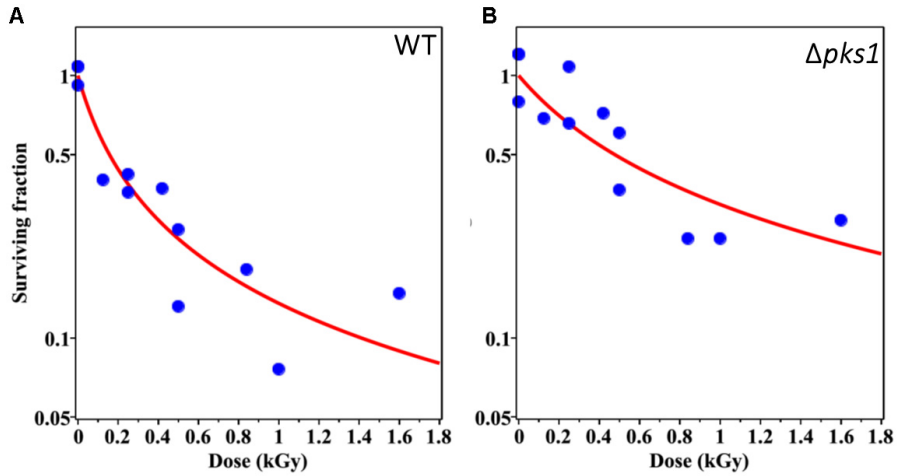


FIGURE 2 | Survival of WT **(A)** and Δ*pks1* **(B)** cells after irradiation with ionizing particles of different LET. Blue circles represent experimentally determined data points from the survival assay and red lines indicate the most optimal (best-fit) survival curve for each strain, using maximum likelihood, when incorporating CFU data at each dose into the survival function (Eq. 3). Dose in kGy is provided on the *x*-axis.

per million counts ranged from 0.84 to 0.95 between biological replicate samples as well as WT and Δ*pks1* cultures under the same condition (IR exposure or control), while the irradiated transcriptome was vastly different from that of controls (**Figure 3A**, **Supplementary Table 3**, and **Supplementary Figure 1**). Initial analysis of differential gene expression with an FDR < 0.05 demonstrated that a range of 4088 (2110 up, 1978 down in the Δ*pks1* deuteron-irradiated culture) to 5305 (2739 up, 2566 down transcripts in the Δ*pks1* proton-irradiated culture) out of the 9268 transcripts predicted to be encoded by the *E. dermatitidis* genome (Chen et al., 2014) were differentially expressed. Overall, 1095 (upregulated) and 969 (downregulated) genes were seen in every sample at this cutoff level, representing 22.2% of the predicted protein-coding genes in this organism (**Figure 3B** and **Supplementary Tables 3, 4**).

We next looked within this shared set of genes for Gene Ontology (GO) categories that were significantly enriched (see the section "Materials and Methods"). This resulted in the following observation (**Figures 3C,D**): genes involved in DNA repair (e.g., DNA replication, DNA recombination, and nucleotide-excision repair) and genes involved in protein catabolism (e.g., ubiquitination) were overrepresented among the upregulated genes (**Figure 3C**). Approximately one third of the genes in the upregulated group (369/1095, 33.7%), moreover, were predicted to encode proteins with no known function.

Among the downregulated genes (**Figure 3D**), biological processes that were significantly enriched were involved in cell growth, including categories such as intracellular trafficking (microtubule-based movement), ATP synthase-coupled proton transport, and translation (ribosomal biogenesis) were particularly enriched. Once again, one third of the genes (340/969, 35.1%) encoded hypothetical proteins (**Supplementary Tables 3, 4**). This is approximately half as numerous as the proportion of unannotated genes within the entire genome (5842/9578, 60.1%), which suggests that this set is enriched with genes involved in well-characterized processes. However, hundreds of genes still remain to be characterized in this organism before a full picture of this response is known. Taken together, these results suggested that *E. dermatitidis* cells respond to charged particle irradiation by activating the DNA repair machinery and removing damaged proteins, as well as inhibiting cell growth.

Comparison of the Transcriptomic Responses of *E. dermatitidis* to Particle and γ-Irradiation

This response was similar to the γ-radiation response (Schultzhaus Z. et al., 2020), something which could be seen at a larger level by plotting the $\log_2$fold-change values of genes significantly regulated by all IR sources on a heatmap (**Figure 4**). However, the heatmap produced by this analysis showed that samples also clearly clustered by IR type (particle and γ-ray), demonstrating that there were notable differences between the two responses that needed further analysis to uncover.

Therefore, we focused on getting more information about the genes that were differentially expressed in each group, particularly those that did not have informative annotations, and compiled lists of genes fitting this description that were also differentially regulated more than five-fold (FDR < 0.05) in either γ-radiation or particle – datasets for comparison (Schultzhaus Z. et al., 2020). Notably, we observed less overlap among these two sets than we expected, with only approximately 12% of the genes shared overall (**Figures 5A,B**). We also were surprised to see that, even after compiling the results from six separate experiments (WT and Δ*pks* cells for three particle exposures), there were many more genes that passed the >5-fold change cutoff in the particle dataset compared with the γ-irradiated dataset ($N = 467$ vs. 166, respectively), indicating that particle irradiation invoked a more dramatic response in the transcriptome.

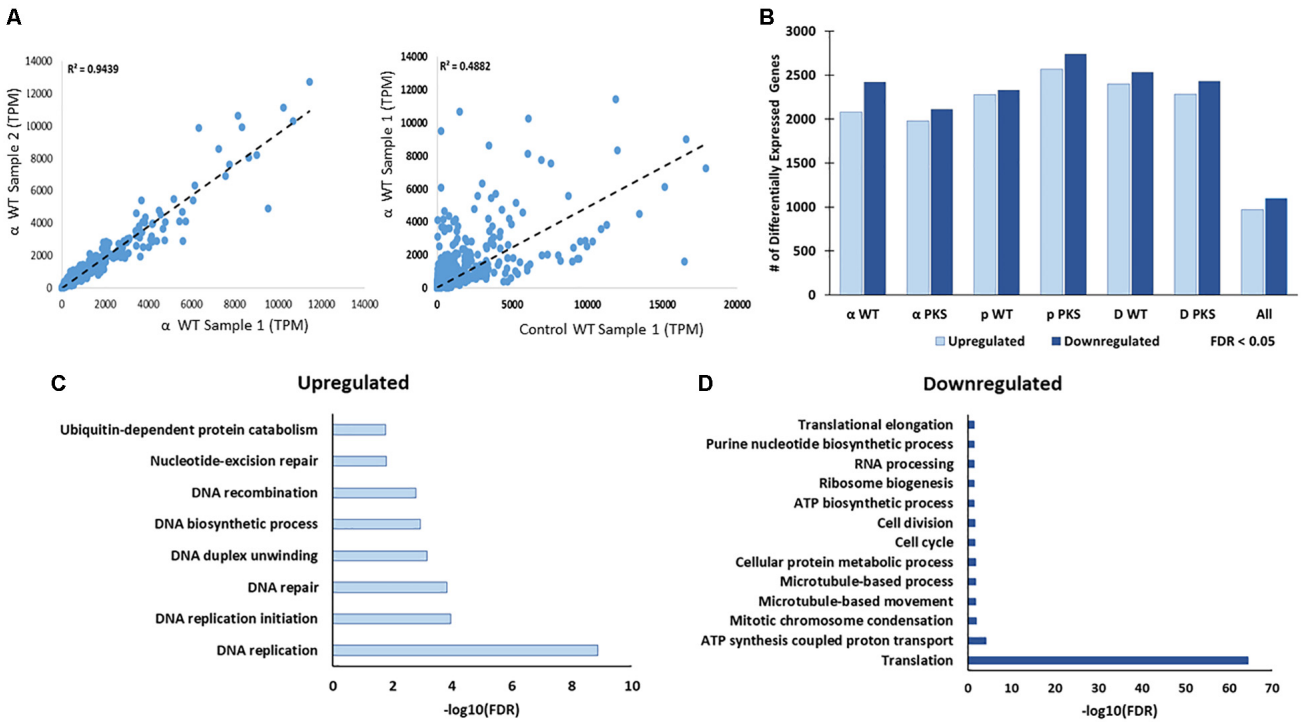


FIGURE 3 | (A) Correlation between Transcripts Per Million (TPM) counts between biological replicate samples within the same condition, between different strains in the same condition, or within the same strain across different conditions. Here, and below, the strains included are the melanized WT and the non-melanized PKS strains, and the condition being exposure to the second-highest dose of particle IR, as detailed in the Section "Materials and Methods" (250 Gy α-particles, 840 Gy protons, and 500 Gy deuterons). **(B)** The number of genes differentially expressed, with a significance cutoff of FDR < 0.05, for each condition, as well as shared in all conditions. **(C)** Gene Ontology-Biological Process categories enriched with the shared, upregulated and **(D)** downregulated gene sets. FDR values provided by FungiFun 2.0. Higher absolute values of –log10(FDR) demonstrate greater significance.

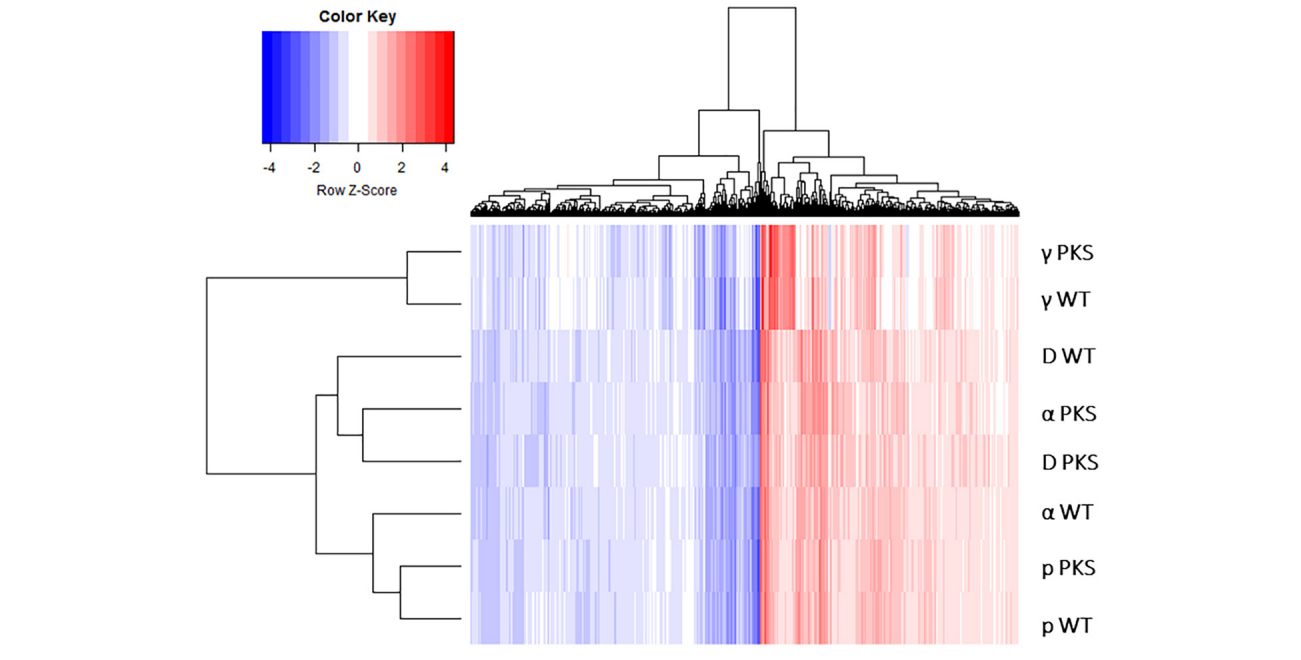


FIGURE 4 | Heatmap depicting the $\log_2$-fold change values of genes within the six particle and the two γ-IR-responding datasets. Genes evaluated were those that were found to be differentially expressed (with an FDR value < 0.05) in all of the six particle datasets.

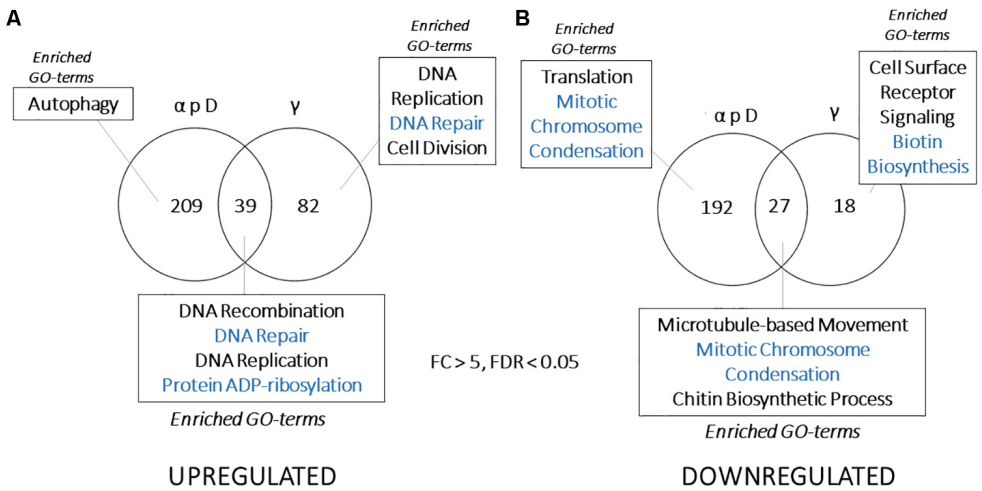


FIGURE 5 | (A) Venn diagram showing overlap between genes that are highly upregulated or **(B)** downregulated (FDR < 0.05, >5 or <–5-fold change in expression) in particle or γ-induced datasets. Categories presented for each section include Gene Ontology-Biological Process categories enriched (FDR < 0.01, from FungiFun 2.0) for each group of genes.

The lack of overlap between these gene sets caused us to reanalyze these data for patterns of enrichment that were specific to each group. This revealed some interesting features. First, DNA repair genes were still enriched among the shared, upregulated set, but DNA replication and repair genes were also enriched within the γ-specific, upregulated group. These proteins included Rad54B, the NHEJ protein Ku70, DNA damage-binding protein CMR1, as well as DNA polymerase α and ε, replication fork protection complex subunit Swi1/Tof1, and the DNA replication regulator SLD2 (**Figure 5A**). Proteins only upregulated by particle irradiation, on the other hand, included autophagy-related proteins (HMPREF1120_02756/ATG8, HMPREF1120_07182/ATG3, and the cysteine protease HMPREF1120_06308/ATG4) and specialized genes potentially involved in sensing and signaling, such as bacteriorhodopsin (HMPREF1120_00264) and a Sir2-like histone deacetylase (HMPREF1120_05820).

Additionally, the group of shared, downregulated genes was small (N = 27), and was enriched for proteins involved in cell growth, cell cycle, and intracellular trafficking, but not enriched for those associated with translation and ribosomal structure (**Figure 5B**). An even smaller set of genes (N = 18) were uniquely downregulated by γ-radiation. On the other hand, translation was highly enriched among the downregulated genes specific to the particle-irradiated gene set. This is notable for two reasons - the antagonistic relationship between protein synthesis and autophagy (Beau et al., 2008; Kraft et al., 2008; Wyttenbach et al., 2008; Hands et al., 2009), and the observation that proteins involved with translation were downregulated by γ-radiation, just not to the extent to pass the cutoff here. Particle irradiation, therefore, induces a similar, but far stronger repression of translation-associated genes than γ-radiation (Schultzhaus Z. et al., 2020).

After this analysis many genes still remained without a predicted function, so we further examined each set with information from the KOG database. Grouping genes into 22 of the KOG categories represented in a substantially decreased the number for which no information was known (from 188/467 to 46/467 for particle irradiation, and from 52/166 to 20/166 for γ-irradiation), and also allowed us to look at the functional composition of the entire *E. dermatitidis* genome. **Figure 6** shows these results. It is immediately apparent that particle-regulated genes embody a greater functional diversity compared to γ-regulated genes. Upregulated genes from the particle irradiation sets include a high proportion annotated as involved in defense mechanisms (e.g., multidrug transporters and general stress-response proteins) while the downregulated gene set from this experiment had a large proportion involved in amino acid transport and chromatin structure compared with the whole genome. The upregulated genes in the γ-IR experiment, however, were dominated by those involved in Replication, Recombination, and Repair, while among downregulated genes there appeared to be an enrichment of those involved in secondary metabolism (**Figure 6**).

In all, although the transcriptomic response to IR broadly involves upregulation of DNA repair and attenuation of growth and protein production, there are some differences observed in the response to particle irradiation, including an upregulation of genes regulating cell death and a stronger downregulation of translation-associated genes.

Genes Most Sensitive to Particle Irradiation Exposure

One goal of this study is to identify genetic loci that robustly respond to IR, which could be used to develop strains that can sense IR and produce various outputs tied to gene expression. With this in mind, we searched for the most promising loci using the shared particle gene through MEME analysis, which searches for conserved promoter sequences in gene sets; STRING analysis,

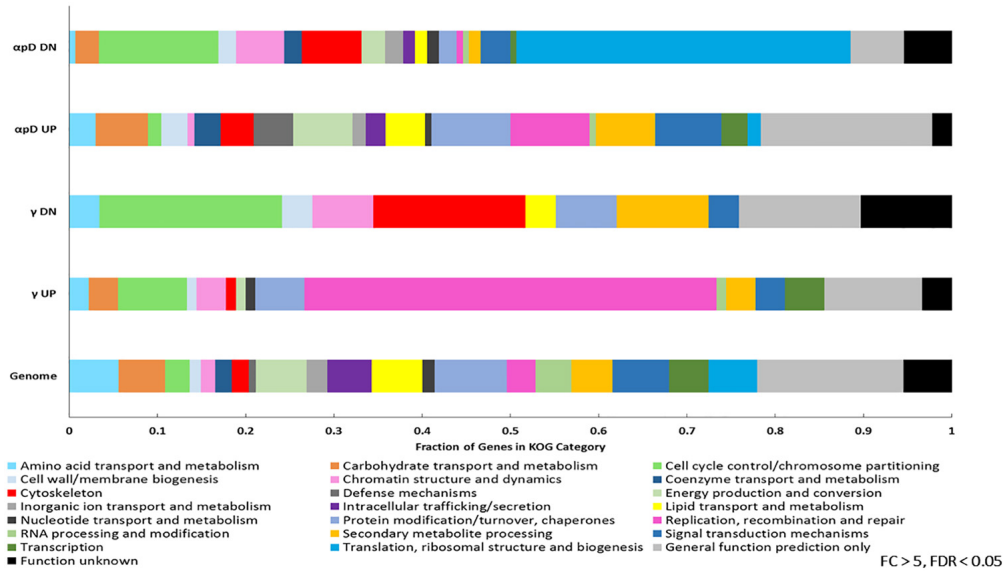


FIGURE 6 | KOG categories identified among the genes differentially expressed >5 or <–5-fold in the particle (α-, proton, or deuteron) or γ-induced datasets, as well as the whole *E. dermatitidis* genome.

which identifies connections between genes within datasets based on published genetic studies; and identification of the genes that showed the strongest change in expression during IR recovery. The output of MEME and STRING analyses are presented in **Supplementary Tables 5–8** and **Supplementary Figure 2**. Briefly, MEME analysis identified a possible 9-nucleotide sequence marking IR-regulated genes, although it was clear that other unidentified elements also exist (**Supplementary Tables 5, 6**). STRING analysis further confirmed the centrality of DNA replication among the upregulated genes, but also identified certain genes that were central to this pathway, such as the Minichromosome Maintenance Protein complex (**Supplementary Figure 2** and **Supplementary Tables 7, 8**).

Next, to identify the most prominent among the differentially expressed genes, we averaged expression data across all strains and particle samples. We identified 19 genes that were differentially expressed with a $\log_2$FC > 5, representing a >32-fold change in expression during recovery from IR exposure (**Table 1**). The most upregulated gene ($\log_2$FC = 6.44 ± 0.77) among all samples was HMPREF1120_01375. It is annotated as a triacylglycerol (TAG) lipase and contains acetyl esterase and lipase domains, potentially associating it with fatty acid catabolism in autophagy (Cingolani and Czaja, 2016; Jaishy and Abel, 2016). Other upregulated genes in this list include HMPREF1120_08154, a homolog of *Saccharomyces cerevisiae* YTA12, a component of the mitochondrial inner membrane mAAA-protease that could be involved in clearing damaged proteins; HMPREF1120_02698, a homolog of *S. cerevisiae* UBP15, which is involved in S-phase entry (Ostapenko et al., 2015); and HMPREF1120_02127, a homolog of *S. cerevisiae* gene YGL039W, a protein involved in reducing aldehydes, which are present in cells undergoing oxidative stress (Singh et al., 2013). Interestingly, HMPREF1120_08448, HMPREF1120_00007, and HMPREF1120_06618 did not contain any predicted functional domains, and we were unable to find any orthologs of these genes in genomes from species outside of the *Exophiala* genus.

Within the group of shared downregulated genes, three genes had no conserved domains or informative orthologs in other species (HMPREF1120_04574, HMPREF1120_03067, HMPREF1120_08344). Of the other six, however, four had functions related to known IR responses, being involved in chromosome dynamics in some way, such as the orthologs of SMC1 and SPC25 (chromosome segregation), the aurora kinase IPL1, and the condensin YCG1. Additionally, the chitin synthase CHS1 and an endoglucanase were the remaining two genes that were on this list of genes repressed during IR exposure recovery. These could be associated with the cessation of growth that we have observed taking place in *E. dermatitidis* cells while damage is repaired or cleared (Schultzhaus Z. et al., 2020). Importantly, 16/19 of these genes were regulated similarly in our γ-regulated dataset, so further studies of IR sensitive promoters may be able to focus on the expression of these few genes in different environments.

Particle-Specific Transcriptomic Responses

Another goal of this study was to identify unique responses to different ionizing particles, so next we looked for patterns specific to certain irradiation datasets. Such loci were not immediately apparent, as the datasets were similar. This is demonstrated by comparing FC values for each significantly regulated gene (FDR < 0.05) shared among the particle groups, which produced R^2 values >0.92 for all pairwise comparisons (**Figure 7**). Moreover, no genes were observed to be upregulated under one condition and downregulated in another. When these gene

TABLE 1 | Genes that were differentially expressed to a level >30-fold or <−30-fold in all particle conditions, including information about their function, conservation, and the average induction or repression levels.

All sources	Gene	Uniprot annotation	Conserved domains	Avg. log_2FC	SD
Up	HMPREF1120_01375	Triacylglycerol lipase	Acetyl esterase/lipase	6.44	0.77
	HMPREF1120_08154	AAA family ATPase	SpoVK/Ycf46/Vps4 family (cell wall/membrane biogenesis, cell cycle control)	6.20	0.73
	HMPREF1120_02698	Ubiquitin thioloesterase	Ubiquitin carboxyl-terminal hydrolase; zinc finger; PIP-kinase	5.84	0.68
	HMPREF1120_08138	Uncharacterized protein	Ankyrin repeat domains	5.84	0.85
	HMPREF1120_08448	Uncharacterized protein	–	5.60	1.84
	HMPREF1120_08742	Dimethylaniline monooxygenase (*N*-oxide forming)	Flavin-containing monooxygenase	5.39	0.64
	HMPREF1120_00007	Uncharacterized protein	–	5.28	0.93
	HMPREF1120_02578	Protein YOP1	TB2/DP1 (deleted in polyposis), HVA22 (abscisic acid-induced) protein family	5.22	0.68
	HMPREF1120_02127	Cinnamyl-alcohol dehydrogenase	Short-chain dehydrogenases/reductases	5.04	0.91
	HMPREF1120_06618	Uncharacterized protein	–	5.02	0.68
Down	HMPREF1120_07157	Structural maintenance of chromosomes protein	SMC superfamily, chromosome segregation ATPase	−5.05	0.88
	HMPREF1120_06748	Aurora kinase, other	STKC Aurora kinase superfamily (chromosome segregation)	−5.18	0.78
	HMPREF1120_01579	Condensin complex subunit 3	Chromosome condensation complex	−5.32	1.26
	HMPREF1120_07981	Chitin synthase 1	Chitin synthase: classes I, II, and III	−5.34	0.86
	HMPREF1120_04574	Uncharacterized protein	–	−5.64	1.02
	HMPREF1120_03688	Uncharacterized protein	Chromosome segregation subunit of the Nuf2-Ndc80 complex	−5.72	0.78
	HMPREF1120_03067	Uncharacterized protein	–	−5.77	0.65
	HMPREF1120_08344	Uncharacterized protein	–	−5.78	0.73
	HMPREF1120_04843	Endoglucanase	Cellulase, glycosyl hydrolase, glucanase	−5.85	1.13

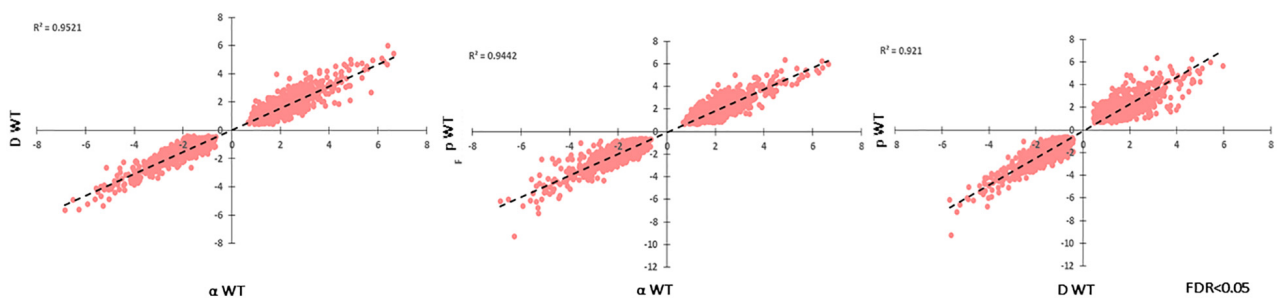


FIGURE 7 | Correlation of log_2FC-values for each significantly regulated gene present in all WT, particle-irradiated gene sets, demonstrating high concordance in the response to the three sources.

groups were minimized to facilitate closer analysis by looking only at genes differentially expressed in each condition with a FC > 5, however, several genes were identified in only one set, and some of these unique gene sets were significantly enriched for certain biological processes (**Figures 8A,B**). For α-particle irradiated samples, genes involved in DNA repair, including DNA Ligase 1 and the helicase RecQ, were upregulated, while genes predicted to be involved in translation were specifically enriched within the α-downregulated genes. Genes upregulated only after proton irradiation were enriched in transmembrane transport, including 13 Major Facilitator Superfamily domain-containing proteins, while genes involved in IMP biosynthesis (purine metabolism) and tRNA aminoacylation were overrepresented among downregulated genes. It is worth noting that protons and α-particles represent the two extremes of particle size, which may account for the higher number of uniquely regulated genes in these two datasets. Finally, within the deuteron set, genes involved in transcriptional regulation were enriched among the upregulated transcripts, including RNA-dependent RNA polymerase, which is thought to be involved in RNA silencing.

To search for more subtle patterns within this data, we again performed STRING analysis to identify pathways that were specific to a certain particle using genes with highly confident links to at least two other genes (probability > 0.9) among all significantly upregulated genes (FDR < 0.05) for each condition (**Supplementary Tables 9, 10**). This provided a detailed list of DNA repair and protein catabolism components induced during the response to particle exposure, including the polymerases

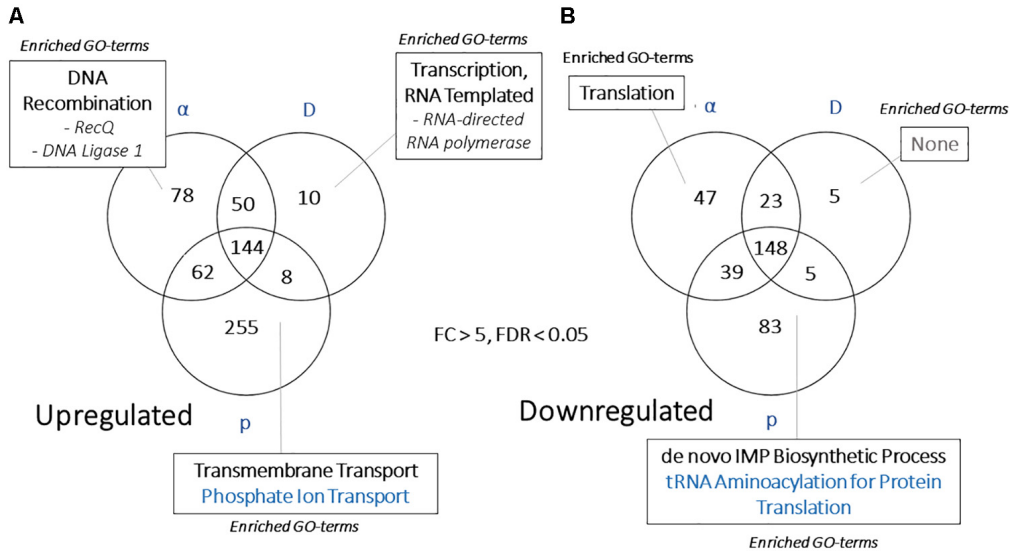


FIGURE 8 | (A) Venn diagram demonstrating the overlap in highly regulated [>5 or **(B)** <–5-fold] genes in each WT, particle-responsive dataset. Categories presented for each section include Gene Ontology-Biological Process categories enriched (FDR < 0.01, from FungiFun 2.0) for each group of genes.

involved in DNA replication (α, δ, and ε), the proteins involved in regulating the cell cycle (Cdc6, 45, and 48), and those involved in DNA repair (for HR – Rad50, 51, 52, MRE11, and Rhp54; for UV excision repair – Rad32; for mismatch repair – MLH1/MSH3/PMS2; and translesion synthesis – Rev1). It also revealed proteins involved in transcriptional regulation, including two histone acetyltransferase-associated categories, the histone chaperone ASF1, and a Jumonji domain-containing protein, which are involved in transcriptional repression through modulating histone methylation (Smith et al., 2006). Finally, it identified other distinct pathways not directly related to DNA or protein damage, including the Myosin ATPase, involved in intracellular trafficking, and nitrogen (glutamine synthase and glutamate dehydrogenase) and carbon (phosphoenolpyruvate carboxykinase) metabolism gene nodes.

Supplementary Tables 11, 12 present the genes represented as highly confident nodes present in only one or two conditions. This analysis provided the most detailed picture of the different ways each source affects *E. dermatitidis*. For example, 18 different categories had two or more genes that were related with high confidence between the α- and deuteron-irradiated samples, which included two unique categories of DNA-associated proteins: the RecQ repair protein and the cohesin complex. In the proton and deuteron datasets, moreover, DNA topoisomerase (DNA replication), the FACT complex (chromatin structure), the RNA helicase (splicing), and Tyrosinase (melanin production) nodes were present only in deuteron and proton datasets. Even more nodes were discovered that were unique to individual irradiation sources. α-particle samples had subunits of the actin-related protein 2/3 complex [which is involved in endocytosis and morphogenesis (Commer and Shaw, 2020)] and synaptobrevin, which is involved in intercellular trafficking. Similarly, the deuteron set contained proteins linked to fimbrin, which is involved in actin cytoskeleton regulation, endocytosis, and morphogenesis (Upadhyay and Shaw, 2008), and the MYST family of histone acetyltransferases. The proton data set had the most unique categories in this analysis (N = 30), and included interesting links such as cystathionine synthase and glutaredoxin genes, which may be involved in the response to increased ROS levels, and several genes involved in mRNA splicing, including the splicing factors *slt11* and *ini1*. This may warrant further exploration of alternative splicing and isoform analysis of *E. dermatitidis* cells recovering from specific sources of IR.

Lastly, we attempted to discover distinct genes that responded to one specific source, which was more challenging. In all, only 15 genes were upregulated >10-fold in response to one type of particle. Of these genes, moreover, only three had predicted functions (**Table 2**). Proton irradiation again had a larger set of genes than α-particle or deuteron irradiation, of which one interesting member was the receptor for Mating Pheromone A, but the overall pattern still held with this sample, with 9/11 genes having no identifiable functional domains. This lack of highly regulated loci unique to any given particle, particularly in the context of the vast transcriptomic changes taking place during the IR exposure response, reiterates that there is substantial overlap in the type of cellular damage that occurs after acute exposure to each of these IR sources, and that much more detailed work is necessary to tease the different effects apart.

Effect of Melanin on the Response to Charged Particle Radiation Exposure

The final analyses we performed were a comparison of the responses of melanized and non-melanized cultures to these particle sources. Because the particle-associated transcriptome was distinct from our previous γ-radiation study, and particles

TABLE 2 | Genes specifically expressed in response to unique particle irradiation sources, including information on function, conservation, and expression.

IR source	Gene	Uniprot annotation	Conserved domains	Log_2FC	SD	FDR
α	HMPREF1120_06214	Uncharacterized protein	–	4.30	1.2	0.013
	HMPREF1120_04615	Uncharacterized protein	Broad specificity phosphatase PhoE	3.36	1.24	0.004
	HMPREF1120_06237	Uncharacterized protein	–	3.44	0.776	0.05
D	HMPREF1120_01524	Uncharacterized protein	–	3.42	1.30	0.007
p	HMPREF1120_03390	Uncharacterized protein	–	5.46	0.117	0.027
	HMPREF1120_05909	Uncharacterized protein		4.88	0.797	0.044
	HMPREF1120_07028	Uncharacterized protein	–	3.71	0.4	1.14e-06
	HMPREF1120_06554	Uncharacterized protein	–	3.62	0.369	3.33e-06
	HMPREF1120_01962	Uncharacterized protein	–	4.42	0.93	1.26e-04
	HMPREF1120_08985	Gluconate 5-dehydrogenase	Short-chain dehydrogenases/reductases	3.42	0.128	0.006
	HMPREF1120_01617	Uncharacterized protein	–	3.38	0.348	6.97e-08
	HMPREF1120_03162	Uncharacterized protein	–	3.32	1.45	0.013
	HMPREF1120_04175	Uncharacterized protein	–	3.45	0.365	1.68e-04
	HMPREF1120_05389	Uncharacterized protein	–	4.34	2.04	0.009
	HMPREF1120_08176	Pheromone A receptor protein	STE3 superfamily/mating pheromone A receptor	3.4	0.991	4.66e-10

are less penetrating, we were prepared to see some difference between the two strains. However, in agreement with the survival data we collected, the responses had substantial overlap, with comparative plots of Log FC between each sample exhibiting R^2 values of ∼0.9 and greater with no genes significantly regulated in opposite directions between any samples (**Figure 9**), and no differences in highly confident STRING nodes (data not shown). However, we did observe some broad differences between the two responses. Specifically, non-melanized cultures had more enriched biological processes compared to the melanized gene set, even though the former contained a smaller group of genes (**Figure 10**). Enriched categories from this analysis included transcription, mismatch repair, protein catabolism, and autophagy in the upregulated, non-melanized set, while additional genes involved in translation (protein folding, ribosomal biogenesis, RNA processing) and growth and energy production (cell cycle, glycolysis) were enriched among the downregulated genes unique to the non-melanized strain. Interestingly, we have observed several other times that the regulation of translation is affected by melanization status, with ribosomal genes preferentially downregulated, under normal conditions, in melanized strains (Robertson et al., 2012; Schultzhaus Z. et al., 2020). These enrichment analyses provide some further credence to the idea that a lack of melanin mediates a complex response in *E. dermatitidis*, but the response is subtle and ancillary to IR recovery.

DISCUSSION

In this study, we obtained a detailed characterization of the *E. dermatitidis* IR response by performing transcriptomic analysis of melanized and non-melanized cultures after exposure to three different types of ionizing particles. Previously, such an approach has been somewhat contentious. The importance of repair gene induction for cell recovery after irradiation, for example has been challenged (Dornfeld and Livingston, 1991; Birrell et al., 2002). However, advances in bioinformatic analysis and transcriptome profiling have demonstrated that such studies actually can at least provide great information about what IR does to cells. For example, a microarray experiment in *C. neoformans* allowed investigators to determine that ergosterol levels affected IR sensitivity, and while it also lead them to a supposed "dead end" in looking at the importance of catalase genes, the extensive upregulation of redox genes indicated that the cells were responding to oxidative stress (Jung et al., 2016). In our case, we saw the upregulation of nearly all types of DNA repair pathways. Not all of these pathways will be essential for resistance to acute IR exposure if they are deleted, as is the case for the NHEJ pathway in *E. dermatitidis*, which is dispensable for γ-ray resistance (Schultzhaus Z. et al., 2020), but the expression patterns of these genes still allows for an understanding of what is occurring in this organism as a result of the exposure, which is the first step toward augmenting resistance, a central goal of radiobiology.

In our prior study on the transcriptomic response of *E. dermatitidis* to γ-radiation, we came to three basic conclusions (Schultzhaus Z. et al., 2020). First, there is little difference in how melanized and non-melanized *E. dermatitidis* strains respond to IR. Each exhibited a similar survival rate, and each demonstrated an almost complete overlap in their gene expression levels during recovery. Second, recovery from γ-radiation in *E. dermatitidis* is dominated by the induction of DNA repair pathways and the repression of translation, the latter of which is likely indicative of cell cycle arrest that is coupled to DNA damage sensing (Weinert and Hartwell, 1988). Third, that species-specific and poorly characterized proteins also have the potential to play a large part in recovery from IR exposure. Almost half of the transcripts responding to IR, in fact, were annotated as hypothetical proteins, and many of the most highly induced transcripts encoded proteins not conserved outside of the *Exophiala*. One of these was in fact found to be essential for IR resistance (Schultzhaus Z. et al., 2020).

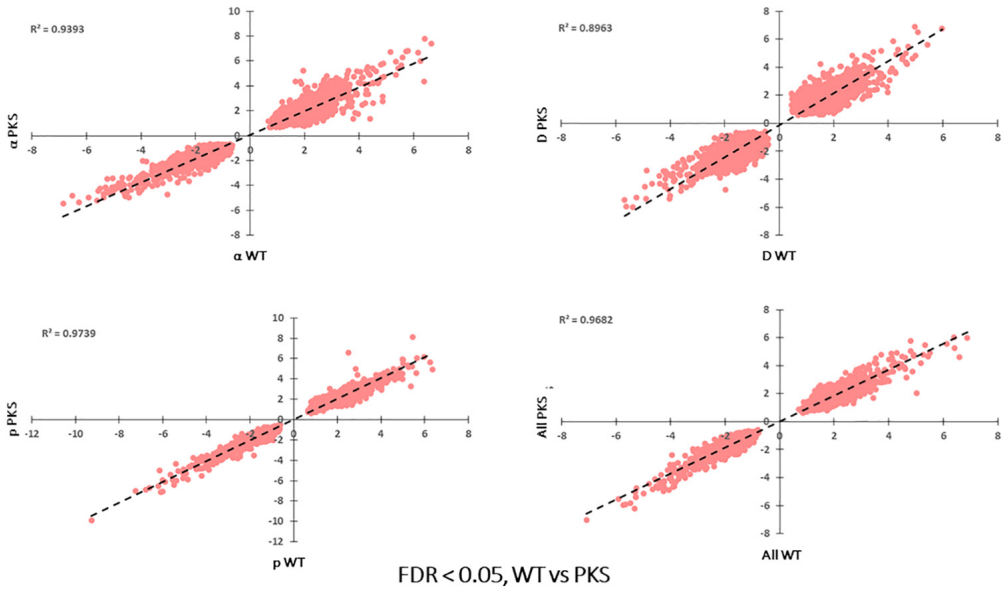


FIGURE 9 | Correlation of log_2FC-values for each significantly regulated gene present WT vs. Δ*pks1*, particle-irradiated gene sets, demonstrating high concordance between the two strains in the response to IR.

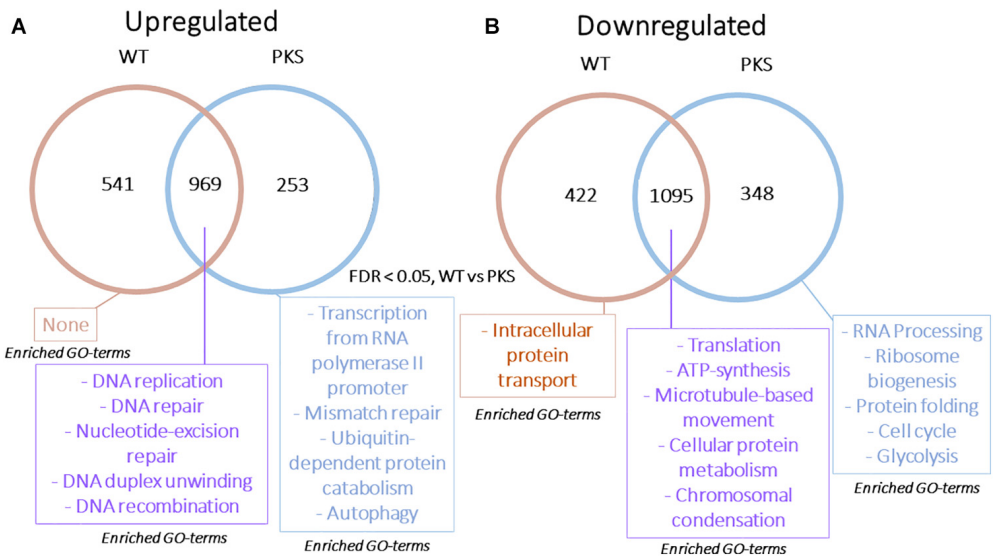


FIGURE 10 | (A) Venn diagram demonstrating the overlap in upregulated or **(B)** downregulated genes in WT and Δ*pks1* particle-responsive datasets. Categories presented for each section include Gene Ontology-Biological Process categories enriched (FDR < 0.01, from FungiFun 2.0) for each group of genes.

We expand upon each of these conclusions with the present study. We reasoned that irradiation by particles, rather than photons, could interact with melanin in a different way, due to the more compact tracks that these particle make through cells (Dadachova et al., 2008; Schweitzer et al., 2009; Upadhyay et al., 2016). However, we once again observed no significant change in survival or gene expression after exposure to any particle source in the absence of melanin. This further supports the conclusion that IR exposure induces a strong and consistent response from the transcriptome that is independent of the presence of melanin, and that melanin does not play a large role in the resistance of this organism to acute IR exposure under the conditions of extremely high, rapid doses that we have used in this study.

It is important to note, however, that many aspects of cellular melanization, including extent, localization, and chemical nature may be affected by the environment, and in turn this may affect how impactful melanin is in mediating the IR recovery response. Using cells with greater amounts of melanin, or changing the irradiation dose rate, could unveil the protective effect seen in some species (Pacelli et al., 2017a). Moreover, there is an

intimate relationship between the components of the cell wall (e.g., carbohydrates such as chitin and glucan, as well as lipids) and melanin, which uses the cell wall as a scaffolding to establish a higher-order structure (Bowman and Free, 2006; Camacho et al., 2017; Chrissian et al., 2020a,b) with emergent properties such as increased free-radical scavenging (Dadachova et al., 2008). This makes the presence of genes involved in cell wall biosynthesis (chitin synthase 1 and glucanase), as well as the TAG lipase especially conspicuous as members of the most differentially expressed transcripts. It is certainly possible that changes inadvertently occur to the protective, strengthening melanin shell around the cell due to this gene regulation after exposure, and this is something that should be explored carefully in the future. There are also some physiological aspects of melanization that need to be studied: we observed a strong transcriptomic response to melanization in *C. neoformans* (Schultzhaus et al., 2019), and here nearly 1000 unique proteins were differentially expressed only in melanized cells in response to particle exposure. These genes represent candidates to concentrate on to determine how melanin interacts with IR, but the observation that they are not enriched for any particular biological function underscores the complex effects melanin has in fungi (Poyntner et al., 2018), something that will take substantial effort to understand.

While the role melanin plays in the *E. dermatitidis* IR response is still obscure, the differences in how photonic and particle irradiation affect this organism have become clearer with this study. The observations we made after breaking down the "hierarchy" of differentially expressed genes in the two groups demonstrates the value of comparative transcriptomic experiments. For example, although all of the canonical processes involved in DNA repair (e.g., DNA synthesis and HR) were observed to be part of the recovery to both types of IR source, the response to particle irradiation involved a more diverse set of genes and an increased magnitude of differential gene expression- 467 were genes differentially expressed with a >5-fold change in every particle-irradiated sample, while this number was only 166 for γ-radiation. Beside being larger in number, the particle-associated genes encompassed a greater diversity of KOG groups, including upregulated genes associated with defense mechanisms (e.g., transporters), lipid and coenzyme transport and metabolism, and energy production and conversion, while there was a conspicuously strong enrichment of ribosomal subunits among the genes that were downregulated after particle exposure. This latter observation was interesting to us, because although we also saw downregulation of translation-associated genes in cultures responding to γ-radiation (Schultzhaus Z. et al., 2020), these were entirely excluded from the list of genes down-regulated five-fold or more. It is possible, then, that the damage that occurs in cells irradiated by ionizing particles requires a stronger cessation of protein synthesis, which may in turn result in massive changes in the proteomic response to these sources, something which we will focus on in follow-up studies.

Another insight we gleaned from the current dataset are the possible roles of autophagy and protein catabolism in recovery from particle irradiation. While DNA is considered to be the prime target for IR-induced damage, and genome instability the major cause of IR-induced cell death, protein oxidation is a known effect of IR-induced free-radical production [for detailed discussions on this topic, see Cabiscol et al. (2000), Braunstein et al. (2009), Daly (2012), Höhn et al. (2013), Radman (2016), Romsdahl et al. (2020), Schultzhaus Z. S. et al. (2020)]. Effects of IR on lipids, including changes in the levels of phospholipids and sterols in the plasma membrane, disruption of regulated membrane transport, and general lipid peroxidation are also known, though to a lesser extent (Stark, 1991; Kiang et al., 2012; Laiakis et al., 2014). Ergosterol biosynthetic genes are involved in IR resistance in *C. neoformans* (Jung et al., 2016, 2017), for example, and membrane permeability appears to be affected in the presence of low dose γ-radiation in *E. dermatitidis* (Robertson et al., 2012). Whether cells are able to deal with this additional, non-genetic damage, or if it also has the potential to render cells incapable of further growth, is unclear. Some studies have demonstrated that proteins involved in the oxidative stress response, such as catalases and superoxide dismutases, play only a minor role in IR resistance, likely because they themselves are damaged rapidly by the oxidative damage that occurs upon exposure (Daly, 2009, 2012; Sharma et al., 2017; Shuryak et al., 2017). In this case, proteins are either protected by the current redox state of the cytoplasm upon irradiation, or they are all damaged and the cell dies. On the other hand, autophagy, which allows cells to turnover damaged macromolecules, may affect IR resistance in eukaryotic cells (Lomonaco et al., 2009; Chaachouay et al., 2011; Höhn et al., 2013). In *S. cerevisiae*, a specific pathway of autophagy was recently found to contribute to DNA repair, and one of the genes involved in that study (ATG8) was also shown to be highly induced here (Eapen et al., 2017). The gene most highly induced by all particle exposures, moreover, was a TAG lipase. In fungi, these enzymes are generally secreted, and are important for modification and breakdown of extracellular lipids during colonization and infection (Gaillardin, 2010; Singh and Mukhopadhyay, 2012; Park et al., 2013). However, the TAG lipase induced here (HMPREF1120_01375) does not possess a signal peptide, so it may be involved in breaking down or otherwise processing intracellular lipid stores that are damaged during irradiation, a process that occurs during autophagy (Christian et al., 2013; van Zutphen et al., 2014). Few studies have looked at autophagy in fungi other than *S. cerevisiae*, but the process is well-characterized in humans, so this particle-associated response should be investigated in further studies.

The accumulation of more RNAseq datasets is, therefore, a useful and simple method for furthering our understanding of radiobiology, especially in organisms with small genomes such as fungi. We anticipate that next, it will begin to help with teasing out differences in the IR response between organisms that vary innate radiation resistance. This has been done, to some extent, in bacteria, where both the extremotolerant *D. radiodurans* and the susceptible *Shewanella oneidensis* have had their radiation-responsive transcriptomes profiled. In this case, distinct mechanisms for dealing with oxidative stress in these organisms were observed (Qiu et al., 2006). Less comparative analysis has been done in fungi, but a recent study has reported on the transcriptional response to both heavy-ion (carbon) beam and X-ray irradiation in *S. cerevisiae* (Guo et al., 2020), which provides a good initial step, as *S. cerevisiae*

is somewhat more susceptible than *E. dermatitidis* to IR. In that study, DNA repair pathways once again dominated the upregulated gene set, and those involved in translation were highly enriched among the downregulated genes. Unlike the data presented above, however, genes involved in the response to oxidative stress were also induced, while autophagy and protein ubiquitination-associated genes were not. The overall number of differentially regulated genes in response to high LET IR was also much smaller for *S. cerevisiae* than here for *E. dermatitidis* (N = 773 vs. >4000 for each condition in the current study). These changes could be affected by the specific conditions used in each study, but that does not disallow them from pointing toward either a more robust or poignant response underlying the higher resistance in *E. dermatitidis*, and at least should direct further studies focused on understanding IR damage in these organisms using more standardized methods.

CONCLUSION

In this study, we have concluded that the transcriptome-wide response to particle irradiation is unique from that of γ-irradiation in the radioresistant yeast *E. dermatitidis*. Specific changes included an induction of autophagy-related genes, a more prominent inhibition of translation, and the identification of hundreds of potentially novel proteins involved in the response to particle exposure. Melanin was found to have little effect on this response, and in fact, the non-melanized mutant was somewhat more resistant to certain doses of radiation exposure. Finally, some loci were identified that consistently respond to IR, which could be used for the future development of IR sensing constructs and strains. These results add to the understanding of how high and low LET radiation affects eukaryotic cells differently and provides a guideline for future transcriptomic and genetic studies on the IR response in fungi.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, GSE152116.

AUTHOR CONTRIBUTIONS

ZS designed and performed the experiments and wrote the manuscript. AC designed and performed the experiments. IS assisted with experiments and analysis of data. ZW designed the experiments, obtained funding, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Defense Threat Reduction Agency Grants HDTRA1-17-1-0013 and HDTRA1-17-1-0020.

ACKNOWLEDGMENTS

The opinions and assertions contained herein are those of the authors and are not to be construed as those of the U.S. Navy, military service at large, or U.S. Government. The authors would like to acknowledge Dr. Seongwon Kim and Dr. Janna Schultzhaus for helpful discussions regarding data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.609996/full#supplementary-material

Supplementary Figure 1 | (A) TPM correlation plots within deuteron, **(B)** proton, and **(C)**α-particle irradiated cultures, within biological replicates and between samples.

Supplementary Figure 2 | STRING network output showing highly confident connections between upregulated genes within shared, upregulated particle genes.

Supplementary Table 1 | Colony forming unit and plating efficiency measurements for melanized (WT) and non-melanized (PKS) strains upon irradiation by each particle type.

Supplementary Table 2 | Best-fit models for each strain detailing parameters for particle irradiation survival curves.

Supplementary Table 3 | Complete output of Salmon and DESeq2 for the particle RNA-seq experiment. Each sheet is labeled with the comparison presented, and within each sheet is the differential expression analysis output, followed by the transcript quantification and length information for both biological replicates of that condition.

Supplementary Table 4 | Genes differentially expressed after irradiation by all particles with fold change values >5 or <−5. Averages of all conditions are also shown, along with the standard deviation.

Supplementary Table 5 | Raw output from MEME analysis of particle-regulated promoters (>5 or <−5-fold) that identified enriched motif associated with particle-responsive genes.

Supplementary Table 6 | Sheet 1: FIMO analysis output from the MEME-Suite, which shows genes with the identified motif present in their promoters throughout the whole *E. dermatitidis* genome. Sheet 2: AME analysis which demonstrates the p-value and E-value for the enrichment of this motif within the highly regulated (>5 or <−5-fold), particle-responsive gene set.

Supplementary Table 7 | STRING output of highly confident connections between genes that were upregulated among all particle sets, including genes making up specific nodes, confidence values for those nodes, and annotations for each gene.

Supplementary Table 8 | The most highly connected, confident nodes identified in the particle-upregulated STRING analysis.

Supplementary Table 9 | STRING output of highly confident connections between genes that were upregulated within each particle dataset, including genes making up specific nodes, confidence values for those nodes, and annotations for each gene.

Supplementary Table 10 | STRING output of highly confident node connection annotations present in all WT particle irradiation-responsive datasets.

Supplementary Table 11 | STRING output of highly confident node connection annotations present in only two WT particle irradiation-responsive datasets.

Supplementary Table 12 | STRING output of highly confident node connection annotations present in only one WT particle irradiation-responsive dataset.

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49.

Battista, J. R., Earl, A. M., and Park, M.-J. (1999). Why is *Deinococcus radiodurans* so resistant to ionizing radiation? *Trends Microbiol.* 7, 362–365. doi: 10.1016/s0966-842x(99)01566-8

Beau, I., Esclatine, A., and Codogno, P. (2008). Lost to translation: when autophagy targets mature ribosomes. *Trends Cell Biol.* 18, 311–314. doi: 10.1016/j.tcb.2008.05.001

Birrell, G. W., Brown, J. A., Wu, H. I., Giaever, G., Chu, A. M., Davis, R. W., et al. (2002). Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8778–8783. doi: 10.1073/pnas.132275199

Bowman, S. M., and Free, S. J. (2006). The structure and synthesis of the fungal cell wall. *Bioessays* 28, 799–808.

Braunstein, S., Badura, M. L., Xi, Q., Formenti, S. C., and Schneider, R. J. (2009). Regulation of protein synthesis by ionizing radiation. *Mol. Cell. Biol.* 29, 5645–5656. doi: 10.1128/mcb.00711-09

Cabiscol, E., Piulats, E., Echave, P., Herrero, E., and Ros, J. (2000). Oxidative stress promotes specific protein damage in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 275, 27393–27398.

Camacho, E., Chrissian, C., Cordero, R. J., Liporagi-Lopes, L., Stark, R. E., and Casadevall, A. (2017). N-acetylglucosamine affects *Cryptococcus neoformans* cell-wall composition and melanin architecture. *Microbiology* 163:1540. doi: 10.1099/mic.0.000552

Carlone, M., Wilkins, D., Nyiri, B., and Raaphorst, P. (2004). TCP isoeffect analysis using a heterogeneous distribution of radiosensitivity. *Med. Phys.* 31, 1176–1182. doi: 10.1118/1.1711415

Chaachouay, H., Ohneseit, P., Toulany, M., Kehlbach, R., Multhoff, G., and Rodemann, H. P. (2011). Autophagy contributes to resistance of tumor cells to ionizing radiation. *Radiother. Oncol.* 99, 287–292. doi: 10.1016/j.radonc.2011.06.002

Chen, Z., Martinez, D. A., Gujja, S., Sykes, S. M., Zeng, Q., Szaniszlo, P. J., et al. (2014). Comparative genomic and transcriptomic analysis of *Wangiella dermatitidis*, a major cause of phaeohyphomycosis and a model black yeast human pathogen. *G3* 4, 561–578. doi: 10.1534/g3.113.009241

Chrissian, C., Camacho, E., Fu, M. S., Prados-Rosales, R., Chatterjee, S., Cordero, R. J., et al. (2020a). Melanin deposition in two *Cryptococcus species* depends on cell-wall composition and flexibility. *J. Biol. Chem.* 295, 1815–1828. doi: 10.1074/jbc.ra119.011949

Chrissian, C., Camacho, E., Kelly, J. E., Wang, H., Casadevall, A., and Stark, R. E. (2020b). Solid-state NMR spectroscopy identifies three classes of lipids in *Cryptococcus neoformans* melanized cell walls and whole fungal cells. *J. Biol. Chem.* 295, 15083–15096. doi: 10.1074/jbc.ra120.015201

Christian, P., Sacco, J., and Adeli, K. (2013). Autophagy: emerging roles in lipid homeostasis and metabolic control. *Biochim. Biophys. Acta* 1831, 819–824. doi: 10.1016/j.bbalip.2012.12.009

Cingolani, F., and Czaja, M. J. (2016). Regulation and functions of autophagic lipolysis. *Trends Endocrinol. Metab.* 27, 696–705. doi: 10.1016/j.tem.2016.06.003

Clark, A. (1991). rec genes and homologous recombination proteins in *Escherichia coli*. *Biochimie* 73, 523–532. doi: 10.1016/0300-9084(91)90124-j

Commer, B., and Shaw, B. D. (2020). Current views on endocytosis in filamentous fungi. *Mycology* doi: 10.1080/21501203.2020.1741471

Critchlow, S. E., and Jackson, S. P. (1998). DNA end-joining: from yeast to man. *Trends Biochem. Sci.* 23, 394–398. doi: 10.1016/s0968-0004(98)01284-5

Cucinotta, F. A., and Durante, M. (2006). Cancer risk from exposure to galactic cosmic rays: implications for space exploration by human beings. *Lancet Oncol.* 7, 431–435. doi: 10.1016/s1470-2045(06)70695-7

Dadachova, E., Bryan, R. A., Howell, R. C., Schweitzer, A. D., Aisen, P., Nosanchuk, J. D., et al. (2008). The radioprotective properties of fungal melanin are a function of its chemical composition, stable radical presence and spatial arrangement. *Pigment Cell Melanoma Res.* 21, 192–199. doi: 10.1111/j.1755-148x.2007.00430.x

Dadachova, E., Bryan, R. A., Huang, X., Moadel, T., Schweitzer, A. D., Aisen, P., et al. (2007). Ionizing radiation changes the electronic properties of melanin and enhances the growth of melanized fungi. *PLoS One* 2:e457. doi: 10.1371/journal.pone.0000457

Dadachova, E., and Casadevall, A. (2008). Ionizing radiation: how fungi cope, adapt, and exploit with the help of melanin. *Curr. Opin. Microbiol.* 11, 525–531. doi: 10.1016/j.mib.2008.09.013

Daly, M. J. (2009). A new perspective on radiation resistance based on *Deinococcus radiodurans*. *Nat. Rev. Microbiol.* 7, 237–245. doi: 10.1038/nrmicro2073

Daly, M. J. (2012). Death by protein damage in irradiated cells. *DNA Repair* 11, 12–21. doi: 10.1016/j.dnarep.2011.10.024

Daly, M. J., Gaidamakova, E. K., Matrosova, V., Vasilenko, A., Zhai, M., Venkateswaran, A., et al. (2004). Accumulation of Mn (II) in *Deinococcus radiodurans* facilitates gamma-radiation resistance. *Science* 306, 1025–1028. doi: 10.1126/science.1103185

DiCarlo, A. L., Tamarat, R., Rios, C. I., Benderitter, M., Czarniecki, C. W., Allio, T. C., et al. (2017). Cellular Therapies for treatment of radiation injury: report from a NIH/NIAID and IRSN workshop. *Radiat. Res.* 188, e54–e75.

Doble, P. A., and Miklos, G. L. G. (2018). Distributions of manganese in diverse human cancers provide insights into tumour radioresistance. *Metallomics* 10, 1191–1210. doi: 10.1039/c8mt00110c

Dornfeld, K. J., and Livingston, D. M. (1991). Effects of controlled RAD52 expression on repair and recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 11, 2013–2017. doi: 10.1128/mcb.11.4.2013

Eapen, V. V., Waterman, D. P., Bernard, A., Schiffmann, N., Sayas, E., Kamber, R., et al. (2017). A pathway of targeted autophagy is induced by DNA damage in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1158–E1167.

Fano, U. (1963). Penetration of protons, alpha particles, and mesons. *Annu. Rev. Nucl. Sci.* 13, 1–66. doi: 10.1146/annurev.ns.13.120163.000245

Feng, B., Wang, X., Hauser, M., Kaufmann, S., Jentsch, S., Haase, G., et al. (2001). Molecular cloning and characterization of WdPKS1, a gene involved in dihydroxynaphthalene melanin biosynthesis and virulence in *Wangiella* (*Exophiala*) *dermatitidis*. *Infect. Immun.* 69, 1781–1794. doi: 10.1128/iai.69.3.1781-1794.2001

Gaillardin, C. (2010). "Lipases as pathogenicity factors of fungi," in *Handbook of Hydrocarbon and Lipid Microbiology*, ed. K. N. Timmis (Berlin: Heidelberg).

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704.

Gruenig, M. C., Lu, D., Won, S. J., Dulberger, C. L., Manlick, A. J., Keck, J. L., et al. (2011). Creating directed double-strand breaks with the Ref protein A novel RecA-dependent nuclease from bacteriophage P1. *J. Biol. Chem.* 286, 8240–8251. doi: 10.1074/jbc.m110.205088

Guo, X., Zhang, M., Gao, Y., Lu, D., Li, W., and Zhou, L. (2020). Repair characteristics and time-dependent effects in response to heavy-ion beam irradiation in *Saccharomyces cerevisiae*: a comparison with X-ray irradiation. *Appl. Microbiol. Biotechnol.* 104, 4043–4057. doi: 10.1007/s00253-020-10464-8

Hada, M., and Georgakilas, A. G. (2008). Formation of clustered DNA damage after high-LET irradiation: a review. *J. Radiat. Res.* 49, 203–210. doi: 10.1269/jrr.07123

Hall, E. J., and Giaccia, A. J. (2006). *Radiobiology for the Radiologist*. Philadelphia, PA: Lippincott Williams & Wilkins.

Hands, S. L., Proud, C. G., and Wyttenbach, A. (2009). mTOR's role in ageing: protein synthesis or autophagy? *Aging* 1, 586–597. doi: 10.18632/aging.100070

Höhn, A., König, J., and Grune, T. (2013). Protein oxidation in aging and the removal of oxidized proteins. *J. Proteom.* 92, 132–159. doi: 10.1016/j.jprot.2013.01.004

Holliday, R. (1975). Further evidence for an inducible recombination repair system in *Ustilago maydis*. *Mut. Res.* 29, 149–153. doi: 10.1016/0027-5107(75)90029-9

Holloman, W. K., Schirawski, J., and Holliday, R. (2007). Towards understanding the extreme radiation resistance of *Ustilago maydis*. *Trends Microbiol.* 15, 525–529. doi: 10.1016/j.tim.2007.10.007

Jaishy, B., and Abel, E. D. (2016). Lipids, lysosomes, and autophagy. *J. Lipid Res.* 57, 1619–1635. doi: 10.1194/jlr.r067520

Jung, K.-W., Lim, S., and Bahn, Y.-S. (2017). Microbial radiation-resistance mechanisms. *J. Microbiol.* 55, 499–507. doi: 10.1007/s12275-017-7242-5

Jung, K.-W., Yang, D.-H., Kim, M.-K., Seo, H. S., Lim, S., and Bahn, Y.-S. (2016). Unraveling fungal radiation resistance regulatory networks through the

genome-wide transcriptome and genetic analyses of *Cryptococcus neoformans*. *MBio* 7:e1483-16.

Kiang, J. G., Fukumoto, R., and Gorbunov, N. V. (2012). *Lipid Peroxidation After Ionizing Irradiation Leads to Apoptosis and Autophagy, Lipid Peroxidation*. London: InTech Open Access Publisher, 261–278.

Kraft, C., Deplazes, A., Sohrmann, M., and Peter, M. (2008). Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. *Nat. Cell Biol.* 10, 602–610. doi: 10.1038/ncb1723

Laiakis, E. C., Strassburg, K., Bogumil, R., Lai, S., Vreeken, R. J., Hankemeier, T., et al. (2014). Metabolic phenotyping reveals a lipid mediator response to ionizing radiation. *J. Proteome Res.* 13, 4143–4154. doi: 10.1021/pr5005295

Lee, M. G., and Yarranton, G. T. (1982). Inducible DNA repair in *Ustilago maydis*. *Mol. Gen. Genet.* 185, 245–250. doi: 10.1007/bf00330793

Loftus, B. J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., et al. (2005). The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307, 1321–1324. doi: 10.1126/science.1103773

Lomonaco, S. L., Finniss, S., Xiang, C., Decarvalho, A., Umansky, F., Kalkanis, S. N., et al. (2009). The induction of autophagy by γ-radiation contributes to the radioresistance of glioma stem cells. *Int. J. Cancer* 125, 717–722. doi: 10.1002/ijc.24402

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Ma, L., Kazama, Y., Hirano, T., Morita, R., Tanaka, S., Abe, T., et al. (2018). LET dependence on killing effect and mutagenicity in the model filamentous fungus *Neurospora crassa*. *Int. J. Radiat. Biol.* 94, 1125–1133. doi: 10.1080/09553002.2019.1524940

Matuo, Y., Izumi, Y., Sakamoto, N. A., Hase, Y., Satoh, K., and Shimizu, K. (2019). Molecular analysis of carbon ion-induced mutations in DNA repair-deficient strains of *Saccharomyces cerevisiae*. *Quantum Beam Sci.* 3:14. doi: 10.3390/qubs3030014

McEntee, K., Weinstock, G., and Lehman, I. (1979). Initiation of general recombination catalyzed in vitro by the recA protein of *Escherichia coli*. *Proc. Natl. Acad. Sci.U.S.A.* 76, 2615–2619. doi: 10.1073/pnas.76.6.2615

Mering, C. V., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261. doi: 10.1093/nar/gkg034

Milisavljevic, M., Petkovic, J., Samardzic, J., and Kojic, M. (2018). Bioavailability of nutritional resources from cells killed by oxidation supports expansion of survivors in *Ustilago maydis* populations. *Front. Microbiol.* 9:990. doi: 10.3389/fmicb.2018.00990

Nikjoo, H., Munson, R. J., and Bridges, B. A. (1999). RBE-LET relationships in mutagenesis by ionizing radiation. *J. Radiat. Res.* 40, S85–S105.

Oliveros, J. C. (2007). *VENNY. An Interactive Tool for Comparing Lists with Venn Diagrams*. Available online at: http://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed March 24, 2020).

Ostapenko, D., Burton, J. L., and Solomon, M. J. (2015). The Ubp15 deubiquitinase promotes timely entry into S phase in Saccharomyces cerevisiae. *Mol. Biol. Cell* 26, 2205–2216. doi: 10.1091/mbc.e14-09-1400

Pacelli, C., Bryan, R. A., Onofri, S., Selbmann, L., Shuryak, I., and Dadachova, E. (2017a). Melanin is effective in protecting fast and slow growing fungi from various types of ionizing radiation. *Environ. Microbiol.* 19, 1612–1624. doi: 10.1111/1462-2920.13681

Pacelli, C., Selbmann, L., Zucconi, L., Raguse, M., Moeller, R., Shuryak, I., et al. (2017b). Survival, DNA integrity, and ultrastructural damage in Antarctic cryptoendolithic eukaryotic microorganisms exposed to ionizing radiation. *Astrobiology* 17, 126–135. doi: 10.1089/ast.2015.1456

Park, M., Do, E., and Jung, W. H. (2013). Lipolytic enzymes involved in the virulence of human pathogenic fungi. *Mycobiology* 41, 67–72. doi: 10.5941/myco.2013.41.2.67

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Poyntner, C., Mirastschijski, U., Sterflinger, K., and Tafer, H. (2018). Transcriptome study of an *Exophiala dermatitidis* PKS1 mutant on an ex vivo skin model: is melanin important for infection? *Front. Microbiol.* 9:1457. doi: 10.3389/fmicb.2018.01457

Priebe, S., Kreisel, C., Horn, F., Guthke, R., and Linde, J. (2014). FungiFun2: a comprehensive online resource for systematic analysis of gene lists from fungal species. *Bioinformatics* 31, 445–446. doi: 10.1093/bioinformatics/btu627

Priebe, S., Linde, J., Albrecht, D., Guthke, R., and Brakhage, A. A. (2011). FungiFun: a web-based application for functional categorization of fungal genes and proteins. *Fungal Genet. Biol.* 48, 353–358. doi: 10.1016/j.fgb.2010.11.001

Qiu, X., Daly, M. J., Vasilenko, A., Omelchenko, M. V., Gaidamakova, E. K., Wu, L., et al. (2006). Transcriptome analysis applied to survival of *Shewanella oneidensis* MR-1 exposed to ionizing radiation. *J. Bacteriol.* 188, 1199–1204. doi: 10.1128/jb.188.3.1199-1204.2006

Radman, M. (2016). Protein damage, radiation sensitivity and aging. *DNA Repair* 44, 186–192. doi: 10.1016/j.dnarep.2016.05.025

Riley, P. (1994). Free radicals in biology: oxidative stress and the effects of ionizing radiation. *Int. J. Radiat. Biol.* 65, 27–33. doi: 10.1080/09553009414550041

Roberts, J. W., Roberts, C., Craig, N. L., and Phizicky, E. (1979). "Activity of the *Escherichia coli* recA-gene product. *Cold Spring Harbor Symp. Quant. Biol.* 43(Pt 2), 917–920.

Robertson, K. L., Mostaghim, A., Cuomo, C. A., Soto, C. M., Lebedev, N., Bailey, R. F., et al. (2012). Adaptation of the black yeast *Wangiella dermatitidis* to ionizing radiation: molecular and cellular mechanisms. *PLoS One* 7:e48674. doi: 10.1371/journal.pone.0048674

Romsdahl, J., Schultzhaus, Z., Chen, A., Liu, J., Ewing, A., Hervey, J., et al. (2020). Adaptive evolution of a melanized fungus reveals robust augmentation of radiation resistance by abrogating non-homologous end-joining. *Environ. Microbiol.* doi: 10.1111/1462-2920.15285 Online ahead of print

RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio Inc: Boston, MA.

Sargentini, N. J., and Smith, K. C. (1986). Quantitation of the involvement of the recA, recB, recC, recF, recJ, recN, lexA, radA, radB, uvrD, and umuC genes in the repair of X-ray-induced DNA double-strand breaks in *Escherichia coli*. *Radiat. Res.* 107, 58–72. doi: 10.2307/3576850

Schultzhaus, Z., Chen, A., Kim, S., Shuryak, I., Chang, M., and Wang, Z. (2019). Transcriptomic analysis reveals the relationship of melanization to growth and resistance to gamma radiation in *Cryptococcus neoformans*. *Environ. Microbiol.* 21, 2613–2628. doi: 10.1111/1462-2920.14550

Schultzhaus, Z., Romsdahl, J., Chen, A., Tschirhart, T., Kim, S., Leary, D. H., et al. (2020). The response of the melanized yeast *Exophiala dermatitidis* to gamma radiation exposure. *Environ. Microbiol.* 22, 1310–1326. doi: 10.1111/1462-2920.14936

Schultzhaus, Z. S., Schultzhaus, J. N., Romsdahl, J., Chen, A., Hervey Iv, W. J., Leary, D. H., et al. (2020). Proteomics reveals distinct changes associated with increased gamma radiation resistance in the black yeast *Exophiala dermatitidis*. *Genes* 11:1128. doi: 10.3390/genes11101128

Schweitzer, A. D., Howell, R. C., Jiang, Z., Bryan, R. A., Gerfen, G., Chen, C.-C., et al. (2009). Physico-chemical evaluation of rationally designed melanins as novel nature-inspired radioprotectors. *PLoS One* 4:e7229. doi: 10.1371/journal.pone.0007229

Sharma, A., Gaidamakova, E. K., Grichenko, O., Matrosova, V. Y., Hoeke, V., Klimenkova, P., et al. (2017). Across the tree of life, radiation resistance is governed by antioxidant Mn2+, gauged by paramagnetic resonance. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9253–E9260.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in *Saccharomyces cerevisiae* is a RecA-like protein. *Cell* 69, 457–470. doi: 10.1016/0092-8674(92)90447-k

Shuryak, I., Bryan, R. A., Nosanchuk, J. D., and Dadachova, E. (2014). Mathematical modeling predicts enhanced growth of X-ray irradiated pigmented fungi. *PLoS One* 9:e85561. doi: 10.1371/journal.pone.0085561

Shuryak, I., Carlson, D. J., Brown, J. M., and Brenner, D. J. (2015). High-dose and fractionation effects in stereotactic radiation therapy: analysis of tumor control data from 2965 patients. *Radiother. Oncol.* 115, 327–334. doi: 10.1016/j.radonc.2015.05.013

Shuryak, I., Matrosova, V. Y., Gaidamakova, E. K., Grichenko, O., Klimenkova, P., Volpe, R. P., et al. (2017). Microbial cells can cooperate to resist high-level chronic ionizing radiation. *PLoS One* 12:e0189261. doi: 10.1371/journal.pone.0189261

Singh, A. K., and Mukhopadhyay, M. (2012). Overview of fungal lipase: a review. *Appl. Biochem. Biotechnol.* 166, 486–520. doi: 10.1007/s12010-011-9444-3

Singh, S., Brocker, C., Koppaka, V., Chen, Y., Jackson, B. C., Matsumoto, A., et al. (2013). Aldehyde dehydrogenases in cellular responses to oxidative/electrophilic stress. *Free Radic. Biol. Med.* 56, 89–101.

Smith, K. M., Phatale, P. A., Bredeweg, E. L., Connolly, L. R., Pomraning, K. R., and Freitag, M. (2006). *Meyers: Encyclopedia of Molecular Cell Biology and Molecular Medicine: Epigenetic Regulation and Epigenomics, Second Edition*. ed. R. A. Meyers, 1–28.

Snel, B., Lehmann, G., Bork, P., and Huynen, M. A. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 28, 3442–3444. doi: 10.1093/nar/28.18.3442

Stark, G. (1991). The effect of ionizing radiation on lipid membranes. *Biochim. Biophys. Acta* 1071, 103–122. doi: 10.1016/0304-4157(91)90020-w

United Nations Scientific Committee on the Effects of Atomic Radiation (2000). *Sources and Effects of Ionizing Radiation: Sources*. New York, NY: United Nations Scientific Committee on the Effects of Atomic Radiation.

Upadhyay, S., and Shaw, B. D. (2008). The role of actin, fimbrin and endocytosis in growth of hyphae in *Aspergillus nidulans*. *Mol. Microbiol.* 68, 690–705. doi: 10.1111/j.1365-2958.2008.06178.x

Upadhyay, S., Xu, X., Lowry, D., Jackson, J. C., Roberson, R. W., and Lin, X. (2016). Subcellular compartmentalization and trafficking of the biosynthetic machinery for fungal melanin. *Cell Rep.* 14, 2511–2518. doi: 10.1016/j.celrep.2016.02.059

van Zutphen, T., Todde, V., De Boer, R., Kreim, M., Hofbauer, H. F., Wolinski, H., et al. (2014). Lipid droplet autophagy in the yeast Saccharomyces cerevisiae. *Mol. Biol. Cell* 25, 290–301. doi: 10.1091/mbc.e13-08-0448

Ward, J. F. (1988). DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability. *Prog. Nucleic Acid Res. Mol. Biol.* 35, 95–125. doi: 10.1016/s0079-6603(08)60611-x

Weinert, T. A., and Hartwell, L. H. (1988). The RAD9 gene controls the cell cycle response to DNA damage in *Saccharomyces cerevisiae*. *Science* 241, 317–322. doi: 10.1126/science.3291120

Wyttenbach, A., Hands, S., King, M. A., Lipkow, K., and Tolkovsky, A. M. (2008). Amelioration of protein misfolding disease by rapamycin: translation or autophagy? *Autophagy* 4, 542–545. doi: 10.4161/auto.6059

Zakrzewska, A., Van Eikenhorst, G., Burggraaff, J. E., Vis, D. J., Hoefsloot, H., Delneri, D., et al. (2011). Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. *Mol. Biol. Cell* 22, 4435–4446. doi: 10.1091/mbc.e10-08-0721

Zhdanova, N. N., Zakharchenko, V. A., Vember, V. V., and Nakonechnaya, L. T. (2000). Fungi from Chernobyl: mycobiota of the inner regions of the containment structures of the damaged nuclear reactor. *Mycol. Res.* 104, 1421–1426. doi: 10.1017/s0953756200002756

Conflict of Interest: 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 © 2021 Schultzhaus, Chen, Shuryak and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fmicb.2020.606612

Gemmatimonas groenlandica sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes

Yonghui Zeng[1,2†], Nupur[3], Naicheng Wu[4], Anne Mette Madsen[5], Xihan Chen[6†], Alastair T. Gardiner[3] and Michal Koblížek[3]*

[1] Department of Environmental Science, Aarhus University, Roskilde, Denmark, [2] Aarhus Institute of Advanced Studies, Aarhus University, Aarhus, Denmark, [3] Centre Algatech, Institute of Microbiology CAS, Třeboň, Czechia, [4] Department of Geography and Spatial Information Techniques, Center for Land and Marine Spatial Utilization and Governance Research, Ningbo University, Ningbo, China, [5] The National Research Centre for the Working Environment, Copenhagen, Denmark, [6] Department of Engineering, Aarhus University, Aarhus, Denmark

The bacterial phylum Gemmatimonadetes contains members capable of performing bacteriochlorophyll-based phototrophy (chlorophototrophy). However, only one strain of chlorophototrophic Gemmatimonadetes bacteria (CGB) has been isolated to date, hampering our further understanding of their photoheterotrophic lifestyle and the evolution of phototrophy in CGB. By combining a culturomics strategy with a rapid screening technique for chlorophototrophs, we report the isolation of a new member of CGB, *Gemmatimonas (G.) groenlandica* sp. nov., from the surface water of a stream in the Zackenberg Valley in High Arctic Greenland. Distinct from the microaerophilic *G. phototrophica* strain AP64^T, *G. groenlandica* strain TET16^T is a strictly aerobic anoxygenic phototroph, lacking many oxygen-independent enzymes while possessing an expanded arsenal for coping with oxidative stresses. Its pigment composition and infra-red absorption properties are also different from *G. phototrophica*, indicating that it possesses a different photosystem apparatus. The complete genome sequence of *G. groenlandica* reveals unique and conserved features in the photosynthesis gene clusters of CGB. We further analyzed metagenome-assembled genomes of CGB obtained from soil and glacier metagenomes from Northeast Greenland, revealing a wide distribution pattern of CGB beyond the stream water investigated.

Keywords: MALDI-TOF MS, bacterial isolation, phototrophy, Gemmatimonadetes, oligotrophic environment

IMPORTANCE

The bacterial phylum Gemmatimonadetes is an important but yet understudied group in natural microbial communities. The isolation of the only phototrophic member of this phylum, *Gemmatimonas phototrophica*, was reported in 2014, which expanded the list of known bacterial phyla capable of performing photosynthesis. Since then, no new phototrophic member of this phylum has been isolated. By applying a novel isolation strategy of combining a mass spectroscopy-based high-throughput profiling method and a rapid screening technique for phototrophic bacterial colonies, we successfully isolated the second phototrophic member of this phylum, *Gemmatimonas*

OPEN ACCESS

Edited by:
Brian P. Hedlund,
University of Nevada, Las Vegas,
United States

Reviewed by:
Marike Palmer,
University of Nevada, Las Vegas,
United States
Javier Pascual,
Darwin Bioprospecting Excellence,
Spain
Lewis Michael Ward,
Harvard University, United States

***Correspondence:**
Yonghui Zeng
zeng@plen.ku.dk;
yonghui.sci@gmail.com

†Present address:
Yonghui Zeng,
Department of Plant
and Environmental Sciences,
University of Copenhagen,
Copenhagen, Denmark
Xihan Chen,
BGI Europe A/S, Copenhagen,
Denmark

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

Received: 15 September 2020
Accepted: 08 December 2020
Published: 15 January 2021

Citation:
Zeng Y, Nupur, Wu N,
Madsen AM, Chen X, Gardiner AT
and Koblížek M (2021)
Gemmatimonas groenlandica sp. nov.
Is an Aerobic Anoxygenic Phototroph
in the Phylum Gemmatimonadetes.
Front. Microbiol. 11:606612.
doi: 10.3389/fmicb.2020.606612

groenlandica, from a stream in Northeast Greenland. Its discovery confirms the widespread presence of phototrophic Gemmatimonadetes bacteria in the environment and raises an intriguing question on the evolutionary history of phototrophy in the phylum Gemmatimonadetes. Distinct from the microaerophilic slow growth rate in *G. phototrophica*, *G. groenlandica* is a strict aerobe and can be readily cultured in liquid medium, opening new possibilities for future strain genetic engineering and detailed photophysiological studies.

STRAIN INFORMATION

Gemmatimonas groenlandica strain TET16^T has been deposited to the Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany under accession no. DSM110279 and to the China General Microbiological Culture Collection Center under accession no. CGMCC1.18661.

INTRODUCTION

Members of bacterial phylum Gemmatimonadetes are widely distributed in natural microbial communities, ranked as one of the nine most abundant phyla found in soils (Janssen, 2006; Youssef and Elshahed, 2009) with a mean abundance of 2.2% of the total soil bacteria (DeBruyn et al., 2011). A more recent survey of 1,706 metagenomes from various environments that were deposited into the MG-RAST server (Wilke et al., 2016) showed that Gemmatimonadetes constitute up to 2.54% total reads with a median value of 0.24% (calculated based on the data in Supplementary Table 1 in Zeng et al., 2016). Gemmatimonadetes are most abundant in soils, wastewater treatment-related samples, biofilms, and plant-associated habitats with the largest proportion (2.54%) reported in an Arctic tundra permafrost metagenome (Zeng et al., 2016). Diversity surveys based on 16S rRNA genes indicate that Gemmatimonadetes are well adapted not only to arid but also to oligotrophic conditions (Hanada and Sekiguchi, 2014).

Despite the widespread distribution of Gemmatimonadetes in the environment, their physiology, ecology and importance in environmental processes are poorly understood (DeBruyn et al., 2011; Hanada and Sekiguchi, 2014). One critical obstacle to an improved understanding of the ecological roles of Gemmatimonadetes is that, since the establishment of the phylum Gemmatimonadetes in 2003 (Zhang et al., 2003), only a few members have been isolated as pure cultures available for detailed studies in the laboratory. To date, the validated type strains in Gemmatimonadetes (updated list accessed at[1]) include *Gemmatimonas (G.) aurantiaca* T-27^T isolated from activated sludge in a wastewater treatment plant (Zhang et al., 2003), *Gemmatimonas phototrophica* AP64^T from a desert lake (Zeng et al., 2015), *Longimicrobium terrae* CB-286315^T from Mediterranean forest soil (Pascual et al., 2016), and *Roseisolibacter agri* AW1220^T from agricultural floodplain soil (Pascual et al., 2018). Studies of this limited number of type strains have revealed some ecologically important metabolisms in Gemmatimonadetes. For instance, *Gemmatimonas aurantiaca* T-27^T is capable of reducing the potent greenhouse gas N_2O under both anaerobic and aerobic conditions (Park et al., 2017; Chee-Sanford et al., 2019); *Gemmatimonas phototrophica* AP64^T is a microaerophilic, facultative photoheterotroph capable of harvesting light energy (Zeng et al., 2015; Koblížek et al., 2020).

Before this study, *G. phototrophica* AP64^T represented the only phototrophic isolate known in the phylum Gemmatimonadetes, possessing type-2 reaction centers that are possibly of proteobacterial origin based on the phylogenies of bacteriochlorophyll biosynthesis genes and the organization of its photosynthesis gene cluster (PGC) (Zeng et al., 2014). Despite the close phylogenetic relationship with purple photosynthetic Proteobacteria (Zeng et al., 2014), the PGC of *G. phototrophica* appears to display a unique feature that the *acsF* gene (involved in BChl biosynthesis pathway) is located between the *bchFNBHLM* and *puhABC* sub-clusters (see Figure 8 in Zeng and Koblížek, 2017), implying an, as yet unknown, evolutionary history of the PGC and phototrophy in CGB. More isolates of CGB are required to test whether the photosynthesis-related genomic and physiological characteristics observed in *G. phototrophica* are common features in all CGB members.

A metagenomic survey using *acsF* (encoding the Mg-protoporphyrin IX monomethyl ester oxidative cyclase) as the marker gene revealed that CGB are widely distributed in various environments, including air, river waters/sediment, estuarine waters, lake waters, biofilms, plant surfaces, intertidal sediment, soils, springs, and wastewater treatment plants, but not in marine systems (Zeng et al., 2016). The wide distribution of CGB in nature and its relatively high abundance among the phototrophic microbial community (0.4–11.9%: Zeng et al., 2016) provide enormous opportunities for the isolation of new CGB members from the environment.

In polar terrestrial environments, phototrophic bacteria have been understudied largely due to the difficulties in sampling and commonly perceived low activities of phototrophic bacteria caused by freezing temperatures and prolonged darkness in winter. A recent bacterial cultivation effort on Antarctic soils identified 330 possibly aerobic anoxygenic phototrophs (Tahon and Willems, 2017), highlighting that polar terrestrial environments could be an untapped source of novel bacterial phototrophs. In this study, we focused on the High Arctic environment in Northeast Greenland with the aim to isolate novel members of CGB. By combining a high-throughput culturomics approach (Lagier et al., 2018) with the rapid screening technique for bacteriochlorophyll-containing colonies (Zeng et al., 2014), we successfully isolated the second chlorophototrophic member of Gemmatimonadetes, *Gemmatimonas groenlandica* TET16^T, from a stream water sample from the Zackenberg Valley. Phenotypic and genotypic comparisons of *G. groenlandica* with *G. phototrophica* allow us to reveal the unique physiological and genomic features in CGB. *G. groenlandica* strain TET16^T represents the first fully aerobic anoxygenic photoheterotroph in the phylum Gemmatimonadetes.

[1]https://lpsn.dsmz.de/

RESULTS AND DISCUSSION

A Culturomics Strategy led to the Isolation of *Gemmatimonas groenlandica* sp. nov. Strain TET16^T

We adopted a culturomics strategy that combined a high-throughput colony screening approach using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) and genome sequencing (Lagier et al., 2018) to search for novel Gemmatimonadetes bacteria in a stream in Northeast Greenland (**Figure 1**). The rapid detection technique of BChl *a* fluorescence emitted from chlorophototrophic colonies (Zeng et al., 2014) allowed us to only load BChl-containing colonies into the MALDI-TOF mass spectrometer. Based on the colony morphology of the only CGB isolate, *Gemmatimonas phototrophica* strain AP64^T (Zeng et al., 2015), the search focus was placed on small colonies with pink or reddish color. With a modest screening effort of ∼500 phototrophic colonies grown on heterotrophic media, we found the candidate CGB strain TET16^T, the colonies of which appeared after 5-week incubation on 1/5 R2A agar supplemented with 8 μg/mL tetracycline. The colony appeared as a circular shape with a size of 1–2 mm. Strain TET16 was the only one in our culture collection that clustered with *G. phototrophica* and *G. aurantiaca* on the dendrogram of MALDI-TOF MS profiles, while distantly related to Alphaproteobacteria (**Figure 2A**).

The complete genome sequence of strain TET16^T confirmed it belongs to the phylum Gemmatimonadetes. The 16S rRNA gene phylogeny placed TET16^T, *G. phototrophica*, and *G. aurantiaca* into the same cluster (**Figure 2B**). The environmental clones that clustered with *G. groenlandica* originate from various environments, including Tibetan lake water, Korean brackish water, French soil, and lake sediment in Japan (**Figure 2B**). A similar pattern in source environments was also observed on the environmental clones clustering with *G. phototrophica* (Zeng et al., 2014), indicating the wide distribution of CGB in natural environments. This was in line with our previous finding from a survey of public metagenomic databases that CGB were present in diverse environments (Zeng et al., 2016). The phylogenetic tree of the *acsF* gene that is involved in bacteriochlorophyll biosynthesis shows that strain TET16^T is more closely related to a pelagic bacterium (20 m deep) of Lake Baikal in Russia (Cabello-Yeves et al., 2018) than to *G. phototrophica* AP64^T, which was isolated from a desert lake in North China (**Figure 2B**). The phylogenomic tree of all

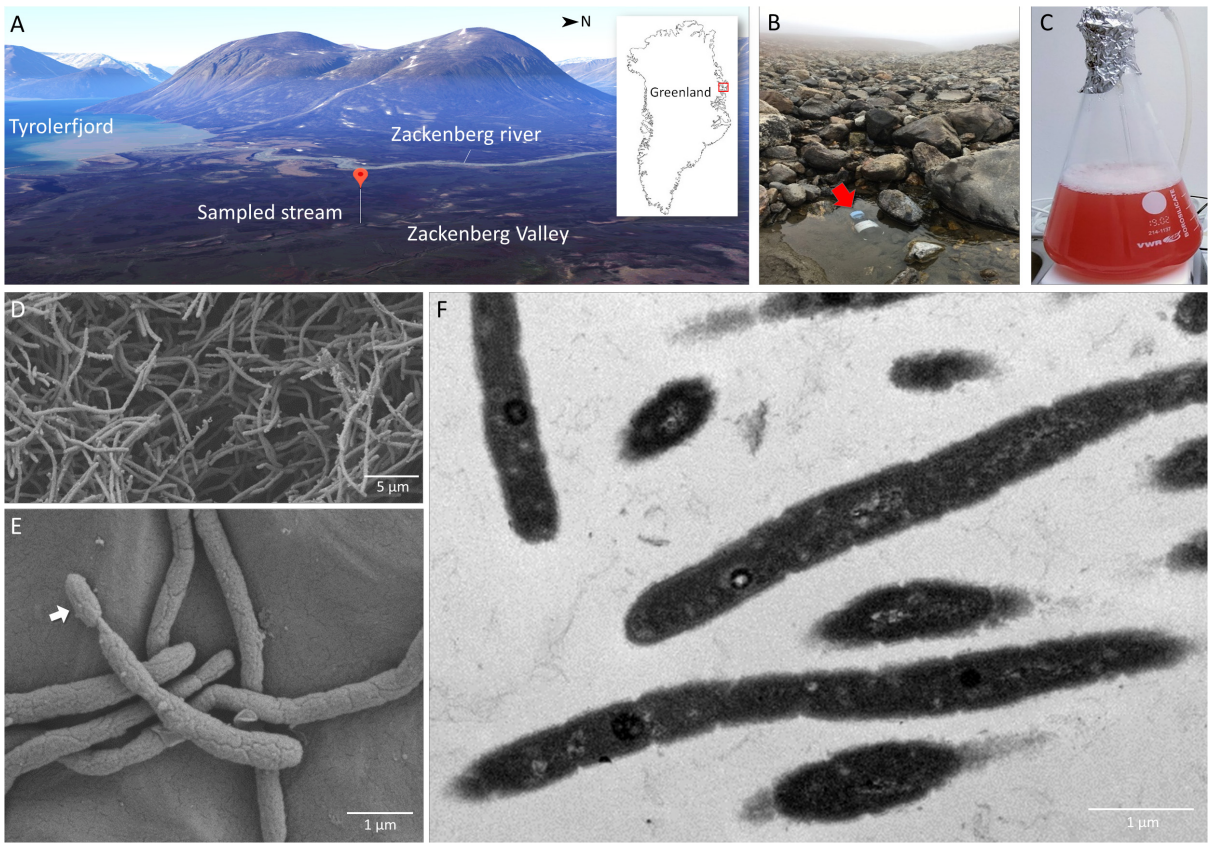


FIGURE 1 | Sampling site at a stream near the Zackenberg Research Station in Northeast Greenland **(A,B)**, 2-week old pure culture of *Gemmatimonas groenlandica* strain TET16^T aerobically grown in a liquid medium **(C)** and microscopic images of the TET16^T cells (**D,E** – scanning electron microscopy; **F** – transmission electron microscopy). The 3D map view of the Zackenberg Valley facing the West was generated from Google Maps.

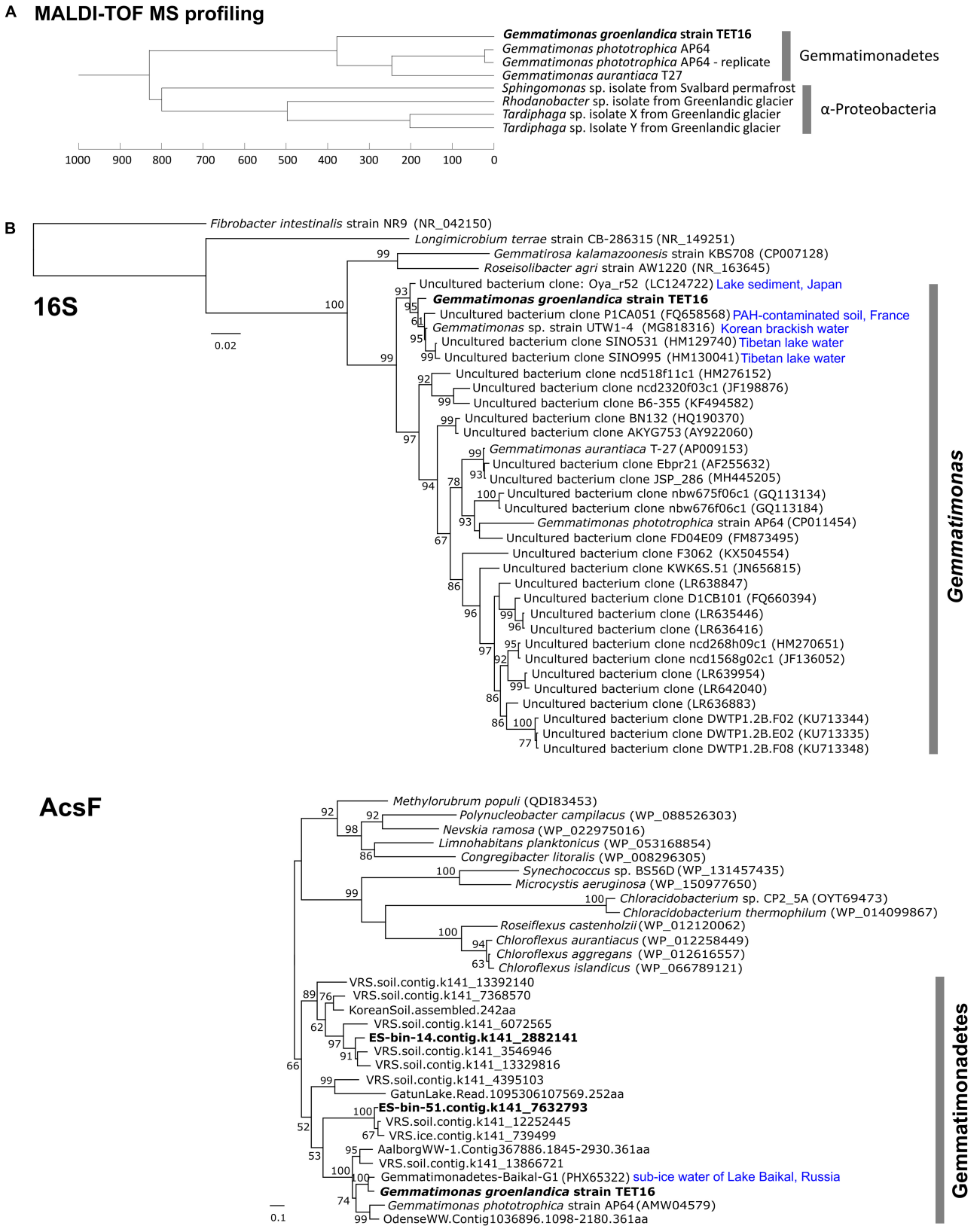


FIGURE 2 | Continued

FIGURE 2 | Dendrogram cluster analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) profiles of the candidate strain TET16^T with reference strains from the phylum Gemmatimonadetes **(A)** and phylogenetic analyses of the 16S rRNA gene and *acsF* gene (encoding aerobic magnesium-protoporphyrin IX monomethyl ester cyclase) of the strain TET16^T **(B)**. MALDI-TOF MS profiles were generated from 2 weeks old colonies on a Bruker's MALDI Biotyper system (see section "Materials and Methods"). A technical replicate of strain AP64^T were performed to assess variations within samples. For phylogenetic analysis, reference sequences were either downloaded from NCBI through BLAST analysis (16S and *acsF*) or retrieved from previous studies (*acsF*). The length cutoffs for reference sequences are 1,375 bp for 16S (i.e., >90% coverage) and 250 residues for *acsF* (i.e., >70% coverage). Type strains and those references closely related to strain TET16^T are highlighted with their source environments shown on the tree. Bars represent nucleotide (16S tree) or amino acid (*acsF* tree) substitution rates. On the *acsF* tree, tBLASTn matches from a Greenlandic glacier and soil metagenomics survey (named as VRS.ice.xxx and VRS.soil.xxx, respectively, details see section "Materials and Methods") were included. The split *acsF* gene from LF-bin-339 was not included.

Gemmatimonadetes-affiliated metagenome-assembled genomes (MAGs) available in the NCBI Genome database and well-characterized isolates also revealed a close relationship between strain TET16^T and *G. aurantiaca* T-27^T and *G. phototrophica* AP64^T, forming a tight cluster on the tree (**Figure 3**).

Strain TET16^T and the two type strains of the same genus, *G. phototrophica* AP64^T and *G. aurantiaca* T-27^T, share 78.4∼78.9% average nucleotide identity (ANI) and 76.9∼78.7% average amino acid identity (AAI) (see **Figure 4A**). The ANI is lower than the threshold of 95∼96% proposed for species delimitation (Kim et al., 2014; Rosselló-Móra and Amann, 2015), while the 16S rRNA gene of strain TET16^T shares 95.7% identity to *G. phototrophica* AP64^T and 95.9% identity to *G. aurantiaca* T-27^T, above the 95% threshold for defining a new genus but below the 98.7% threshold for defining a new species (Chun et al., 2018). Furthermore, the phylogenomic analysis of all existing Gemmatimonadetes isolates and MAGs in the NCBI genome database showed that strains TET16^T, AP64^T, and T-27^T were closely related to each other, forming the tight *Gemmatimonas* cluster (**Figure 3**). The genome-based taxonomic tool GTDB-Tk also classified TET16^T into the *Gemmatimonas* genus with a relative evolutionary divergence (RED) value of 0.934 at the genus level, which was higher than the median RED value of 0.902 calculated from all genera in the database. These lines of evidence, together with the high genome-level synteny between strain TET16^T and the other two *Gemmatimonas* species (**Figure 4B**), supports that strain TET16^T represents a new species within the genus *Gemmatimonas*, named *Gemmatimonas groenlandica*, with its source location of Greenland (Groenland in Danish) designated as the species name.

Slow growth and formation of tiny colonies are common characteristics of the hitherto cultured Gemmatimonadetes bacteria (see summarized characteristics in Pascual et al., 2018). Thus, they can be easily outcompeted by fast growers during initial enrichment from the environment, which largely explains why there are so few pure cultures formally described in the phylum Gemmatimonadetes. Some technological innovations have been introduced to circumvent these issues. For example, the use of a diffusion sandwich system consisting of an array of 384 miniature diffusion chambers and a sample dilution-based high-throughput approach successfully led to the isolation of *Longimicrobium terrae* (Pascual et al., 2016) and *Roseisolibacter agri* (Pascual et al., 2018), respectively. In this study, we show the use of antibiotics is also an effective means to recover Gemmatimonadetes diversity in pure cultures. Although *G. groenlandica* strain TET16^T was initially isolated from an agar plate supplemented with tetracycline, our antibiotics susceptibility test showed that, interestingly, its growth was completely inhibited by tetracycline. Its genome also lacks genes for degrading tetracycline or transporting tetracycline outside the cell. It is likely that during the initial incubation, tetracycline inhibited the growth of bacterial cells surrounding TET16^T cells, and the ensuring depletion of tetracycline by these surrounding cells created a favorable micro-niche for TET16^T to grow in the area where tetracycline was removed through passive diffusion. Given the high light sensitivity of tetracycline, the other possibility is that tetracycline initially inhibited fast growers but over time was photo-degraded, and then the slow growing TET16^T cells started to propagate.

Gemmatimonas groenlandica Is an Aerobic Anoxygenic Phototroph

In contrast to the microaerophilic *G. phototrophica* AP64^T, *G. groenlandica* TET16^T grows well in liquid T21 medium under fully aerobic conditions (reaching stationary phase within 7∼10 days). It can also grow under microaerophilic conditions (∼10% O_2) albeit much more slowly. Fermentative growth was not observed under anaerobic conditions. Growth did not occur under photoautotrophic and chemoautotrophic conditions using sulfide and thiosulfate as electron donors and $NaHCO_3$ as carbon source. Therefore, chemoorganoheterotrophic and photoheterotrophic are the preferred growth modes with the ability to utilize various carbon sources (**Table 1**) under aerobic, light or dark conditions.

At the genome level, phototrophic *G. groenlandica* and *G. phototrophica* appear to be more distantly related than *G. phototrophica* and non-phototrophic *G. aurantiaca* as a large inversion only occurred in *G. groenlandica* (**Figure 4B**), which may suggest a distinct evolutionary history in the local CGB in Greenland. Nonetheless, *G. groenlandica* and *G. phototrophica* show identical organization of photosynthesis-related genes in their PGCs with high DNA sequence identities (70–100%) (**Figure 4C**), implying that they share a common ancestor and, therefore, probably similar structures and properties in their reaction centers. However, surprisingly, the two species show different *in vivo* absorption spectra in the near infra-red range. Two peaks (819 and 866 nm) occurred in *G. phototrophica* corresponding to its double concentric ring of LH complexes (Dachev et al., 2017). However, only one absorption peak (863 nm) appears in *G. groenlandica* (**Figure 5A**). This indicates that the light harvesting system in *G. groenlandica* may have a different structure, as further evidenced by the difference in their major carotenoid composition, where the

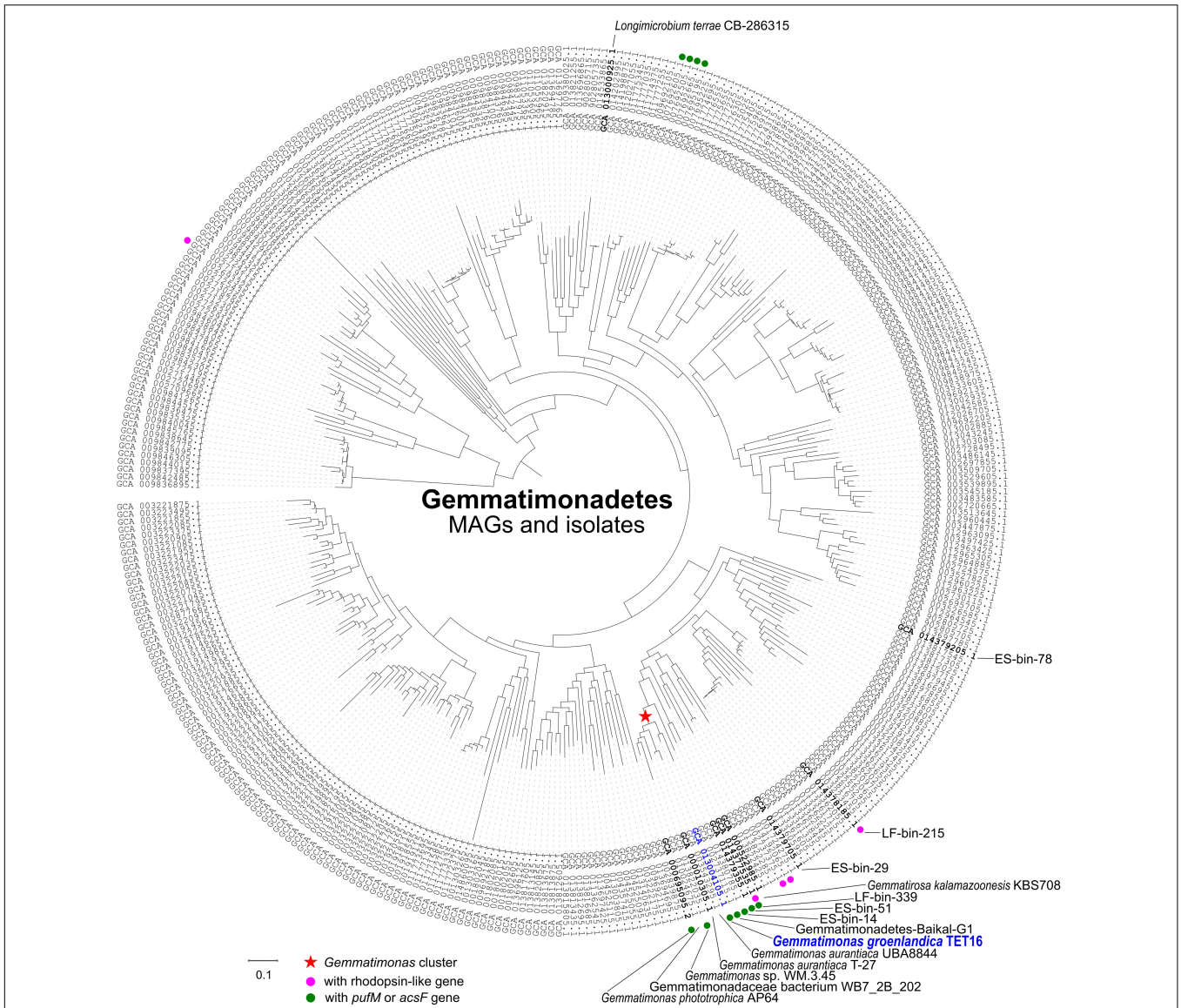


FIGURE 3 | A maximum likelihood phylogenomic tree of the metagenome-assembled genomes (MAGs) and isolates of the phylum Gemmatimonadetes based on aligned 36,076 amino acid positions from 400 most universal marker genes (see section "Materials and Methods"). All 364 Gemmatimonadetes genomes were downloaded from the NCBI microbial genomes portal. The tree was rooted with the genome of *Fibrobacter succinogenes* strain S85 (GenBank assembly no. GCA 000146505.1). The *Gemmatimonas* cluster, type strains, and the MAGs discussed in this study were highlighted on the tree. Bar represents the amino acid substitution rate. The genomes with rhodopsin-like gene or *pufM*/*acsF* gene were marked to show the distribution of phototrophic Gemmatimonadetes.

putative pigment (2*S*,2′*S*)-oscillol 2,2′-di-(α-L-rhamnoside) that dominates in *G. phototrophica* only constitutes a minor fraction in *G. groenlandica* (**Figure 5B**).

In line with the preferred aerobic lifestyle of *G. groenlandica*, its genome only contains the aerobic version of Mg-protoporphyrin IX monomethyl ester oxidative cyclase (*acsF*) and lacks the anaerobic version encoded by the *bchE* gene, whereas both *acsF* and *bchE* genes exist in the genome of *G. phototrophica* (**Table 2**). Similarly, the anaerobic version of coproporphyrinogen oxidase (*hemN*) that is involved in BChl biosynthesis pathway is also absent in *G. groenlandica*. Additionally, none of the denitrification genes that are commonly involved in bacterial anaerobic respiration, including nitrate reductase, nitrite reductase, nitric oxide reductase, and nitrous oxide reductase, were found in its genome. To cope with oxidative stresses associated with the aerobic lifestyle, *G. groenlandica* possesses an expanded gene repository for scavenging reactive oxygen species compared to *G. phototrophica* (**Table 2** and **Figure 6**), including catalase (KatE), superoxide dismutase [Cu-Zn SOD], chloroperoxidase, and organic hydroperoxide resistance gene, as well as bacteriophytochrome for fine regulation of cellular metabolisms in response to light, which is one of the major causes for the generation of radical oxygen species inside cells. The limited arsenal for handling

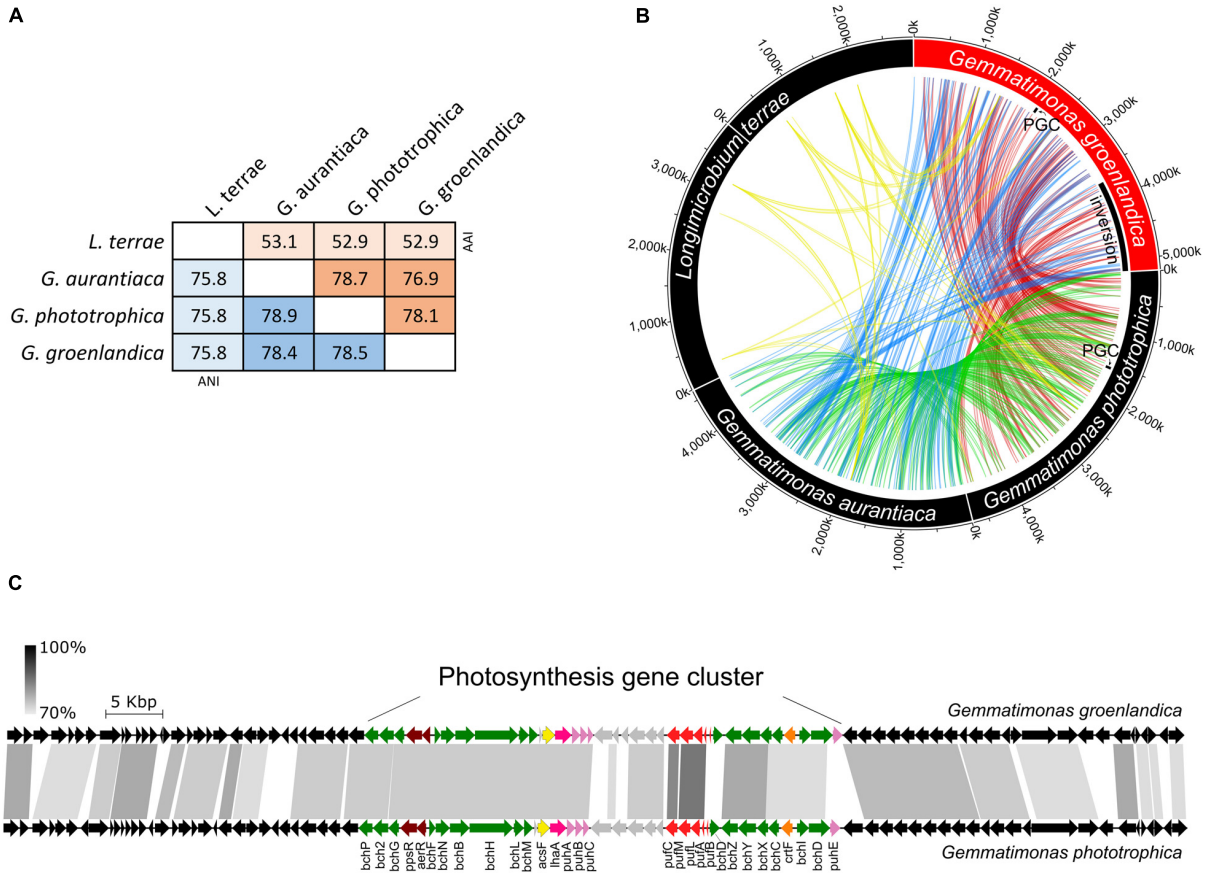


FIGURE 4 | Genome-level sequence similarity between *Gemmatimonas groenlandica* TET16^T and the other three type strains in the phylum Gemmatimonadetes including *Gemmatimonas (G.) phototrophica* AP64^T, *G. aurantiaca* T-27^T, *Longimicrobium (L.) terrae* CB-286315^T. **(A)**, average nucleotide identities (ANI) and average amino acid identities (AAI). **(B)**, genome similarity assessed by connecting highly conserved regions (>80% identity, >2 kb) between each pair. Note that the *L. terrae* genome contains two chromosomes. Any plasmid that may exist was excluded from the analysis. The loci of photosynthesis gene clusters (PGCs) in *G. groenlandica* and *G. phototrophica* are marked. An inversion event of a large genomic region in *G. groenlandica* was identified and marked. **(C)**, Conserved genome synteny of the PGCs (~42 kb) and neighboring regions (~30 kb in each side; see details in **Supplementary Figure 1**) in *G. groenlandica* and *G. phototrophica*. Connecting gray blocks represent matched regions by BLASTn analysis with the cutoffs of >70% sequence identity and >500 bp in length.

oxidative stresses in *G. phototrophica* could explain its sluggish or halted growth when exposed to fully aerobic conditions (Zeng et al., 2015).

Metabolic reconstruction based on genome sequences reveals that *G. groenlandica* is an aerobic anoxygenic phototroph. In addition to the complete PGC and various oxidative stress response genes, it has complete sets of genes for glycolysis and citric acid cycle for generating ATP and NAD(P)H as well as respiratory complexes I-IV for generating a proton gradient across the cytoplasmic membrane (**Figure 6**). It also possesses genes encoding type IV pili and flagella. The major difference between *G. groenlandica* and *G. phototrophica* was the presence of the pentose phosphate pathway in *G. groenlandica*, which could potentially provide it with additional NADPH. *G. phototrophica* appears to be better equipped for carbon storage than *G. groenlandica* by possessing phosphoenolpyruvate synthase for gluconeogenesis. Despite that growth was not observed under low oxygen tension (<1%) conditions in both species, their respiratory complex IVs have both high-affinity (*aa3*-type) and low-affinity (*cbb3*-type) terminal cytochrome *c* oxidase (*cyt c*) (**Figure 6**). The high-affinity *cyt c* is likely used to meet basic cellular needs for energy when facing microaerobic conditions.

Aerobic anoxygenic phototrophs (AAPs) is a functional group of bacteria widely distributed in natural environments, utilizing cyclic photophosphorylation to generate ATP without the need for external electron donors and obtaining organic carbon sources from the environment (Koblížek, 2015). Distinct from their purple photosynthetic relatives, AAPs are unique in their capacity to synthesize BChl *a* exclusively under aerobic conditions. All members of AAPs identified to date belong to the phylum Proteobacteria. The discovery of the fully aerobic phototroph of *G. groenlandica* expands AAPs further into the phylum Gemmatimonadetes. The structure and physical properties of the reaction centers in *G. groenlandica* and its photophysiology warrant further investigations, for instance, how light contributes to the growth of TET16^T and how light influences its carbon metabolisms.

TABLE 1 | Comparison of phenotypic, physiological and genomic characteristics of the two phototrophic members in the phylum Gemmatimonadetes, *Gemmatimonas groenlandica* strain TET16^T and *Gemmatimonas phototrophica* strain AP64^T.

	***G. groenlandica* TET16^T**	***G. phototrophica* AP64^T**
Cell		
shape	short to long rod	short to long rod
color	pink to red	red
capsule-like structure	yes	yes
fission mode	binary, budding	binary, budding
Genome		
GC%	65.1%	64.4%
size (Mbp)	5.179	4.717
plasmid	n. d.	n. d.
Growth		
temp. range	15–32°C	20–30°C
pH range	6.5–9.0	6.0–9.0
pH optimum	7.3	7.5–8.0
oxygen requirement	fully aerobic, 10% O_2 tolerant	obligate microaerophilic
carbon source	yeast extract, saccharin, salicin, adonitol, trehalose, dulcitol, rhamnose, pyruvate, glucose	yeast extract
non-utilized carbon source	xylose, ribose, erythritol, turanose, cellobiose, melibiose, lyxose, arabinose	casamino acids, sodium succinate, sodium acetate, sodium pyruvate, potato starch, sucrose, L-glutamic acid, L-leucine, L-arginine, L-alanine, L-isoleucine, L-arabinose, D-sorbitol, D-mannitol
resistant to antibiotics	bacitracin, chloramphenicol, nystatin	ampicillin, penicillin, paramycin sulfate, polymyxin B sulfate, nystatin
susceptible to antibiotics	neomycin, amoxicillin, tetracycline, amphotericin B	neomycin, vancomycin, bacitracin, gentamicin
in liquid media	yes	not observed
in prolonged darkness	yes*	yes
Chemotaxonomy**		
fatty acid	C15:0 iso, C15:1 ω6c	C16:1, C14:1, C18:1 ω9c
polar lipid	Phosphatidylethanolamine, aminolipid, diphosphatidylglycerol	–
quinone	MK-8, MK-9	MK-8

*Data from strain AP64^T were retrieved from previous studies (Zeng et al., 2014, 2015). n. d., not detected. *observed for a period of 3 weeks. **only major compositions are reported.*

Unique but Conserved Photosynthesis Gene Cluster in CGB

The discovery of the second CGB member *G. groenlandica* enabled us to identify the common and unique features in the photosynthesis gene clusters of CGB (**Figure 7A**). The PGCs of *G. groenlandica* and *G. phototrophica* are identical in terms of gene content and organization and sub-cluster orientation, including the hypothetical genes located between the *puh* (reaction center assembly proteins) and *puf* (reaction center proteins) operons (**Figure 4C**). Similar patterns in operon organization were observed in some incomplete Gemmatimonadetes PGCs reconstructed from active sludge metagenomes (Zeng et al., 2016) and a Lake Baikal's surface water metagenome (Cabello-Yeves et al., 2018). The PGC of *G. groenlandica* appear to be more closely related to MAG Gemmatimonadetes-Baikal-G1 (Cabello-Yeves et al., 2018) since their PGCs share a higher number of genes that are of >90% amino acid identity (**Figure 7B**), including *acsF*, *bchCDHILXYZ*, *crtF*, and *pufLMC*. In contrast, only *bchL* and *pufLMA* genes between *G. groenlandica* and *G. phototrophica* and *pufL* gene between *G. phototrophica* and the Baikal MAG are >90% identical. Together, this evidence indicates that CGB share a conserved PGC likely originating from a common ancestor.

Gemmatimonadetes PGCs (Gemma-PGCs) contain the same gene sub-clusters as Proteobacteria, including *bchP2G*, *bchFNBHLM*, *lhaA-puhABC*, *pufBALMC*, *crtF-bchCXYZ*, and *bchID*, but they differs in their orientations (**Figure 7C**). No proteobacterial PGC was found to show completely identical orientations of each sub-cluster and the same relative positions of these six sub-clusters to those in Gemma-PGC. The relative location of *acsF* and *bchO* genes are also unique in Gemma-PGC with *acsF* consistently located between the *bchFNBHLM* and *lhaA-puhABC* sub-clusters and *bchO* between *pufBALMC* and *bchCXYZ*. As environmental metagenomics data are exploding, these unique features could serve as convenient markers for identifying Gemma-PGC from metagenomic contigs.

Aerobic CGB Also Exist in Northeast Greenland's Soil and Glacier

Quantification of aerobic CGB in environmental samples remains a challenge. Given the high synteny and high sequence identities of the PGCs of *G. groenlandica* and *G. phototrophica*, it

TABLE 2 | Key differences in the presence of genes related to bacteriochlorophyll (BChl) biosynthesis (only including three enzymes that have both aerobic and anaerobic versions) and genes related to oxidative stress response between *Gemmatimonas phototrophica* and *Gemmatimonas groenlandica*.

Functional category		Gene	Presence and GenBank access. no.	
Enzyme	Function		*G. phototrophica*	*G. groenlandica*
BChl biosynthesis pathway				
Coproporphyrinogen oxidase	Coproporphyrinogen III → Protoporphyrinogen IX	*hemF* (aerobic) *hemN* (anaerobic)	WP_075071451 WP_075071577	QJR35625 *n. d.*
Protoporphyrinogen oxidase	Protoporphyrinogen IX → Protoporphyrin IX	*hemJ** (aerobic) *hemG* (anaerobic)	WP_026851012 WP_053334334	QJR36952 QJR35626
Mg-protoporphyrin IX monomethyl ester oxidative cyclase	Mg-protoporphyrin IX → Divinylprotochlorophyllide	*acsF* (aerobic) *bchE* (anaerobic)	WP_082821546 WP_026851009	QJR35611 *n. d.*
Oxidative stress response				
Bacteriophytochrome, two-component sensor histidine kinase	Light sensing, regulation	*BphP*	*n. d.*	QJR35953 QJR37607
Bacteriophytochrome, heme oxygenase	Biosynthesis of the biliverdin chromophore	*BphO*	*n. d.*	QJR37606
Superoxide dismutase [Cu-Zn]	reactive oxygen species scavenging	CuZnSOD	*n. d.*	QJR38173
Superoxide dismutase [Mn]	reactive oxygen species scavenging	MnSOD	WP_169804576	QJR34760

A complete set of shared and unique genes is presented in ***Supplementary Table 1****. n.d., gene homolog not detected. *previously annotated as hemY, revised as hemJ according to Kato et al. (2010).*

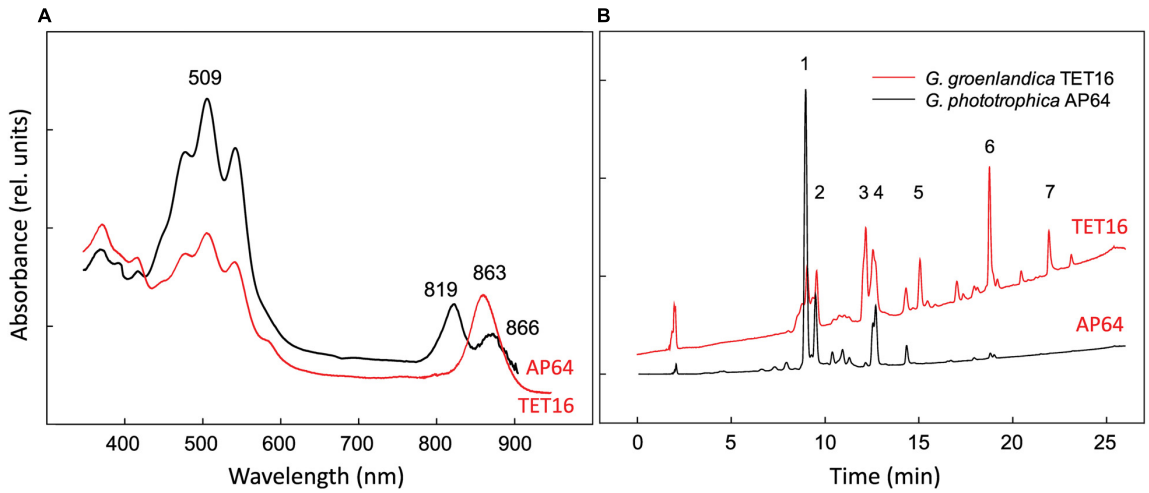
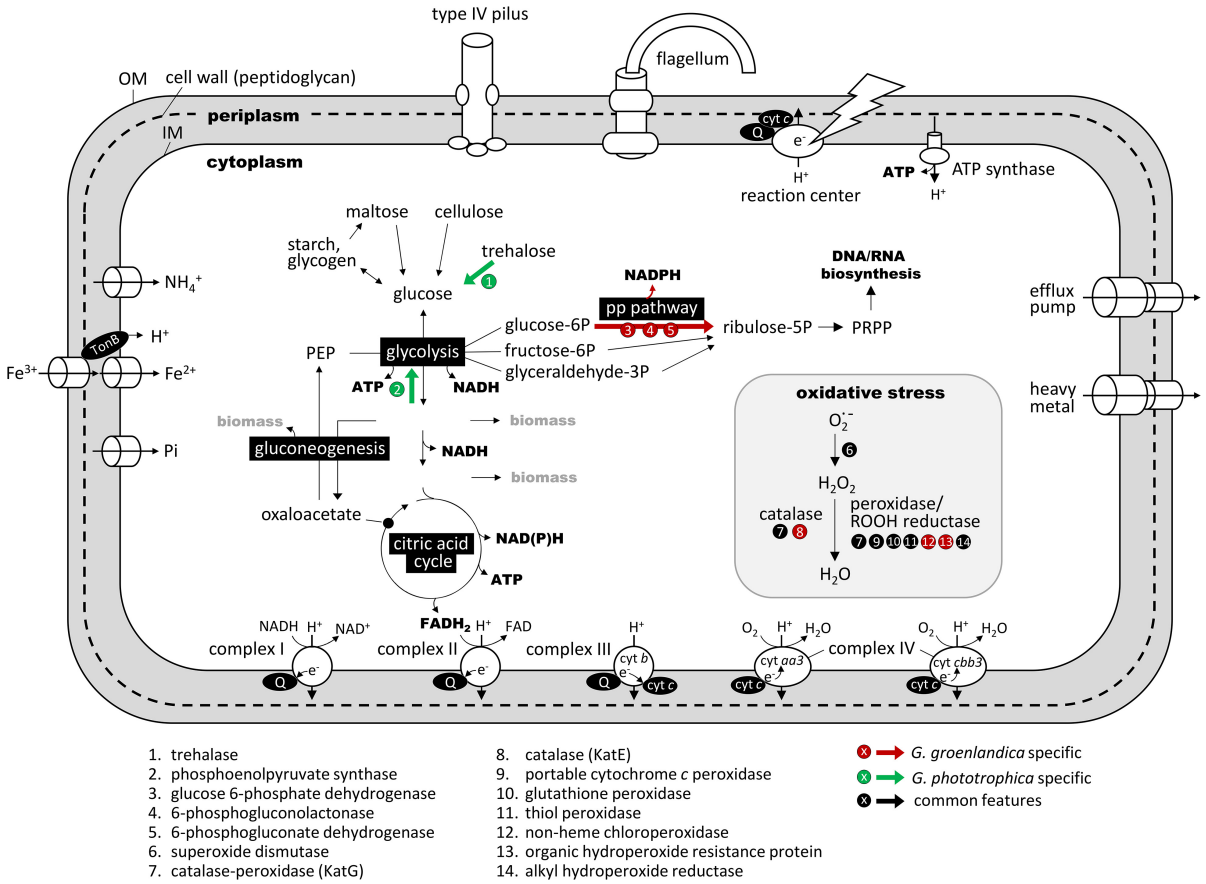


FIGURE 5 | Comparison of *in vivo* absorption spectra **(A)** and HPLC elution profiles of pigments recorded at 490 nm **(B)** of *Gemmatimonas groenlandica* TET16^T and *Gemmatimonas phototrophica* AP64^T. Identified peaks: 1 and 2, putative (2S,2′S)-oscillol 2,2′-di-α-L-rhamnoside; 3: unknown carotenoid, λ_{max} = 317,(470),495,527 nm; 4: unknown keto-carotenoid, λ_{max} = 499 nm; 5–6: spirilloxanthin-like carotenoids, λ_{max} = 315,(470),493,526 nm; 7: unknown carotenoids, λ_{max} = 470,500 nm.

is practically impossible to distinguish between aerobic CGB represented by *G. groenlandica* and microaerophilic CGB like *G. phototrophica* using conventional approaches that rely on biomarker gene phylogenies. Instead, based on our comparison of anaerobic BChl biosynthetic genes and oxidative stress response genes in these two species (see above), we propose that the lack of *bchE* and presence of *BphP* and *BphO* could be strong indicators of a query genome belonging to aerobic CGB.

We applied these criteria and examined the six Gemmatimonadetes MAGs previously assembled from 460G-base shotgun reads of a surface soil sample (with prefix ES-bin) and a glacial ice sample (with prefix LF-bin) near the Villum Research Station (81°36′ N, 16°40′ W) in Northeast Greenland (Zeng et al., 2020). Three bins (ES-bin-14, ES-bin-51, and LF-bin-339) were found to contain at least one *puf* gene, one *puh* gene, and one *bch* gene (**Table 3**), indicating they belong to CGB. All the three MAGs possess *acsF*, which is, however, split into two fragments in LF-bin-339, whereas all lack the *bchE* gene. ES-bin-51 and LF-bin-339 contain both *BphP* and *BphO*. Despite the small dataset and the incompleteness of the MAGs (**Table 3**),



1. trehalase
2. phosphoenolpyruvate synthase
3. glucose 6-phosphate dehydrogenase
4. 6-phosphogluconolactonase
5. 6-phosphogluconate dehydrogenase
6. superoxide dismutase
7. catalase-peroxidase (KatG)
8. catalase (KatE)
9. portable cytochrome *c* peroxidase
10. glutathione peroxidase
11. thiol peroxidase
12. non-heme chloroperoxidase
13. organic hydroperoxide resistance protein
14. alkyl hydroperoxide reductase

G. groenlandica specific
G. phototrophica specific
common features

FIGURE 6 | Comparison of key metabolic pathways and functions related to the aerobic photoheterotrophic lifestyle in *Gemmatimonas groenlandica* and *Gemmatimonas phototrophica* based on genome annotations. The comparison is based on the "KEGG metabolic pathway reconstruction" function of the RAST (Rapid Annotations using Subsystems Technology) web server (https://rast.nmpdr.org/; Aziz et al., 2008). A complete set of shared and unique genes between these two genomes is shown in **Supplementary Table 1**. OM, outer membrane; IM, cytoplasmic membrane; PRPP, phosphoribosyl diphosphate; Q, quinone pool; *cyt*, terminal cytochrome oxidase; PEP, phosphoenolpyruvate; PRPP, phosphoribosylpyrophosphate; PP pathway, pentose phosphate pathway. Red arrow indicates the pathway that only occurs in *Gemmatimonas groenlandica* and green arrows are those only present in *Gemmatimonas phototrophica*. Pathways and functions drawn in black and white are predicted in both genomes. Note that the respiratory complex IV has two types of cytochrome *c* as terminal electron acceptor. The *aa3*-type has a low affinity for O_2 and the *cbb3*-type is a high-affinity version.

these lines of evidence may suggest that aerobic CGB are more prevalent than microaerophilic CGB in the environment, consistent with the hypothesis that aerobic CGB can generate energy more efficiently and thus may cope more effectively with cold and nutrient stresses in supraglacial environments.

On the AcsF phylogenetic tree (**Figure 2B**), the two aerobic CGB MAGs, ES-bin-14, and ES-bin-51, were placed on the branches that are distinct from *G. groenlandica* and *G. phototrophica*. However, on the phylogenomic tree (**Figure 3**), MAGs ES-bin-14, ES-bin-51 , and LF-bin-339 are closely related to the members of the *Gemmatimonas* cluster, likely representing novel CGB species in the same genus or family. Four MAGs (GenBank assemblies GCA_007692605.1, GCA_007692505.1, GCA_007692665.1, and GCA_007695195.1) from a high-altitude alkaline salt lake in the Cariboo Plateau in Canada (Zorz et al., 2019) were also found to contain the *pufM* or *acsF* gene and therefore they belong to CGB, albeit distantly related to the *Gemmatimonas* cluster (**Figure 3**). Intriguingly, MAGs LF-bin-215 and LF-bin-339 also contain a rhodopsin (Rho)-like gene. A proteorhodopsin-like gene is also present in the Gemmatimonadetes MAG Baikal-G1 (non-CGB) assembled from a Lake Baikal metagenome (Cabello-Yeves et al., 2018), potentially providing an additional energy source. However, their function as proton-pump rhodopsins has not yet been verified. Given the low relative abundance of Gemmatimonadetes in the environment, of which only a minor fraction are phototrophic, the ecological function of phototrophic Gemmatimonadetes is likely minor and it is more probable that they only serve as part of a rare microbial biosphere, providing ecosystems with persistent microbial seeds, functional diversity, and ecological resilience (Lynch and Neufeld, 2015; Jousset et al., 2017).

Interestingly, a Gemmatimonadetes MAG (named CSSed162cmB_429) recently assembled from a hypersaline soda lake sediment metagenome (top layer, 0–2 cm) was found to contain genes coding for type-2 RC, type I RubisCO (*rbcLM*), and phosphoribulokinase (*prkB*) (Vavourakis et al., 2019),

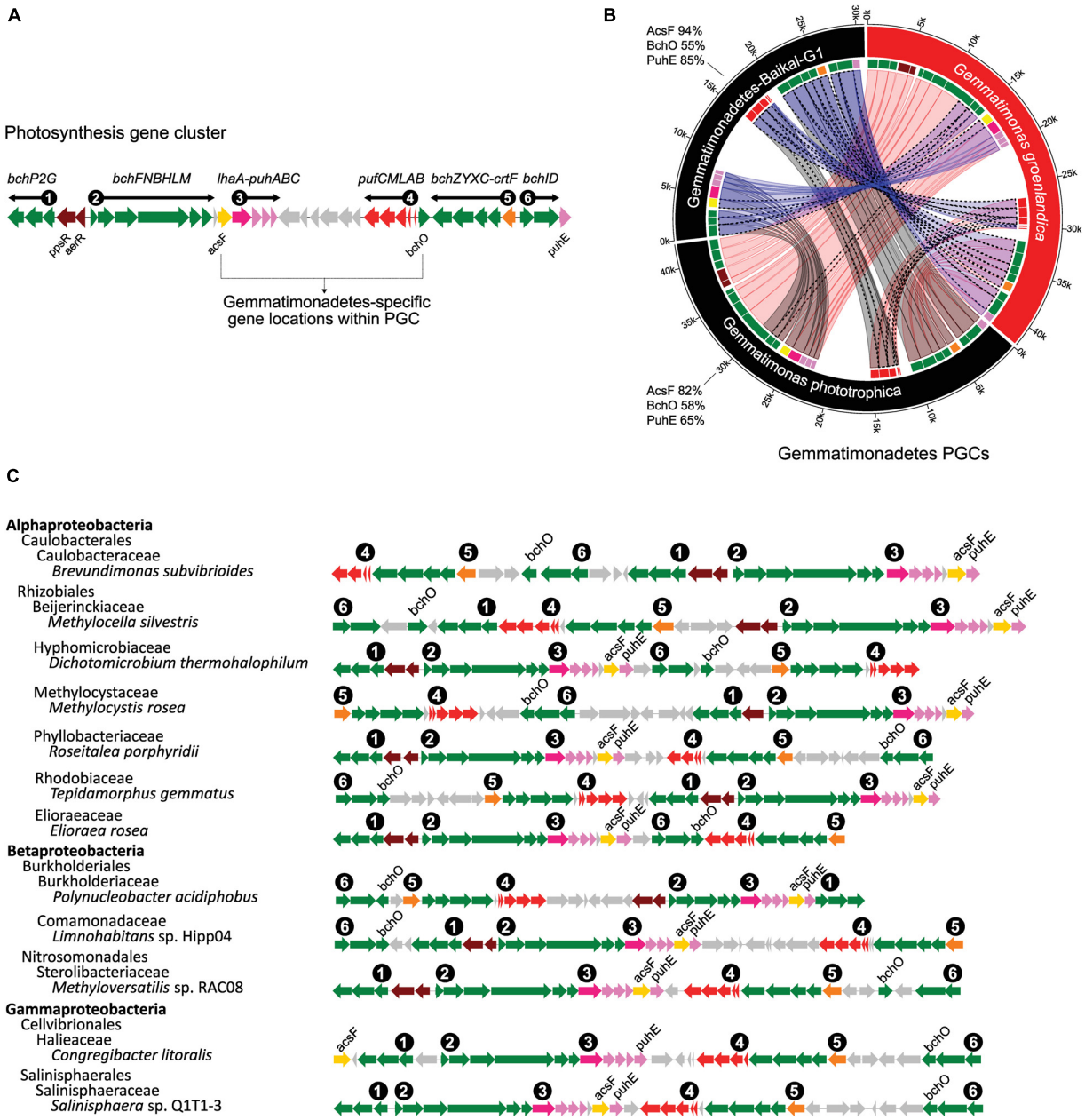


FIGURE 7 | Conserved and unique gene organization of photosynthesis gene cluster (PGC) in chlorophototrophic Gemmatimonadetes bacteria. **(A)**, six gene sub-clusters commonly found in photosynthesis gene clusters. *bch* and *acsF*, bacteriochlorophyll biosynthesis genes; *puf*, genes encoding reaction center proteins; *puh*, genes encoding reaction center assembly proteins; *crt*, carotenoid biosynthesis genes; *lhaA*, light-harvesting complex I assembly protein; *ppsR* and *aerR*, regulation related genes. Gray-colored genes are hypothetical ORFs with unknown function. The unique locations of *acsF* and *bchO* are highlighted. **(B)**, comparison of three known Gemmatimonadetes PGCs including two isolates (*G. phototrophica* and *G. groenlandica*) and a MAG, Gemmatimonadetes-Baikal-G1, reconstructed from a metagenome of the Baikal Lake in Russia (Cabello-Yeves et al., 2018). The same genes are connected by ribbons. A dotted outline of a ribbon represents >90% protein sequence identity for the gene pair. The protein sequence identities of AcsF, BchO and PuhE in reference to *G. groenlandica* are shown next to the species arc. The fragmented Gemma-PGCs in the MAGs of this study were not included. **(C)** Representative PGC architecture in various proteobacterial genomes supporting the uniqueness of PGC in Gemmatimonadetes. Genomes were top tBlastn hits against NCBI's RefSeq genome database using fused AcsF-BchO-PuhE protein sequences (see section "Materials and Methods"). One genome from each group at the class level was downloaded from NCBI. Only the top scoring genome from each class was kept as a representative.

indicating a photoautotrophic potential in this CGB member. *G. phototrophica* also contains a RubisCO-like gene (GenBank accession no. WP_026848175) but lacks the *prkB* gene, and no RubisCO homolog was found in *G. groenlandica*. Phylogenetic analysis showed that the RubisCO-like gene of *G. phototrophica* groups into the IV-Photo cluster (Tabita et al., 2007) with all

TABLE 3 | Metagenome-assembled genomes (MAGs) of Gemmatimonadetes origin from the "Lille Firn" glacier soil and ice metagenomes in Northeast Greenland.

MAG	GenBank assembly accession ID	# contigs	Genome size (bp)	GC %	Compl %	Cont %	Key functional genes											
							puf	puh	bch	acsF	bchE	hemF	hemN	hemJ	hemG	BphP	BphO	RubisCO	Rho-like
ES-bin-14	GCA_014380785.1	633	5,349,535	60.2	79.18	6.59	+	+	+	+	–	+	–	–	–	–	–	–	–
ES-bin-29	GCA_014379705.1	758	5,043,933	67.5	67.91	1.1	–	+	–	–	–	–	–	–	–	–	–	+	–
ES-bin-51	GCA_014379355.1	293	4,821,585	65.3	83.65	2.39	+	+	+	+	–	+	–	–	–	+	+	–	–
ES-bin-78	GCA_014379205.1	339	4,113,156	60.3	89.05	7.69	–	–	–	–	–	–	–	–	–	–	–	–	–
LF-bin-215	GCA_014378185.1	808	3,680,986	65.7	73.65	4.4	–	–	–	–	–	–	–	–	–	–	–	–	+
LF-bin-339	GCA_014377535.1	254	3,787,951	65.7	90.61	3.30	+	+	+	+	–	+	+	–	+	+	+	–	+

Statistics data were retrieved from a previous study by Zeng et al. (2020). The presence of selected key functional genes was investigated in this study by tBLASTn search in each MAGs using corresponding G. groenlandica genes as queries. The BLAST results that pass the threshold of E < e-5, coverage >40%, and identity >40%) were regarded as a positive match. +, presence of the gene or at least one gene in the category of puf (reaction center-related proteins), puh (reaction center assembly proteins) or bch (bacteriochlorophyll biosynthesis gene). acsF/bchE, Mg-protoporphyrin IX monomethyl ester oxidative cyclase (aerobic/anaerobic); hemF/hemN, coproporphyrinogen oxidase (aerobic/anaerobic); hemJ/hemG, protoporphyrinogen oxidase (aerobic/anaerobic); BphP and BphO, bacteriophytochrome; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; Rho-like, rhodopsin-like. Compl, completeness; Cont, contamination.

members coming from phototrophic Proteobacteria or Chlorobi (**Figure 8**). It is unclear whether the last common ancestor of the phylum Gemmatimonadetes was a photoautotroph or the photoautotrophic capacity evolved later by HGT. However, this opens a possibility that the evolution of the phylum Gemmatimonadetes might resemble that of the phylum Proteobacteria, where all members were supposed to have originated from a photoautotrophic purple bacterium and the photosynthetic capacity has been lost many times, resulting in various non-photosynthetic lineages (Woese, 1987; Battistuzzi et al., 2004). Over the evolutionary course, *G. groenlandica* may represent aerobic CGB that have adapted to modern fully oxygenated surface environments, whereas *G. phototrophica* represents a more primitive species undergoing evolutionary transitioning from anoxic to oxic environments. More complete genomes from various lineages of CGB are needed to decipher the evolutionary puzzle of phototrophy in this phylum.

CONCLUDING REMARKS

Since our accidental discovery of *G. phototrophica*, the first chlorophototrophic member of the phylum Gemmatimonadetes (Zeng et al., 2014), molecular data accumulated to date suggest CGB are widespread in the environment (Zeng et al., 2016; Cabello-Yeves et al., 2018; Vavourakis et al., 2019). We circumvented the slow growth nature of CGB by employing a target screening strategy that led to the successful isolation of the second member of CGB as a pure culture. The use of antibiotics during the initial enrichment and the selection of a cold low-biomass environment in Greenland for the cultivation attempt appear to be the key to our success. Our strategy also demonstrates the power of combining MALDI-TOF MS and colony infrared imaging techniques in discovering novel chlorophototrophs from nature. The new CGB member of *G. groenlandica* sp. nov. provides an additional model microorganism as a strictly aerobic anoxygenic phototroph in this phylum that readily grows in liquid medium. This trait is not seen in the microaerophilic slow grower *G. phototrophica*, and may prove rather important for future genetic engineering and detailed photophysiological studies.

TAXONOMY

Gemmatimonas groenlandica sp. nov. [groen.lan'di.ca. Gr. n. pertaining to the isolation source of Greenland (Groenland in Danish)] is a bacteriochlorophyll *a*-containing bacterium isolated from the stream water in Northeast Greenland. Cells are short to long rods, contain capsule-like structures, and divide in a binary fission mode with budding occasionally observed (**Figure 1E**). The colonies display a pink-to-red color and cultures turn reddish in stationary phase under fully aerobic conditions. The temperature range for growth is between 15 and 32°C, with optimum at 20–25°C and growth occurs at pH between 6.5 and 9.0 with an optimum at pH 7.3. Cells appear intolerant to NaCl as growth was inhibited even in the

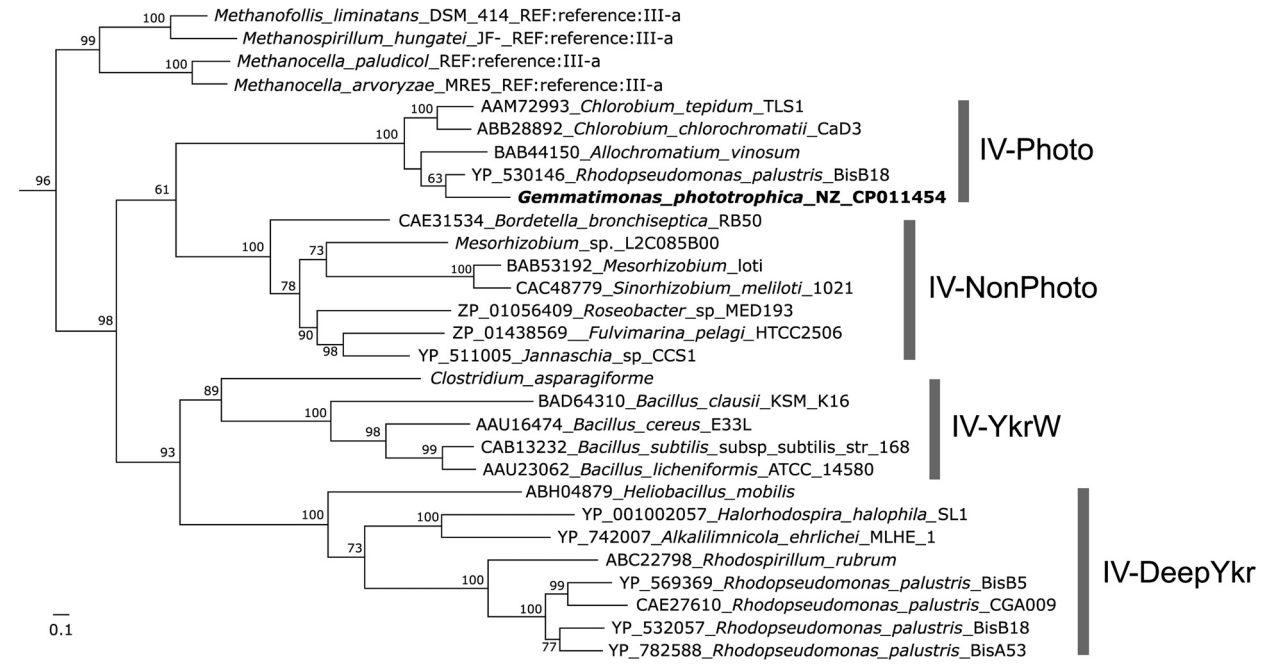


FIGURE 8 | Maximum-likelihood tree of RubisCO-like protein (RLP) gene (GenBank accession no. WP_026848175) in *Gemmatimonas phototrophica*. RubisCO sequences collected and classified by Jaffe et al. (2019) were used as references. Bar represents amino acid residue substitution rate. Bootstrap values between 0.5 and 1 are shown on the branches.

presence of 0.1% NaCl. It prefers growth in aerobic conditions but can also grow slowly under microaerophilic conditions (10% O_2). Fermentative growth was not observed under anaerobic conditions. Growth did not occur under photoautotrophic and chemoautotrophic conditions using sulfide and thiosulfate as electron donor and $NaHCO_3$ as the sole carbon source. Chemoorganoheterotrophic and photoheterotrophic growth modes are preferred and various carbon sources are utilized under aerobic, light or dark conditions. Cells are resistant to bacitracin, chloramphenicol, and nystatin but susceptible to neomycin, amoxicillin, tetracycline, and amphotericin B. The substrates utilized as carbon source/electron donor under photo- or chemoheterotrophic condition include saccharin, salicin, adonitol, trehalose, dulcitol, rhamnose, pyruvate, glucose and yeast extract, but not xylose, ribose, erythritol, turanose, cellobiose, melibiose, lyxose, and arabinose. Only yeast extract (0.5 g L^{-1}) can be utilized as nitrogen source, but not nitrite, nitrate, glutamine, ammonium ion, and casamino acids. Addition of vitamins is not necessary for growth. The dominant fatty acids are C15:0 iso and C15:1 ω6c (**Supplementary Figure 2**) and the dominant polar lipids are phosphatidylethanolamine, aminolipid and diphosphatidylglycerol (**Supplementary Figure 3**). The major respiratory quinones are MK-8 and MK-9 (**Supplementary Figure 4**). The genomic GC content is 65.1% and the genome size is 5,179,092 bp. The type strain, TET16^T (= DSM 110279^T and CGMCC 1.18661^T), was isolated from the surface water of a stream in the Zackenberg Valley in Northeast Greenland.

MATERIALS AND METHODS

Sampling and Screening of Isolates by MALDI-TOF MS

Sampling was done in late August 2017 at a small branch of a stream (74°28′12.5″N, 20°31′02.0″W) close to the Zackenberg Research Station (74°28′10.4″N, 20°34′30.1″W) in Northeast Greenland. This area has a mean annual air temperature of −9.2°C and an annual precipitation of 203 mm with August being the warmest month (mean 5.1°C) (Hasholt and Hagedorn, 2000). The stream is part of the tributaries of the Zackenberg River with very low concentrations of dissolved organic matter (0.6 mg C L^{-1}) and soluble reactive phosphate (8.4 μg P L^{-1}) (Pastor et al., 2019). Surface water was sampled and filled into a 50 mL bottle which was kept under 4°C until transport to the laboratory in Denmark about 1 month later. The water sample was diluted 1:10 and then 100 μL of dilution was plated onto a 1/5 strength R2A agar plate (Difco). To increase the diversity of cultured bacteria, the following antibiotics were used individually: 8 μg/mL tetracycline (TET), 20 mg L^{-1} piperacillin sodium salt, 20 mL L^{-1} streptomycin, 8 mg L^{-1} gentamicin, and 20 mg L^{-1} kanamycin. For agar plates supplemented with antibiotics, 100 μL of the original water sample was plated.

Agar plates were incubated under room temperature and normal laboratory indoor light condition for two to 10 weeks until colonies formed.

To screen for potential phototrophic Gemmatimonadetes bacteria, a two-step strategy was adopted to increase the chance of success with manageable labor efforts: (1), only small (slow growers), pinkish or reddish colonies were considered for MALDI-TOF MS screening based on previous knowledge accumulated from *Gemmatimonas phototrophica* strain AP64, which was isolated from a desert lake (Zeng et al., 2015); (2), the MALDI-TOF MS fingerprints of *G. phototrophica* and *G. aurantiaca* were used as references and only colonies that formed a tight cluster with these two references in the MALDI-TOF MS fingerprinting analysis were considered for further verification as phototrophic Gemmatimonadetes by genome sequencing and measurement of absorption spectra.

All target colonies were subjected to MALDI-TOF MS fingerprinting analysis using the Microflex LT system (Bruker Daltonics, Bremen, Germany) following the procedure described previously (Zervas et al., 2019). Briefly, a toothpick was used to transfer a small amount of a test colony onto the target plate (MSP 96 polished steel, Bruker), which was evenly spread out and formed a thin layer of biomass on the steel plate. The sample was then overlaid with 70% formic acid and allowed for air dry before the addition of 1 μL MALDI-MS matrix solution (α-cyano-4-hydroxycinnamic acid, Sigma-Aldrich). The standard method "MBT_AutoX" was applied to obtain proteome profiles within the mass range of 2 – 20 kDa using the flexControl software (Bruker). The flexAnalysis software (Bruker) was used to smooth the data plot, subtract the baseline and generate main spectra (MSP), followed by a hierarchical clustering analysis using the MALDI Biotyper Compass Explorer software, which produced a dendrogram as output for visual inspection of similarities between samples. For defining different groups at strain/species level, an empirical distance cutoff of 50 was used. There was no consensus on the cutoff at genus or above levels, which varies greatly among different bacterial groups. We used lab-maintained cultures of Proteobacteria as negative controls.

Genome Sequencing, Phylogeny, and Comparative Genomics

Genomic DNA of the selected isolate TET16^T (from a TET-supplemented agar plate) was extracted from cells harvested from 1/5 R2A agar plates after 2–3 weeks growth using the EasyPure bacterial genomic DNA kit (TransGen Biotech, Beijing, China) and was sequenced both on an Illumina NextSeq 500 platform in house and on a PacBio Sequel platform at BGI Hong Kong using a 20K library method for SMART cell with the manufacture's standard protocols. A total of 3,097,562 Illumina reads (PE 150) and 366,471 PacBio reads were generated. For quality control, the Illumina reads were trimmed at the left end for 10 bases due to irregularities in GC content and at the right end for 30 bases to remove adaptors and irregular bases, while all PacBio reads were used for following *de novo* hybrid-assembly. The gap-free complete genome was assembled using Unicycler (ver. 0.4.8) in a hybrid mode with default settings (Wick et al., 2017). The genome of strain TET16^T was annotated with the NCBI's prokaryotic genome annotation pipeline (GenBank accession no. CP053085).

The 16S rRNA gene and *acsF* gene were retrieved from the TET16 genome and aligned with reference sequences. The 16S rRNA gene reference sequences were downloaded from the NCBI *nr* database through BLASTn search (>97% identities, >1,375 bp, equivalent to >90% coverage). The reference sequences for *acsF* include the sequences downloaded from the NCBI *nr* database through tBLASTn search (>250 amino acids, equivalent to >70% coverage) and those used in two previous studies (Zeng et al., 2016, 2020). For phylogeny inference, sequences were first aligned with MAFFT v7.471 (Katoh and Standley, 2013) using the Q-INS-i algorithm for 16S rRNA genes, which takes secondary structure information of RNA into account and the G-INS-I algorithm for AcsF protein sequences. The most appropriate evolutionary model was determined using ModelTest-NG (Darriba et al., 2020). Then, the phylogenetic tree was built with RAxML-NG (Kozlov et al., 2019) using the nucleotide model GTR + I + G4 for 16S rRNA genes and the amino acid model LG + I + G4 for AcsF protein sequences both with 1,000 bootstrap replicates. The tree was visualized in the Geneious Prime environment (Biomatters, New Zealand).

To demonstrate the ancient connection of phototrophic Gemmatimonadetes bacteria to purple phototrophic bacteria, phylogenetic analysis of the RubisCO-like gene identified in the complete genome of *Gemmatimonas phototrophica* AP64^T (GenBank accession no. CP011454; Zeng et al., 2016) but not described before (Zeng et al., 2014) were also performed in this study. The well-classified RubisCO reference sequences were retrieved from the study by Jaffe et al. (2019). Multiple sequence alignment and tree inference were conducted using the same method as described above for the AcsF phylogeny.

The average nucleotide identity (ANI) and average amino acid identity (AAI) between *Gemmatimonas groenlandica* TET16^T (this study) and other three type strains in the phylum Gemmatimonadetes that have genome sequences publicly available, including *G. phototrophica* (Zeng et al., 2015), *G. aurantiaca* (Zhang et al., 2003) and *Longimicrobium terrae* (Pascual et al., 2016), were calculated using FastANI (ver. 1.3[2]; Jain et al., 2018) and CompareM (ver. 0.1.1[3]), respectively. The two plasmids in the genome of *Longimicrobium terrae* strain CB-286315 were removed prior to the calculation. Whole genome-level synteny of these four genomes were also calculated using the Easyfig program (ver. 2.2.3; Sullivan et al., 2011) and a circos plot was created with Circa[4]. The synteny of PGCs (∼42 kb) and flanking regions (∼ 30 kb) in *G. groenlandica* and *G. phototrophica* was calculated and visualized using Easyfig.

The unique gene organization features in the PGC of chlorophototrophic Gemmatimonadetes bacteria was evaluated by comparing the two complete PGCs of *G. groenlandica* and *G. phototrophica* and an incomplete but continuous PGC from a Gemmatimonadetes MAG reconstructed from the metagenome

[2]https://github.com/ParBLiSS/FastANI

[3]https://github.com/dparks1134/CompareM

[4]http://omgenomics.com/circa

of Lake Baikal, Russia (Cabello-Yeves et al., 2018) with various proteobacterial PGCs. The reference PGCs from Proteobacteria were chosen based on tBlastn hits using fused AcsF-BchO-PuhE protein sequences against NCBI's RefSeq genome database. The reasons for using a fused protein sequence as the query are (1), to select for complete and continuous PGCs, instead of fragmented PGC with parts distantly located on a chromosome as often occurred in purple bacterial genomes (Nagashima and Nagashima, 2013); (2), these three genes' locations within PGC are more flexible (Zeng and Koblížek, 2017) and thus more susceptible to evolutionary pressure compared to other PGC genes that form sub-clusters and, therefore, they are more likely to reflect species evolution, as has been demonstrated on the *acsF* gene (Boldareva-Nuianzina et al., 2013; Zeng et al., 2014). The top scoring genome from each group at the class level was downloaded and compared to Gemmatimonadetes PGCs.

For the metabolic reconstruction of *G. groenlandica* and *G. phototrophica*, the genomes annotated by NCBI's Prokaryotic Genome Annotation Pipeline were uploaded to the RAST web server (Aziz et al., 2008) for re-annotation with the original gene prediction information retained. The "Function based Comparison" and "KEGG Metabolic Analysis" functions of RAST were used to analyze both shared and different key metabolic pathways related to a photoheterotrophic life strategy, including central carbon metabolism, energy production, key transporters and membrane structures, and oxidative stress response. The predicted unique functions in one genome were confirmed by tBLASTn searching for homologs in the other genome. If no homologs above the threshold (E < e-05, alignment coverage >40%) were found, the gene queried was designated as a unique gene.

The phylogenomic tree of strain TET16^T was constructed as follows. The protein FASTA files (*_protein.faa.gz) of all Gemmatimonadetes-affiliated genomes including MAGs and isolates (as of October 2020) were downloaded from the NCBI microbial genomes portal via FTP. The PhyloPhlAn pipeline v3.0.58 (Asnicar et al., 2020) was used to automatically retrieve 400 most universal marker genes from each input genome, multi-align each marker gene, concatenate alignments, and infer the phylogenomic tree. The configuration file for the pipeline was generated using the following command "*phylophlan_write_config_file -d a -o gemma_config.cfg –db_aa diamond –map_dna diamond –map_aa diamond –msa mafft –trim trimal –tree1 fasttree –tree2 raxml –verbose.*" The resulting concatenated alignment includes 36,076 amino acid positions. The genome of *Fibrobacter succinogenes* strain S85 (GenBank assembly no. GCA 000146505.1) was used as the outgroup. The programs FastTree and RAxML (Stamatakis, 2014) were used to build the trees using the PhyloPhlAn 3.0 database in an accurate mode with the diversity level set as medium. Due to high computational cost, bootstrapping on the output RAxML tree was not performed. Instead, the refined phylogeny (the RAxML best tree) produced by RAxML starting from the FastTree phylogeny was selected as the final phylogenomic tree. The tree was edited online at the website of iTOL (Letunic and Bork, 2019). The genome-based taxonomy of TET16^T was computed using the command *classify_wf* of the GTDB-Tk tool kit (ver 1.4.0, release R95; Parks et al., 2018).

Morphology, Phenotypic and Chemotaxonomic Characterization

Strain TET16^T grows well on standard R2A agar and in corresponding R2B liquid media. The medium established for optimal growth has been designated as T21 and contains (L^{-1} 0.5 g yeast extract, 0.5 g peptone, 1.0 g K_2HPO_4, and 0.5 g pyruvate with a modified SL-8 trace element solution (refer to DSMZ medium 1222) as followed (final conc. L^{-1}): 5.2 mg Na_2-EDTA, 2.09 mg $FeSO_4 \times 7\ H_2O$, 190 μg $CoCl_2 \times 6\ H_2O$, 122 μg $MnCl_2 \times 4\ H_2O$, 70 μg $ZnCl_2$, 24 μg $NiCl_2 \times 6\ H_2O$, 36 μg $Na_2MoO_4 \times 2\ H_2O$, 62 μg H_3BO_3, 17 μg $CuCl_2 \times 2\ H_2O$, 266 μg $SrCl_2$, 14.7 mg $CaCl_2 \times 2\ H_2O$, and 20.3 mg $MgCl_2$. The pH was adjusted to 7.25–7.3 by addition of 1M HCl solution. The colonies on solid agar plates were observed after 3–5 days of incubation at 23°C aerobically under 12/12 h light/dark regime. Cell imaging was performed using a JEOL JSM-7401F scanning electron microscope (SEM) and a JEOL JEM-1010 transmission electron microscope (TEM) with standard protocols at the Laboratory of Electron Microscopy, Biology Centre of ASCR, České Budějovice, Czechia[5].

Cell growth of the strain TET16^T at different temperatures (4, 10, 15, 18, 20, 22, 25, 30, and 35°C) and pH (4, 5, 6, 7, 8, 9, and 10) was examined using T21 media. The following pH buffer solutions were used: acetic acid/sodium acetate for pH 4-6, K_2HPO_4/KH_2PO_4 for pH 6-8, sodium bicarbonate/sodium carbonate for pH 9-10. Growth on various NaCl concentrations (0.1, 0.5, 1, 2, 3, 4, and 5) (w/v) was investigated. All the physiological experiments were performed in T21 broth media under 12/12 h light/dark regime. Fermentative growth and anaerobic phototrophic growth was assayed as described previously (Zeng et al., 2015) using T21 media. The microaerobic growth condition (1% or 10% O_2) was created by purging an anaerobic jar with commercially purchased gases made by mixing air with pure nitrogen gas in a corresponding ratio.

Antibiotics tests were performed on T21 agar plates supplemented with the following antibiotics (25 mg L^{-1} as working concentration unless stated otherise): bacitracin, chloramphenicol, nystatin (100 mg L^{-1}), neomycin, amoxicillin, tetracycline (15 mg L^{-1}), and amphotericin B (15 mg L^{-1}). Carbon source utilization was carried out at 23°C using the carbon-free minimal media containing K_2HPO_4 (0.5 g L^{-1}), 1 mL trace element solution SL-8 supplemented with one of the following carbon sources (5 mM): saccharin, salicin, adonitol, trehalose, dulcitol, rhamnose, pyruvate, D-glucose, xylose, ribose, erythritol, turanose, cellobiose, melibiose, lyxose, and arabinose. For nitrogen source tests, the minimal media was modified as (L^{-1}): 0.5 g K_2HPO_4, 0.5 g pyruvate, 1 mL trace element solution SL-8 added with one of following nitrogen sources: NH_4Cl (5 mM), KNO_3 (5 mM), glutamine (5 mM), KNO_3 (5 mM), $NaNO_2$ (5 mM), yeast extract (0.5 g L^{-1}), and casamino acids (0.5%). For vitamin tests, the T21 medium was supplemented with one of the following vitamins (final

[5]http://triton.paru.cas.cz/old-lem

conc. L^{-1}): biotin (15 μg), folic acid (100 μg), pyridoxin HCl (15 μg), PABA (300 μg), niacin (100 μg), thiamine HCl (500 μg), riboflavin (100 μg), nicotinamide (500 μg), and B12 (15 μg).

BChl *a* fluorescence from colonies in the near infrared region was initially detected with a lab-assembled infra-red colony imaging system as described before (Zeng et al., 2014). The pigment composition was further analyzed and confirmed using high-performance liquid chromatography (HPLC). The cells were harvested from 5 to 6 days old T21 liquid media by centrifugation (10,000 × *g* for 3 min). The pellet was extracted with 100% methanol. 20 μL of the mix was injected into Nexera LC-40 HPLC system (Shimadzu, Japan) equipped with Kinetex 2.6 μm C8 100Å column (150 mm × 4.6 mm, Phenomenex) heated at 40°C. A binary solvent system was used: A, 25% 28 mM ammonium acetate + 75% methanol; B, 100% methanol at a constant flow rate 0.8 mL min^{-1}. BChl *a* and carotenoids were observed at 770 and 490 nm, respectively.

Respiratory quinones were extracted with 1 mL 7:2 (vol:vol) aceton:methanol mixture. The debris was removed by 3-min centrifugation in an Eppendorf desktop centrifuge at the top speed. The quinones were analyzed using Prominence-i LC-2030C HPLC system equipped with UV-VIS diode-array detector (Shimadzu Inc., Japan). Respiratory quinones were separated on a heated Luna 3 μm C18(2) 100Å 150 × 4.6 mm column (Phenomenex Inc., United States) using binary solvent system: A – 100% methanol; B – 10:3 methanol/heptane (vol:vol). The eluted quinones were detected at 275 nm and identified based on the retention time and absorption spectra. Natural menaquinones extracted from *Micrococcus luteus*, and purchased ubiquinone-10 were used as control standards. Analysis of polar lipids and fatty acids were carried out by the identification service and Dr. Brian Tindall, at DSMZ (Braunschweig, Germany). The *in vivo* absorption spectra were recorded on a Shimadzu UV2600 spectrophotometer.

Gemmatimonadetes MAGs From High Arctic Greenland

We previously reported six MAGs of Gemmatimonadetes origin in a metagenomics study of high arctic soil and glacier in Northeast Greenland (Zeng et al., 2020). Detailed analysis of these MAGs other than general description was not carried out in that study. Here we further present the functional genes and phylogenetic data of these Gemmatimonadetes MAGs with the aim to assess potential importance and metabolic diversity of phototrophic Gemmatimonadetes bacteria in Greenlandic environments. The six MAGs include ES-bin-14, ES-bin-29, ES-bin-51, and ES-bin-78 that were assembled from the exposed surface soil metagenome (designated ES) at the "Lille Firn" glacier (designated LF, GPS: 81.566° N, 16.363° W) close to the Villum Research Station in Northeast Greenland and the two bins LF-bin-215 and LF-bin-339 that were assembled from the LF surface ice metagenome (see more details on sampling in Zeng et al., 2020). The genome annotations of the MAGs were downloaded from the NCBI microbial genome portal via FTP.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

YZ conceived the study. YZ wrote the manuscript with input from Nupur and MK. MK analyzed the respiratory quinones. Nupur performed the phenotypic and physiological characterization and HPLC with help from YZ and MK. NW carried out the fieldwork. YZ and AM performed the MALDI-TOF MS-related work. Nupur and ATG optimized the cultivation media. YZ and XC assembled and analyzed the genomes and reconstructed the phylogenies. All authors read and approved the final version.

FUNDING

This work was supported through the Villum Experiment grants (Nos. 17601 and 32832) and a Marie Skłodowska-Curie AIAS-COFUND fellowship (EU-FP7 Program, under the Grant Agreement No. 609033) to YZ. MK, ATG, and Nupur were supported by the PhotoGemm+ project no. 19-28778X financed by the Czech Science Foundation.

ACKNOWLEDGMENTS

We thank Tina Thane for the help on Illumina sequencing, Thanassis Zervas for the initial help on hybrid assembly, and Jason Dean for his help with electron microscopy. We are grateful to the three reviewers from a previous submission for their critical comments that helped improve this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.606612/full#supplementary-material

Supplementary Figure 1 | Mauve genome alignment of *G. groenlandica* and *G. phototrophica* demonstrating the highly conserved genomic regions housing the photosynthesis gene cluster.

Supplementary Figure 2 | DSMZ report on the fatty acids composition in *G. groenlandica* TET16^T.

Supplementary Figure 3 | Polar lipids in *G. groenlandica* TET16^T by thin-layer chromatography.

Supplementary Figure 4 | Respiratory quinones in *G. groenlandica* TET16^T by liquid chromatography.

Supplementary Table 1 | Comparison of the genome annotations of *G. groenlandica* and *G. phototrophica*.

REFERENCES

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* 11:2500.

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 Genom.* 9:75. doi: 10.1186/1471-2164-9-75

Battistuzzi, F. U., Feijao, A., and Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* 4:44. doi: 10.1186/1471-2148-4-44

Boldareva-Nuianzina, E. N., Bláhová, Z., Sobotka, R., and Koblížek, M. (2013). Distribution and origin of oxygen-dependent and oxygen-independent forms of Mg-protoporphyrin monomethylester cyclase among phototrophic *proteobacteria*. *Appl. Environ. Microbiol.* 79, 2596–2604. doi: 10.1128/aem.00104-13

Cabello-Yeves, P. J., Zemskaya, T. I., Rosselli, R., Coutinho, F. H., Zakharenko, A. S., Blinov, V. V., et al. (2018). Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl. Environ. Microbiol.* 84:e02132-17.

Chee-Sanford, J., Tian, D., and Sanford, R. (2019). Consumption of N2O and other N-cycle intermediates by *Gemmatimonas aurantiaca* strain T-27. *Microbiol. SGM* 165, 1345–1354. doi: 10.1099/mic.0.000847

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

Dachev, M., Bína, D., Sobotka, R., Moravcová, L., Gardian, Z., Kaftan, D., et al. (2017). Unique double concentric ring organization of light harvesting complexes in *Gemmatimonas phototrophica*. *PLoS Biol.* 15:e2003943. doi: 10.1371/journal.pbio.2003943

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. doi: 10.1093/molbev/msz189

DeBruyn, J. M., Nixon, L. T., Fawaz, M. N., Johnson, A. M., and Radosevich, M. (2011). Global biogeography and quantitative seasonal dynamics of gemmatimonadetes in soil. *Appl. Environ. Microbiol.* 77, 6295–6300. doi: 10.1128/aem.05005-11

Hanada, S., and Sekiguchi, Y. (2014). "The phylum gemmatimonadetes," in *The Prokaryotes – Other Major Lineages of Bacteria and the Archaea*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer), 677–681. doi: 10.1007/978-3-642-38954-2_164

Hasholt, B., and Hagedorn, B. (2000). Hydrology and geochemistry of river-borne material in a high arctic drainage system, Zackenberg, Northeast Greenland. *Arctic Antarct. Alpine Res.* 32, 84–94. doi: 10.1080/15230430.2000.12003342

Jaffe, A. L., Castelle, C. J., Dupont, C. L., and Banfield, J. F. (2019). Lateral gene transfer shapes the distribution of RuBisCO among candidate phyla radiation bacteria and DPANN archaea. *Mol. Biol. Evol.* 36, 435–446. doi: 10.1093/molbev/msy234

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 1–8.

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi: 10.1128/aem.72.3.1719-1728.2006

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., et al. (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* 11, 853–862. doi: 10.1038/ismej.2016.174

Kato, K., Tanaka, R., Sano, S., Tanaka, A., and Hosaka, H. (2010). Identification of a gene essential for protoporphyrinogen IX oxidase activity in the *Cyanobacterium synechocystis* sp. *PCC*6803. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16649–16654. doi: 10.1073/pnas.1000771107

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, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0

Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* 39, 854–870. doi: 10.1093/femsre/fuv032

Koblížek, M., Dachev, M., Bína, D., Nupur, L., Piwosz, K., and Kaftan, D. (2020). Utilization of light energy in phototrophic Gemmatimonadetes. *J. Photochem. Photobiol. B.* 213:112085. doi: 10.1016/j.jphotobiol.2020.112085

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Lagier, J. C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., et al. (2018). Culturing the human microbiota and culturomics. *Nat. Rev. Microbiol.* 16, 540–550. doi: 10.1038/s41579-018-0041-0

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259.

Lynch, M. D., and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* 13, 217–229. doi: 10.1038/nrmicro3400

Nagashima, S., and Nagashima, K. V. (2013). Comparison of photosynthesis gene clusters retrieved from total genome sequences of purple bacteria. *Adv. Bot. Res.* 66, 151–178. doi: 10.1016/b978-0-12-397923-0.00005-9

Park, D., Kim, H., and Yoon, S. (2017). Nitrous oxide reduction by an obligate aerobic bacterium, *Gemmatimonas aurantiaca* strain T-27. *Appl. Environ. Microbiol.* 83, e502–e517.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Pascual, J., Foesel, B. U., Geppert, A., Huber, K. J., Boedeker, C., Luckner, M., et al. (2018). *Roseisolibacter agri* gen. nov., sp nov., a novel slow-growing member of the under-represented phylum Gemmatimonadetes. *Intern. J. Syst. Evol. Microbiol.* 68, 1028–1036. doi: 10.1099/ijsem.0.002619

Pascual, J., Garcia-Lopez, M., Bills, G. F., and Genilloud, O. (2016). *Longimicrobium terrae* gen. nov., sp nov., an *Oligotrophic bacterium* of the under-represented phylum Gemmatimonadetes isolated through a system of miniaturized diffusion chambers. *Intern. J. Syst. Evol. Microbiol.* 66, 1976–1985. doi: 10.1099/ijsem.0.000974

Pastor, A., Freixa, A., Skovsholt, L. J., Wu, N., Romaní, A. M., and Riis, T. (2019). Microbial organic matter utilization in high-arctic streams: key enzymatic controls. *Microb. Ecol.* 78, 539–554. doi: 10.1007/s00248-019-01330-w

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

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

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tabita, F. R., Hanson, T. E., Li, H., Satagopan, S., Singh, J., and Chan, S. (2007). Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol. Mol. Biol. Rev.* 71, 576–599. doi: 10.1128/mmbr.00015-07

Tahon, G., and Willems, A. (2017). Isolation and characterization of aerobic anoxygenic phototrophs from exposed soils from the Sof. Rondane Mountains, East Antarctica. *Syst. Appl. Microbiol.* 40, 357–369. doi: 10.1016/j.syapm.2017.05.007

Vavourakis, C. D., Mehrshad, M., Balkema, C., Van Hall, R., Andrei, A. Ş, Ghai, R., et al. (2019). Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biol.* 17:69. doi: 10.1186/s12915-019-0688-7

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., et al. (2016). The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Res.* 44, D590–D594. doi: 10.1007/8623_2015_119

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

Youssef, N. H., and Elshahed, M. S. (2009). Diversity rankings among bacterial lineages in soil. *ISME J.* 3, 305–313. doi: 10.1038/ismej.2008.106

Zeng, Y., Baumbach, J., Barbosa, E. G. V., Azevedo, V., Zhang, C., and Koblížek, M. (2016). Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environ. Microbiol. Rep.* 8, 139–149. doi: 10.1111/1758-2229.12363

Zeng, Y., Chen, H., Madsen, A. M., Zervas, A., Nielsen, T. K., Andrei, A., et al. (2020). Potential rhodopsin and bacteriochlorophyll-based dual phototrophy in a high Arctic glacier. *mBio* 11:e02641-20. doi: 10.1128/mBio.02641-20

Zeng, Y., Feng, F., Medová, H., Dean, J., and Koblížek, M. (2014). Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7795–7800. doi: 10.1073/pnas.1400295111

Zeng, Y., and Koblížek, M. (2017). "Phototrophic Gemmatimonadetes: a new "purple" branch on the bacterial tree of life," in *Modern Topics in the Phototrophic Prokaryotes*, ed. P. C. Hallenbeck (Cham: Springer), 163–192. doi: 10.1007/978-3-319-46261-5_5

Zeng, Y., Selyanin, V., Lukeš, M., Dean, J., Kaftan, D., Feng, F., et al. (2015). Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium *Gemmatimonas phototrophica* sp. nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. *Intern. J. Syst. Evol. Microbiol.* 65, 2410–2419. doi: 10.1099/ijs.0.000272

Zervas, A., Zeng, Y., Madsen, A. M., and Hansen, L. H. (2019). Genomics of aerobic photoheterotrophs in wheat phyllosphere reveals divergent evolutionary patterns of photosynthetic genes in *Methylobacterium* spp. *Genome Biol. Evol.* 11, 2895–2908. doi: 10.1093/gbe/evz204

Zhang, H., Sekiguchi, Y., Hanada, S., Hugenholtz, P., Kim, H., Kamagata, Y., et al. (2003). *Gemmatimonas aurantiaca* gen. nov., sp nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phyl. nov. *Intern. J. Syst. Evol. Microbiol.* 53, 1155–1163. doi: 10.1099/ijs.0.02520-0

Zorz, J. K., Sharp, C., Kleiner, M., Gordon, P. M., Pon, R.T., Dong, X., et al. (2019). A shared core microbiome in soda lakes separated by large distances. *Nat. Commun*. 10, 1–10.

Conflict of Interest: XC is currently employed at BGI Europe A/S, Denmark.

The remaining 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 © 2021 Zeng, Nupur, Wu, Madsen, Chen, Gardiner and Koblížek. 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

REVIEW
published: 10 February 2021
doi: 10.3389/fmicb.2021.630013

Bioprospecting of Novel Extremozymes From Prokaryotes—The Advent of Culture-Independent Methods

Maksim Sysoev[1], Stefan W. Grötzinger[1], Dominik Renn[1], Jörg Eppinger[1,2], Magnus Rueping[1,2] and Ram Karan[1]**

[1] KAUST Catalysis Center (KCC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia,
[2] Institute for Experimental Molecular Imaging, University Clinic, RWTH Aachen University, Aachen, Germany

OPEN ACCESS

Edited by:
André Antunes,
Macau University of Science and Technology, China

Reviewed by:
Giovanna Cristina Varese,
University of Turin, Italy
Judith Maria Braganca,
Birla Institute of Technology and Science, India
Simona Di Gregorio,
University of Pisa, Italy

***Correspondence:**
Magnus Rueping
magnus.rueping@kaust.edu.sa;
Magnus.Rueping@rwth-aachen.de
Ram Karan
ram.karan@kaust.edu.sa

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

Received: 16 November 2020
Accepted: 21 January 2021
Published: 10 February 2021

Citation:
Sysoev M, Grötzinger SW, Renn D, Eppinger J, Rueping M and Karan R (2021) Bioprospecting of Novel Extremozymes From Prokaryotes—The Advent of Culture-Independent Methods. Front. Microbiol. 12:630013. doi: 10.3389/fmicb.2021.630013

Extremophiles are remarkable organisms that thrive in the harshest environments on Earth, such as hydrothermal vents, hypersaline lakes and pools, alkaline soda lakes, deserts, cold oceans, and volcanic areas. These organisms have developed several strategies to overcome environmental stress and nutrient limitations. Thus, they are among the best model organisms to study adaptive mechanisms that lead to stress tolerance. Genetic and structural information derived from extremophiles and extremozymes can be used for bioengineering other nontolerant enzymes. Furthermore, extremophiles can be a valuable resource for novel biotechnological and biomedical products due to their biosynthetic properties. However, understanding life under extreme conditions is challenging due to the difficulties of *in vitro* cultivation and observation since > 99% of organisms cannot be cultivated. Consequently, only a minor percentage of the potential extremophiles on Earth have been discovered and characterized. Herein, we present a review of culture-independent methods, sequence-based metagenomics (SBM), and single amplified genomes (SAGs) for studying enzymes from extremophiles, with a focus on prokaryotic (archaea and bacteria) microorganisms. Additionally, we provide a comprehensive list of extremozymes discovered via metagenomics and SAGs.

Keywords: extremophile, culture-independent methods, metagenomics, SAG, halophiles, thermophiles, psychrophile

INTRODUCTION

A mounting paradigm shift toward using sustainable resources has stimulated exploring new efficient approaches in technological processes (Raddadi et al., 2015; Lamers et al., 2016; Krüger et al., 2018). Enzymes, as natural catalysts, have shown remarkable abilities that have revolutionized the chemical, biotechnological, bioremediation, agricultural, and pharmaceutical industries (Martin and Vandenbol, 2016; Wiltschi et al., 2020). However, the narrow range of stability of most described biocatalysts from mesophilic organisms limits their use for many applications (Raddadi et al., 2015). Enzymes derived from microorganisms thriving under harsh conditions, called extremophiles, can overcome these restrictions, and today, such biocatalysts are in higher demand than ever before (Karan et al., 2012a; Rizk et al., 2012; Johnson, 2014; Raddadi et al., 2015; Sarmiento et al., 2015; Singh et al., 2016; Jorquera et al., 2019). The diversity of

extreme environments promises to reveal biocatalysts capable of withstanding harsh industrial conditions, providing better efficiency with the lower environmental burden (Elleuche et al., 2014; Grötzinger et al., 2017). Extremophiles are present in all three domains of life (bacteria, archaea, and eukarya) (Rothschild and Mancinelli, 2001; Jorquera et al., 2019). According to their natural habitats, extremophiles are classified into thermophiles, alkaliphiles, acidophiles, halophiles, and others (Karan et al., 2012a,b; Raddadi et al., 2015). Therefore, compared to their mesophilic equals, extremozymes are usually able to perform reactions under a broader range of conditions (Karan et al., 2011; Reed et al., 2013; Sarmiento et al., 2015). This extended activity range often allows extremophiles to identify as polyextremophiles, as they tolerate multiple extreme conditions (Karan and Khare, 2010; Rekadwad and Khobragade, 2017; Karan et al., 2019). Therefore, polyextremophiles are perfect candidates as a source of novel enzymes for industrial needs (Raddadi et al., 2015; Sarmiento et al., 2015; Lamers et al., 2016; Krüger et al., 2018).

However, only a small percentage of all microorganisms can be grown in a laboratory environment. This is particularly the case for archaea, representing a large group of prokaryotic extremophiles (Kaeberlein et al., 2002; Zengler et al., 2002; Buerger et al., 2012). Improving computing and bioinformatics technologies have made it possible to study the "dark matter" of the microbial world by looking at the genome data extracted from microbial habitats (Hedlund et al., 2014).

The advent of next-generation sequencing allowed researchers to sequence thousands of microorganisms in parallel, and due to the high sensitivity of these genome-based methods, now we can study microbes with a very low abundance that may be overlooked by other methods (Grötzinger et al., 2014; Laurence et al., 2014; Karan et al., 2020). Accordingly, sequence-based metagenomics (SBM) was born, which studies microorganisms by randomly shearing environmental DNA, sequencing it, and assembling the reads (Hugenholtz and Tyson, 2008).

Further advances allowed researchers to study the genome from single cells (Grötzinger et al., 2014, 2017; Akal et al., 2019; Karan et al., 2019; Vogler et al., 2020). Single amplified genome (SAG) technology separates individual cells before analyzing their DNA, thus giving us information about each cell's genome instead of bulk metagenomes. SAG technology specifically allows whole-genome assemblies from small-sample volumes with low cell yields and low cell abundance compared to those of other cells within a given sample. This technology has proven to be especially useful for studying extremophiles, as their environment makes their cultivation and consequent genomic study particularly complicated (Karan et al., 2019). As the coherent successor of the SBM approach, SAG solved some bottlenecks by introducing a cell-sorting step, making the subsequent sequence analysis of a complex sample easier and more straightforward (Grötzinger et al., 2014). Although SBM and SAG technologies have been available for years, the expected boost in biotechnology has not been realized (Martin and Vandenbol, 2016). Despite the progress in next-generation sequencing technologies, relatively few new extremozymes have been discovered and functionally characterized using culture-independent methods. This shortfall is mainly, in addition to the high sequencing costs, because of the lack of reliable, functional annotation of the genomic data caused by the low amount (0.09%) of experimentally described genes (Grötzinger et al., 2014, 2017; Attrill et al., 2019).

Generally, annotation algorithms depend on existing functional annotations, thus giving reliable results for already well-studied branches of life.

Therefore, culture-based methods are still needed to study a novel enzyme's function and structure, confirmed by several studies (Martin and Vandenbol, 2016; Speda et al., 2017).

Nevertheless, the lack of a specific database limits our choice for the right expression model to study the enzyme of interest. Consequently, the combination of these issues creates a vicious loop resulting in low detection and usage of extremozymes. This review aims to highlight the importance of extremozyme from prokaryotes research for industry by showing how recent developments in sequencing, computation and bioinformatics solve the problems associated with data mining for extremozymes and outline the annotation difficulties, which must be overcome in the future.

ENZYMES: A SUSTAINABLE SOURCE FOR GREEN CHEMISTRY

Enzymes are broadly used in biotechnology and a variety of industries (e.g., agriculture, food, textiles, chemicals, pharmaceuticals, and biofuels) as catalysts, therapeutic agents, analytic reagents, and diagnostic tools (Illanes et al., 2012; Robinson, 2015; Ramesh et al., 2020). Enzymes have broadened the horizon of potential applications by allowing us to perform chemo- and regioselective reactions, which is a big struggle for current chemical techniques (Rasor and Voss, 2001; Schäfer et al., 2007). In general, enzymatic reactions are safer, faster, less hazardous, and generate less waste, thus following the twelve rules of green chemistry (Anastas and Eghbali, 2010). Especially today, in a world with 7.5 billion people and an expected 9.8 billion 30 years from now (United Nations, Department of Economic and Social Affairs, Population Division, 2017), it is crucial to utilize the high potential of enzymes for biotechnological applications and green chemistry to reduce humanity's overconsumption of resources (Anastas and Eghbali, 2010; Sheldon, 2016).

Although enzymes have vast potential in biotechnological applications, they have been used only in very few specific reactions. Most described enzymes can be used only for a limited number of industrial processes (Herbert, 1992; Elleuche et al., 2014). This limitation is caused by the narrow ranges of enzymatic stability, including a majority of essential parameters for chemical reactions, such as temperature, pressure, pH, and the use of organic solvents. For example, although water is considered the solvent of life, it is a poor solvent for synthetic reactions (Aitken and Brown, 1969). Organic solvents are used not only to increase the solubility of hydrophobic substrates but also to shift the thermodynamic equilibrium from hydrolysis

to condensation and suppress water-dependent side reactions (Carrea and Riva, 2000).

A number of genetic (Gatti-Lafranconi et al., 2010) and chemical (Siddiqui et al., 2009) modifications together with immobilization strategies (Mukhopadhyay et al., 2015) have been developed to overcome these restrictions (Stepankova et al., 2013). These strategies aim to increase the enzyme's stability or decrease the denaturing effect of the reaction conditions. To date, there is no ultimate mechanism to increase the stability of enzymes. A point mutation can drastically affect the enzyme's efficiency and stability, although it is difficult to see any apparent trends or patterns to fully control a given enzyme (Vieille and Zeikus, 2001; Sarmiento et al., 2015). It is essential to identify mutagenesis strategies to alter enzymes from mesophilic organisms to achieve stability in extreme physicochemical conditions, which is a challenging task (Reed et al., 2013). Although various stabilizing immobilization strategies exist, several hurdles, such as a limited increase in stability, change in reaction rate, alteration of stabilities under other conditions, and potential influence on the specificity of the enzymes, have not allowed these approaches to be used in a broader range (Coker, 2016).

EXTREMOPHILES

An alternative approach to achieve biocatalysis in extreme physicochemical conditions is to use enzymes derived from organisms that thrive in extreme conditions (Jorquera et al., 2019). These organisms are called extremophiles and live in harsh environments of elevated temperatures (thermophiles), salt concentration (halophiles), pressure (barophiles), osmotic compound content (osmophiles), heavy metal content (metalophiles), and radiation (radiophiles); acidic or basic pH (acid or alkaliphiles); extreme dryness (xerophiles); extreme cold (psychrophiles); or a combination of different extremes (polyextremophiles) (**Figure 1**; Kristjánsson and Hreggvidsson, 1995; Gerday and Glansdorff, 2009; Gabani and Singh, 2013; Coker, 2016; Horikoshi, 2016).

Halophiles and Xerophiles

Halophiles are microorganisms that require elevated salt concentrations to grow. There are four main classifications of halophiles based on their optimal NaCl concentration for growth: slight halophiles (0.2 M), moderate halophiles (0.5–2.5 M), borderline extreme halophiles (2.5–4.0 M), and extreme halophiles (4.0–5.9 M). Halophilic organisms use several adaptation mechanisms to live in these conditions. These include modified electrostatic charge of their proteins, balancing osmotic pressure either by compatible solutes, such as betaine and ectoines, or with chloride and potassium uptake into the cells by transporters (primary or secondary) and the coordinated action of bacteriorhodopsin and ATP synthase (Karan et al., 2012a).

Halophilic organisms and their enzymes are used in several different fields, including production of fermented foods, manufacturing of solar salt from seawater, leather industries, environmental bioremediation and textile, pharmaceutical (Karan and Khare, 2010; Oren, 2010; Karan et al., 2012a; de Lourdes Moreno et al., 2013; Akal et al., 2019). The biochemical properties of the molecules, enzymes, and compatible solutes synthesized by halophiles also present potential implications in fine chemicals, medicines, and bioimplants (Irwin, 2010; Karan et al., 2012a,b; Chen and Jiang, 2018; Jin et al., 2019). One of the main selling points of these enzymes is their stability in such solvents as benzene, toluene or chloroform, which are frequently used in different industrial environments (Oren, 2010). Halophiles are a potential source of novel extremozymes like amylases, proteases, nucleases, cellulases, chitinases, xylanases, esterases, alcohol dehydrogenases and lipases (Karan et al., 2012a; Coker, 2016). For example, nuclease from *Micrococcus varians* has long been used commercially for production of a flavoring agent 5′-guanylic acid (5′-GMP) due to its efficiency in degrading RNA at 60°C and 12% (w/v) salt (Kamekura et al., 1982). Several other enzymes have been heterologously expressed and characterized (**Table 1**). Interestingly, some of the halophilic enzymes display polyextremophilicity i.e., stability toward more than one extreme condition e.g., high salt, elevated or low temperature, alkaline or acidic pH and non-aqueous medium having great potential application for industrial and biotechnological processes (Grötzinger et al., 2017; Akal et al., 2019; Karan et al., 2020).

Along with the enzymes, halophiles possess a number of bioactive molecules that found its application in different areas. The biocompatible solute ectoine has potential use in respiratory medicine (Galinski et al., 1985) and can reduce cell stress effects in nanoparticle-induced lung inflammation by inhibiting the signals (Sydlik et al., 2009). Some halophiles, such as *Halobacterium* and *Haloferax*, were found to accumulate polyhydroxyalkanoates (PHAs), which are a family of biopolyesters with diverse uses in the medical, environmental, and industrial fields (Kirk and Ginzburg, 1972; Quillaguaman et al., 2010; Chen and Patel, 2012). Halophiles, such as *Natronococcus occultus* and *Naloterrigena hispanica,* have been found to generate antimicrobial peptides and diketopiperazines (Charlesworth and Burns, 2015; Coker, 2016). Some haloarchaea produce internal protein gas vesicle nanoparticles (GVNPs), which have been tested as a vaccine scaffold platform and ultrasonic contrast agents (Stuart et al., 2001; Shapiro et al., 2014; DasSarma and DasSarma, 2015). Notably, halophilic nature makes cell lysis, hence farming of the enzymes and other molecules, easier. Since halophiles require high salt to maintain their cell integrity, lowering the salt concentration in the buffer system disrupts the cell wall of most halophiles.

Xerophiles are microorganisms that can survive and grow in arid conditions with water activity $a_w < 0.75$ by forming spores that help them mitigate environmental stress. Adaptive mechanisms are connected to water loss prevention and increased water retention through the accumulation of compatible solutes, production of extracellular polymeric substances (EPSs), adaptations on the cell membrane to retain intracellular water, and synthesis of DNA repair proteins (Lebre et al., 2017). These unique adaptations allow xerophiles to be used in microbial electrochemical systems (Dopson et al., 2015) or in next-generation industrial biotechnology, where they can

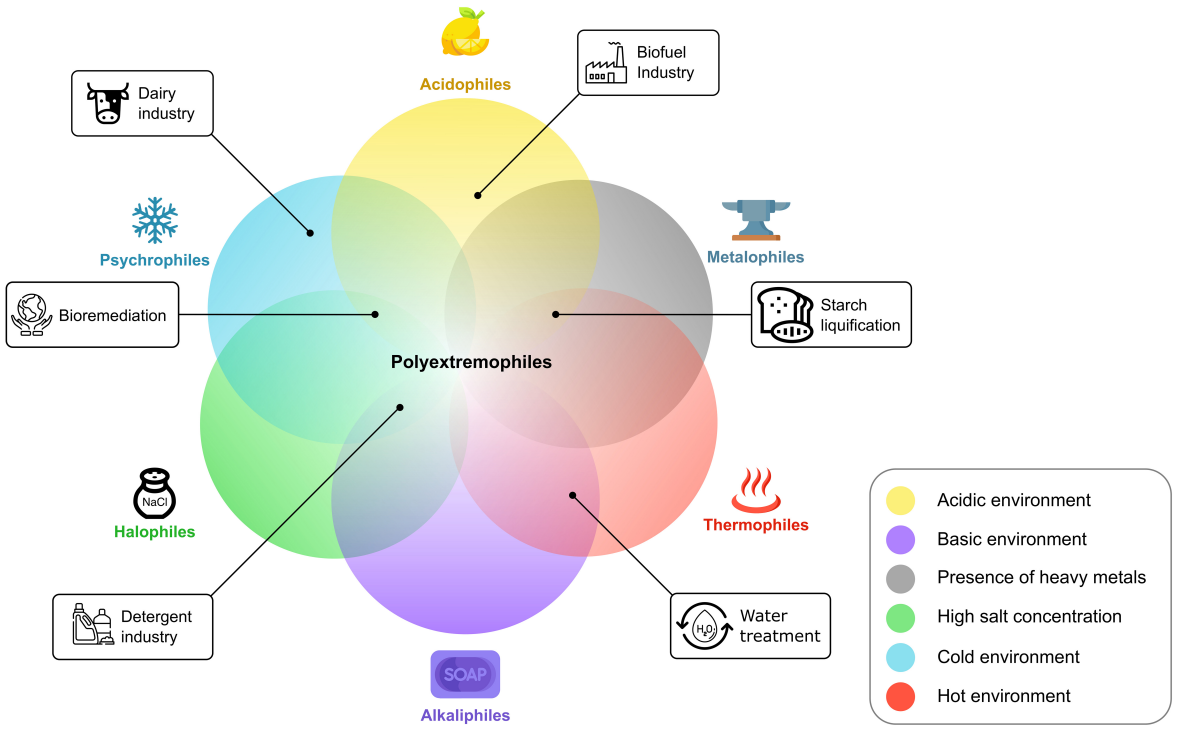


FIGURE 1 | Selection of different extremophiles. The corresponding image displays the extreme condition(s) in which they thrive. Overlaps are examples of potential polyextremophiles and are not limited to the displayed configurations.

be used for treating long-chain fatty acids, cellulose, chitin, rubbers, or other compounds (Chen and Jiang, 2018).

Thermophiles and Psychrophiles

Thermophiles are heat-loving microorganisms with an optimum growth temperature of 45°C or above, while hyperthermophiles grow at temperatures above 80°C (Sarmiento et al., 2015). Thermophiles and hyperthermophiles exist in various natural ecosystems, such as geothermal waters, hot springs, volcanoes, deep-sea hydrothermal vents, and other ecosystems with high-temperature parameters. Thermophiles have thermostable proteins and cell membranes that do not become denatured at high temperatures, and some may also resist proteolysis (Sarmiento et al., 2015). Notably, polymerases from thermophiles have laid the foundation for the discovery of polymerase chain reaction (PCR), a technique that has become crucial in medicine and research. Today we may find various enzymes from thermophiles (e.g., from *Thermus aquaticus* and *Pyrococcus furiosus*) finding their use in PCR due to their stability and reasonable cost (Brock, 1997; Bruins et al., 2001; Irwin and Baird, 2004). Along with polymerases, a lot of different thermophilic enzymes, such as lipases, laccases, and xylanases, are also on the market, thus making the industrial process much more environmentally friendly (Damiano et al., 2003; Atalah et al., 2019).

Psychrophiles are cold-loving microorganisms that can grow at temperatures between −20 and 20°C. Psychrophiles possess diverse adaptive molecular mechanisms to survive and thrive at such low temperatures. Psychrophilic bacteria have increased (i) unsaturated fatty acids, cyclopropane-containing fatty acids, and short-chain fatty acids in their membranes, which prevent the loss of membrane fluidity; (ii) cold-shock proteins (CSPs) and chaperones to protect the synthesis of RNA and proteins; (iii) antifreeze proteins (AFPs) that bind to ice crystals and create a state of thermal hysteresis; and (iv) mannitol and other compatible solutes that act as cryoprotectants to prevent cell damage by ultraviolet (UV) radiation and ice formation (Sarmiento et al., 2015; De Santi et al., 2016a).

Psychrophiles have a promising future in pharmaceuticals and medicine since their cell membranes hold surfactants capable of sustaining stability at low temperatures (Cavicchioli et al., 2011). Psychrophiles, such as *Pandalus borealis*, *Euphausia superba*, *Moraxella* species, and *Flavobacterium* species, have been found to produce anticancer and antitumor agents (Margesin and Feller, 2010). Psychrophilic enzymes display high catalytic activity, stability at low temperatures, and pronounced heat lability and may offer useful industrial and biotechnological applications in various domains, such as pharmaceutical science, molecular biology, textiles, paper, food, feed technologies, detergents, and cosmetics (Margesin and Feller, 2010; Cavicchioli et al., 2011; Sarmiento et al., 2015; Dhaulaniya et al., 2019; Al-Ghanayem and Joseph, 2020).

Acidophiles and Alkaliphiles

Acidophiles are microorganisms that grow at an optimum pH < 3 (Baker-Austin and Dopson, 2007). Acid-tolerant

TABLE 1 | List of extremozymes discovered with metagenomics and single amplified genomes.

Enzyme	Source	Heterologous host	Screening approach
		Function-based metagenomics (FBM)	
β-Glucosidase	Hydrothermal spring	*E. coli*	Schröder et al., 2014
Carboxylesterase	Hot vent sediment	*E. coli*	Placido et al., 2015
Esterase	Glacier soil	*E. coli*	De Santi et al., 2015, 2016a,b
Esterase	Hot spring water, Sediment, and compost	*E. coli, Thermus Thermophilus*	Leis et al., 2015
Esterase	Hot spring water	*E. coli*	López-López et al., 2015
Hormone-sensitive lipase	Permafrost	*E. coli*	Petrovskaya et al., 2016
Lipase	Hot spring sediment	*E. coli*	Sahoo et al., 2017
Lipase	Hot spring water	*E. coli*	Yan et al., 2017
Phospholipase	Hydrothermal vent	*E. coli*	Fu et al., 2015
Serine protease	Hot spring sediment	*E. coli*	Singh et al., 2015
		Sequence-based metagenomics (SBM)	
α-Galactosidase	Hot spring water	*E. coli*	Schröder et al., 2014
Endoglucanase	Hot spring sediment	*E. coli*	Zhao et al., 2017
β-Galactosidase	Hot spring water	*E. coli*	Liu et al., 2015
Esterase	Hot spring mud	*E. coli*	Zarafeta et al., 2016
Mercuric reductase	Brine pool	*E. coli*	Sayed et al., 2014
Amine transferase	Hot spring	*E. coli*	Ferrandi et al., 2017
Epoxide hydrolase	Hot spring	*E. coli*	Ferrandi et al., 2015
		Single amplified genomes (SAG)	
Alcohol dehydrogenases	Brine pool	*Hfx. volcanii*	Grötzinger et al., 2017; Akal et al., 2019
Carbonic anhydrase	Brine pool	*Halobacterium* sp. NRC-1	Vogler et al., 2020
Glucose dehydrogenase	Brine pool	*Halobacterium* sp. NRC-1	Grötzinger et al., 2014; Karan et al., 2019
2-hydroxy dehydrogenase	Brine pool	*Halobacterium* sp. NRC-1	Grötzinger et al., 2014; Karan et al., 2019
Protease	Brine pool	*Halobacterium* sp. NRC-1, *E. coli*	Grötzinger et al., 2014; Karan et al., 2019

microbes have optimum growth at pH > 5 but are still active in lower pH environments. Acidophiles maintain their cytoplasmic pH close to neutrality to protect acid-labile cellular constituents by active pumping of protons (proton flux system); by decreased permeability of the cell membrane, which helps to suppress the entry of protons into the cytoplasm; and by improved protein and DNA repair systems compared to those of neutrophiles (Baker-Austin and Dopson, 2007). Acidophilic enzymes can block the activity of matrix metallopeptidases (MMPs), which are essential for tumor metastasis (Irwin, 2010). MMP inhibitors from an acidophilic *Penicillium* species isolated from Berkeley Pit Lake promise a therapeutic approach for cancer (Stierle et al., 2006). Proteolytic enzymes from acidophiles have also been reported as nonallergenic preservatives in medicines (Sharma et al., 2012). Amylolytic enzymes, such as trehalase isolated from acidophilic *Sulfolobus solfataricus,* are used in medicine as preservatives and stabilizers (Schiraldi et al., 2002).

Alkaliphiles are microorganisms that grow in alkaline environments with a pH > 9, usually showing optimal growth at pH ∼10. Alkaliphiles may coexist with neutrophiles under mild basic pH conditions and live in specific extreme environments. Alkaliphilic bacteria possess molecular mechanisms that compromise the activation of both symporter and antiporter systems. Electrogenic antiporters produce an electrochemical gradient of Na^+ and H^+, and the symporter system enables the uptake of Na^+ and other solutes into the cells (Chinnathambi, 2015). Alkaliphilic enzymes have found use in several applications, such as tannery water treatment (Li et al., 2017), food, cosmetics, and pharmaceutical production (Horikoshi, 2016).

Radiophiles

Radiophiles or radiation-resistant extremophiles thrive in high oxidative stress and radiation environments, including UV radiation, gamma radiation and X-ray radiation, and have potential applications in therapeutics pharmacology and biotechnology. Direct intense or prolonged exposure to different forms of radiation, such as UV radiation, can lead to mutagenic and cytotoxic DNA lesions, resulting in various types of human cancers (Gabani and Singh, 2013). Primary and secondary metabolic products from radiophiles can protect the organism's DNA and can be used to manufacture anticancer drugs, antioxidants, and sunscreens (Raddadi et al., 2015). Extremolytes, mycosporin-like amino acids (MAAs) from the red alga *Porphyra rosengurttii*, are commercially utilized to enhance the UV-protective properties of sunscreens (de la Coba et al., 2009) and are therapeutic candidates as preventive agents in UV radiation-induced cancers, such as melanoma. Bacterioruberin isolated from *Halobacterium* and *Rubrobacter* and deinoxanthin isolated from *Deinococcus radiodurans* are other therapeutic candidates for cancer diseases (Singh and Gabani, 2011; Choi et al., 2014).

Polyextremophiles

Polyextremophiles i.e., microorganisms growing preferentially under multiple extremes, have developed features that allow them to thrive in harsh environments such as Deep Lake Antarctica, where the temperature reaches as low as −20°C and stays liquid only due to extreme salt concentrations (DasSarma et al., 2013; Karan et al., 2013, 2020). Deep-sea anoxic brine pools at the bottom of the Red Sea are another type of polyextremophilic environment (Antunes et al., 2011). These brine pools are extreme in different physicochemical parameters and vary drastically, with temperatures ranging from 22.6 to 68.2°C and NaCl concentrations varying from 2.6 to 5.6 M (Antunes et al., 2011). Additionally, they show a characteristic sharp brine-seawater interface, with steep gradients of dissolved oxygen, density, pH, salinity, and temperature (Emery et al., 1969; Ross, 1972; Anschutz and Blanc, 1995). Because of this variation, brine pools offer a multitude of habitats for different kinds of extremophiles.

Polyextremophiles have a diverse range of uses and applications. Halothermophiles and halopsychrophiles are promising sources of useful enzymes (Elleuche et al., 2014). Enzymes derived from these polyextremophilic microbes possess particularly attractive properties for biotechnology, namely, their function at high salt concentrations and high or low temperatures (Sarmiento et al., 2015). Elevated temperatures are used to shift equilibria, distill products, increase reaction's speed, liquefy compounds, and eliminate microbial contamination (Grötzinger et al., 2017). Low temperatures can save energy (e.g., in washing processes, bioremediation, or food processing), avoid labile or volatile compound production (e.g., in biotransformations or food processing) or prevent bacterial growth. Therefore, cold-active enzymes have great potential for various biotechnological processes. Increasing the salt concentration in solution decreases the water activity, thus mimicking aqueous-organic solvent mixtures. Therefore, halophilic enzymes generally retain high activity and stability in high salt and organic or nonaqueous media (Karan and Khare, 2010, 2011; Karan et al., 2011, 2012a). In turn, organic solvents increase the solubility of hydrophobic substrates and alter the hydrolytic and kinetic equilibria (Sellek and Chaudhuri, 1999). Traditional mesophilic enzymes lose their native structure and thus catalytic activity in organic solvents, limiting their use. However, the high stability of halophilic enzymes toward salt is associated with tolerance to low water activity, such as in mixtures of aqueous and organic or nonaqueous media (Sellek and Chaudhuri, 1999), which emphasizes their high potential for biocatalysis. Cold-active or heat-stable enzymes operating in high-salt or organic solvents are of interest for the sustainable production of value-added chemicals. The world is currently looking for microorganisms with new enzymes, such as hydrolases, amylases, cellulases, peptidases, and lipases (Dumorne et al., 2017). Some examples include alkalithermophilic serine proteases from *Alkalibacillus* sp. NM-Da2, which can potentially be applied in different biotechnological and pharmaceutical industries (Abdel-Hamed et al., 2016), and alkalipsychrophilic esterase from the marine bacterium *Rhodococcus* sp., which can be used in the food industry since the process catalyzed by an esterase can be stopped by increasing the temperature, thus saving food properties (De Santi et al., 2014). The polyextremophilic characteristics of halophilic amylases potentially make them efficient catalysts under alkaline pH and high salinity in processes such as starch hydrolysis and applications such as detergent production, the food industry, and bioremediation (Ali et al., 2014; Kumar et al., 2016; Rekadwad and Khobragade, 2017).

MINING ENZYMES FROM EXTREME ENVIRONMENTS

Enzymes derived from extremophiles, so-called extremozymes, can catalyze chemical reactions in harsh conditions, such as those found in industrial processes (Sarmiento et al., 2015). The current global market for industrial enzymes was 9.9 billion USD in 2019 and is expected to grow to 14.9 billion USD by 2027 (Grand View Research, 2020)[1], in which novel extremozymes could play a significant role and further expand this market.

Extremophiles thrive in extreme habitats, including salt lakes, deep-sea vents, acidic sulfurous lakes, alkaline lakes, hot springs, Arctic and Antarctic waters, and alpine lakes (Kristjánsson and Hreggvidsson, 1995; Gerday and Glansdorff, 2009). The challenges in mining the enzymatic potential of extremophilic locations are the hurdle of mimicking these harsh conditions in the laboratory (Sarmiento et al., 2015), the remote areas, and the unusually low cell density, causing a minimal amount of biomass yields (Ferrer et al., 2007). As a result, despite the high scientific potential and industrial value of extremozymes, very little is known about their structure and function. Therefore, most enzymes currently used in industry originate from either fungi or mesophilic bacteria (Elleuche et al., 2014).

Microbial Dark Matter—The Big Unknown

According to recent estimations, Earth is home to an upward range of 10^{12} (1 trillion) microbial species, with microorganisms being the most abundant, widespread, and taxonomically, metabolically, and functionally diverse organisms (Locey and Lennon, 2016). During the past decade, high-throughput sequencing in combination with advanced bioinformatics algorithms has allowed for enhanced insight into microbial taxa and expanded the estimation of the global microbial load by orders of magnitude via projects such as the "Earth Microbiome project" (EMP), which analyzed more than 200,000 environmental samples (Gilbert et al., 2014). Out of the 10^{12} microbial species, only approximately 10^5 have been sequenced (0.00001%), 10^4 have been cultured, and even the EMP, which also uses rRNA sequences for identification, has cataloged fewer than 10^7 species (0.001%), of which 29% were detected only twice (Locey and Lennon, 2016). It is estimated that approximately 99.999% of microbial taxa remain undiscovered (Locey and Lennon, 2016). Focusing on the $\sim 10^4$ cultivated species and looking at the highest order, among the 60 significant lines of

[1]https://www.grandviewresearch.com/industry-analysis/enzymes-industry, last access date: 2/1/2021 4:43 p.m.

descent (phyla or divisions) that are known within the archaeal and bacterial domain (Hugenholtz and Kyrpides, 2009), 50% are uncultured and make up the "microbial dark matter" (Marcy et al., 2007). Furthermore, 88% of all isolated microbes are members of only four bacterial phyla, *Proteobacteria, Firmicutes, Actinobacteria*, and *Bacteroidetes* (Rinke et al., 2013).

The Challenge of Cultivation

Traditional cultural methods employ the ability of microorganisms to grow in a specific laboratory environment. Historically, these methods have been used in microbiology, including selective or differential media, microscopy, Gram-staining, and biochemical tests. These methods are sensitive, reliable, inexpensive, and provide qualitative and quantitative results on the bacterial populations (Nowrotek et al., 2019). However, only a small fraction of all microbes can be grown in a laboratory setting. Thus, although cultural methods are a well-established, simple, and inexpensive way of isolating, detecting, and quantifying microorganisms, the methods lack speed and are labor-intensive with the risk of contamination. Based on this legacy, microbial culturomics has emerged as a tool that complements metagenomics data and gives another way to determine composition of microbial populations. Several studies have shown that culture-dependent and culture-independent methods often deliver different results (Steven et al., 2007; Carraro et al., 2011; Stefani et al., 2015; Nowrotek et al., 2019; Rego et al., 2019). For example, using 212 different culture conditions, such as temperature, various oxygen levels, and selective media, many researchers have isolated 340 different bacterial species (Lagier et al., 2012). Interestingly, the metagenomics approach could identify only 51 species out of 340 cultured species in this study. Therefore, it is crucial to combine both culture-dependent and culture-independent ways to study a given microbiome (Steven et al., 2007; Lagier et al., 2016; Nowrotek et al., 2019; Sarhan et al., 2019).

The enormous lack of isolates is based on the fact that very few microbes are cultivable, e.g., only as few as 0.001–0.1% of the microbes found in seawater can be cultivated under laboratory conditions (Amann et al., 1995). Therefore, archaea enzymes are typically studied after expression in a heterologous host, mostly *Escherichia coli* (*E. coli*). However, this approach is impossible for some archaeal proteins due to low expression rates, inaccurate protein folding, or the lack of functionally necessary posttranslational protein modifications in eubacterial hosts. Besides, it is challenging to identify archaeal enzymes directly from genome samples in bacterial host-based functional screening assays, as archaeal promoter structures differ from bacterial ones (Zweerink et al., 2017). A direct *in vivo* investigation of archaeal enzymes in an archaeal strain or an archaeal host with the right expression machinery of the respective proteins represents a valuable alternative. For this reason, some archaeal expression models have been used, such as *Halobacterium salinarum*/NRC-1 or *Haloferax volcanii*. Using these expression systems, researchers have studied archaeal TATA-binding proteins, discovered 16S gene transfer between bacteria and archaea (Baliga et al., 2000; Fuchsman et al., 2017), and several different archaeal enzymes have been studied using these archaeal expression systems (**Table 1**). However, this approach requires technically sensitive and straightforward technologies to detect active enzymes and a methodology to express proteins in an archaeal host heterologously. Ultimately, as extremophiles preferentially live in harsh conditions, it is difficult to satisfy all of these microbes' extreme requirements. However, with the progress in computational methods and improving bioinformatics algorithms, it is now possible to study the mysterious "dark matter" of the microbial world using culture-independent approaches, including bulk metagenomics and SAGs.

Function-Based Metagenomics (FBM), Sequence-Based Metagenomics (SBM), and Single Amplified Genomes (SAGs) as a Gateway to Novel Species

Culture-independent methods developed and utilized over the last few years, such as metagenomics, have recently gained momentum. Metagenomics analyses rely on the direct isolation of genomic DNA from the environment. These analyses can be either sequence-based (i.e., putative enzymes are discovered based on their conserved sequences) or function-based (i.e., functional enzymes are found based on the expressed features, such as a specific enzyme activity).

Function-Based Metagenomics (FBM)

The function-based metagenomics (FBM) approach is based on cloning random environmental DNA into expression hosts, such as *E. coli,* forming a library. This library is then screened for function by different assay methods (Rashid and Stingl, 2015). The main challenge of this method is the need for hosts to create functional expression libraries and sufficient screening methods (Rashid and Stingl, 2015). The alternative is the sequence-based approach, which historically was based on sequence homology and used a colony hybridization technique to screen metagenomic clones using an oligonucleotide primer or probes for the target gene. In a subsequent step, the desired gene was amplified by PCR and cloned into suitable expression vectors. This technique led to discovering novel sequences similar to existing known sequences and may help find enzymes more efficiently (Raghunathan et al., 2005; Marcy et al., 2007).

However, metagenomics approaches give little information about expression and regulation of genes in the environment. Recently, metatranscriptomics approaches have been used to reveal gene expression profiles and ecophysiology of natural microbial communities (Hua et al., 2015; Tripathy et al., 2016). Combined genomic and transcriptomic analyses can discover and characterize the relative transcriptional levels of multiple genes and disclose the functional diversity in microbial communities. Such approach already has been utilized for studying extremophiles in acidophilic conditions and provided important data on *Ferrovum* population and its metabolic potentials and gene expression profile (Tripathy et al., 2016).

In the end, metatrascriptomics approach allowed researchers to peak into the ecological role of these unexplored, but potentially important microorganisms.

Sequence-Based Metagenomics (SBM)

SBM became more attractive approach with the advent of high-throughput sequencing, when the main biases and bottlenecks of environmental sampling along with the requirement of DNA cloning before sequencing were removed (Bohmann et al., 2014). Today, the DNA from metagenomic samples can be sequenced without cloning. Furthermore, sometimes it is possible to extract desired gene sequences directly from metagenomics data and synthesize the genes *de novo* after bioinformatic annotation and codon optimization if required. These developments shifted the bottleneck of culture-independent sequencing, mainly to bioinformatics-based sequence alignment, data management, and, in particular, annotation algorithms. Annotation algorithms initially relied on sequence homology-based methods and, therefore, depended on existing functional annotations, giving excellent results only for already well-studied life branches. Therefore, it is advisable to use additional methods to account for this point. Different annotation approaches arose over time, still based on existing functional annotations but offering more degrees of freedom regarding the DNA sequences. Big data analysis requires tremendous computing powers, and for a long time, computers were not ready to handle that kind of load (Fan et al., 2014). Since introduction of the first supercomputer in 1964, the performance has dramatically increased, and now the fastest supercomputer is more than 100 times more powerful than those from a decade ago (**Figure 2A**). Along with the development of powerful computers and advanced computing algorithms, the cost of sequencing has been continuously decreasing, and in the last two decades, it has dropped by a factor of over one million (Wetterstrand, 2013; **Figure 2B**). Increased availability opened up multiple possibilities to study genomes, and several technologies appeared, including first-, second-, and third-generation sequencing. Due to the rapid development of next-generation sequencing, shotgun and high-throughput versions of SBM have become more popular. The SBM method allows for the sequencing of environmental DNA (Thomas et al., 2012) by randomly shearing all metagenomic DNA from one sample, a method depending on the cloning step into vectors for amplification, and sequencing the vector DNA, followed by assembly via algorithms (Rashid and Stingl, 2015).

Despite many benefits, SBM also suffers from limitations and challenges. Not only does the short read length of the DNA sequencing and the size and complexity of the sequence data pose analytic and informatics challenges (Rashid and Stingl, 2015), but also the ability to assemble independent genomes is drastically reduced at both very low abundances and increasing genomic heterogeneity (Albertsen et al., 2013). These challenges result in a low probability of identifying rare populations, occurring in less than 1% of the total metagenomic DNA (Kunin et al., 2008). This effect comes mainly from cross-strain assemblies (Rinke et al., 2014) that originate from different organisms but can also come from unwanted host DNA or environmental contamination (Schmieder and Edwards, 2011) that cannot be avoided. Hence, despite high expectations, metagenomics data have not yet led to the expected boost in biotechnology (Chistoserdova, 2010).

Single Amplified Genomes (SAGs)

SAGs gain the momentum due to the breakthrough of single-cell genome sequencing that has alleviated many bottlenecks in metagenomics by physically separating the genomic material of uncultured cells (Raghunathan et al., 2005; Kvist et al., 2007; Marcy et al., 2007). The **Figure 3** shows a schematic comparison of the critical steps in SBM and SAGs screening using SAGs. In recent years, single-cell genome sequencing has become a highly accessible tool (Chi, 2014). In part, this utility is due to the improvements that have taken place in the protocols for DNA or RNA isolation, leading to more simplified procedures. Additionally, several industrial kits have become available, leading to the amplification of genetic material from single cells (Chi, 2014). To prepare samples for further operations, they need to be separated from the bulk cell mass. There are several different methods for sample preparation of single cells for single-cell genomics (SCG) (Blainey, 2013). These methods include serial dilution (Zhang et al., 2006), micromanipulation (Woyke et al., 2010), optofluidics (optical tweezing in conjunction with microfluidics) (Landry et al., 2013), laser-capture microdissection of tissue samples (Frumkin et al., 2008) and fluorescence-activated cell sorting (FACS) (Swan et al., 2011; Dupont et al., 2012; McLean et al., 2013; Wilson et al., 2014), ultimately resulting in individually sorted samples enriched with a particular microbe into 96-well plates, with a minimal amount of either or both host and environmental contamination. FACS has become the most popular method due to its high performance and the ability to separate individual environmental cells based on various cellular properties (e.g., size, fluorescence, and granularity). Moreover, FACS can also be used to study populations (Thompson et al., 2013). The enriched pool of symbiont cells can undergo whole-genome amplification (WGA) followed by sequencing, yielding a population genome assembly or a homogeneous draft assembly in a clonal population. However, FACS has some technical problems for the sorting of microbial cells. Because it is impossible to confirm cell identity visually, sometimes non-cellular fluorescent particles present in environmental samples can be sorted along with targeted microbial cells (Davey and Kell, 1996; Müller and Nebe-Von-Caron, 2010). For the same reason, FACS also retains a low efficiency in recovering rare cells (Yamamura et al., 2005; Yoshimoto et al., 2013; Nakamura et al., 2016).

After sorting, cells need to be lysed (e.g., with an alkaline solution, detergent, or heat). Furthermore, to obtain DNA contigs, the sample undergoes multiple displacement DNA amplification to sequence genomic DNA from a single microorganism. Subsequently, microorganisms, or unamplified single DNA fragments can be screened in each well using PCR with specific gene primers. WGA can also be performed using bacteriophage Φ29 polymerase to identify specific genes in a genome. Finally, it is vital to monitor DNA contamination to

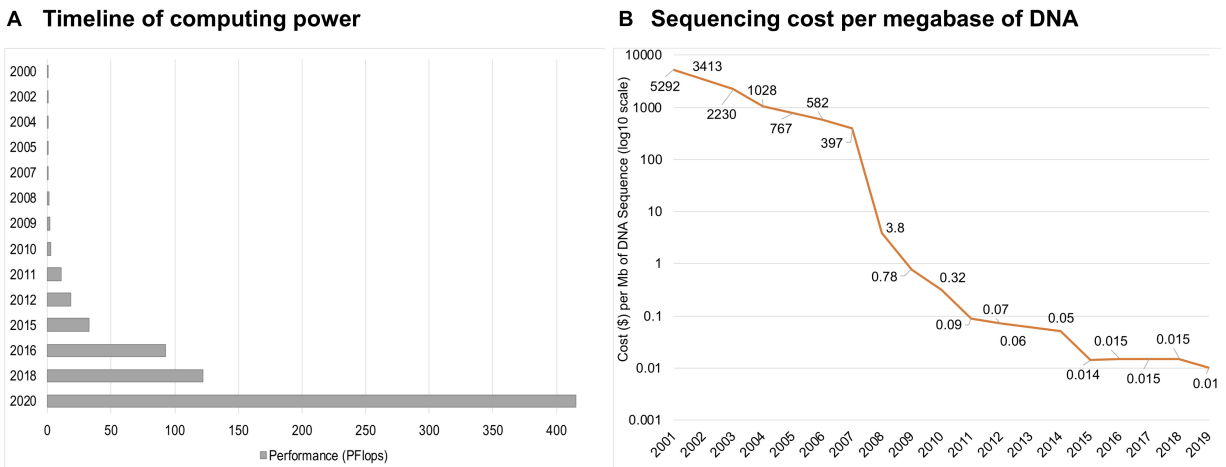


FIGURE 2 | The timeline of computing power shows a dramatic increase over the last two decades **(A)**. Sequencing cost per megabase (Mb; one million bases) of DNA in USD **(B)**.

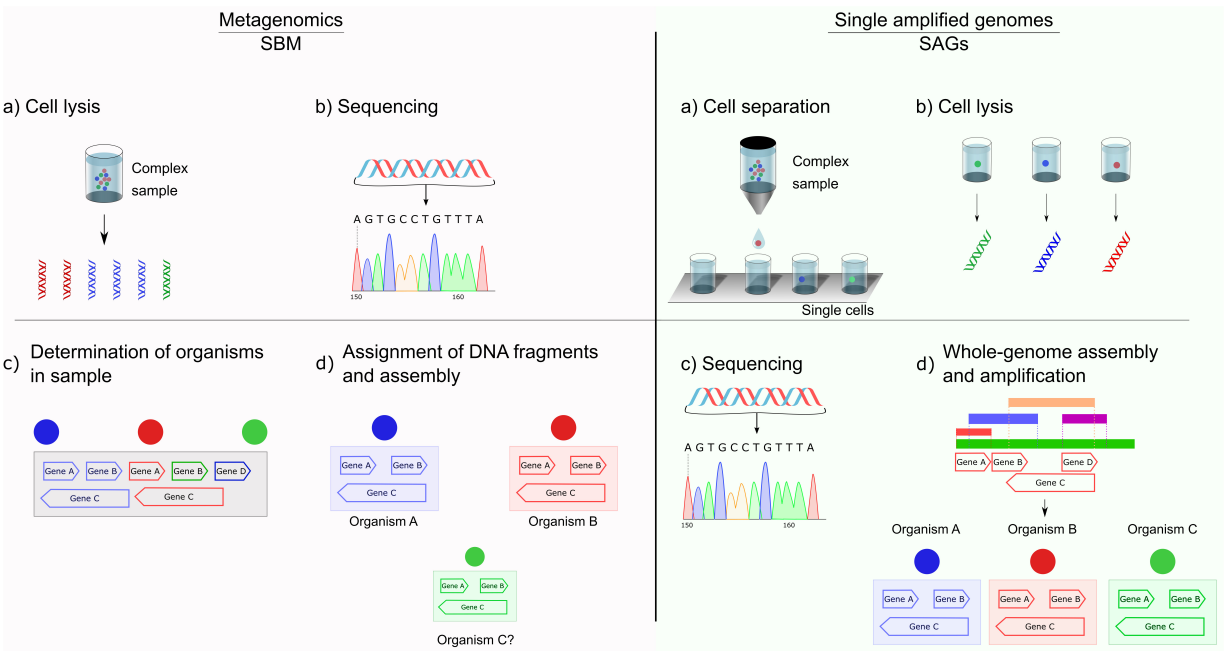


FIGURE 3 | Schematic comparison of sequence-based metagenomics (SBM) and single amplified genomes (SAGs). In contrast to the multitude of sequences obtained in metagenomics, SAG allows the correct assignment of one genome to one sample. Different amounts of diverse species result in an uneven distribution of DNA fragments for metagenomics, drastically increasing complexity in the assignment of DNA fragments to potential genomes and their assembly.

ensure that all reagents and equipment used are decontaminated (Giddings and Newman, 2015).

The Challenge of Gene Function Annotation

Although the accessibility of uncultured species' genomes has increased, computational challenges hinder discovering novel extremozymes using culture-independent methods. These hindrances are not unique to extremophiles and affects all organisms, albeit every group has its own unique hurdles contributing to the major cause. The major challenge in mining uncultured organisms' genomic data is the reliable large-scale annotation of DNA sequences from non-mesophilic organisms with low sequence homology to experimentally described organisms. Annotation by homology faces a significant dilemma: annotation reliability is reciprocally coupled to protein diversity. Therefore, the difference between gene sequences negatively correlates with the overall homology to any related gene. Moreover, studying extremozymes amplifies not only in

culture-based approaches but also culture-independent methods. First reason is linked to the fact that extremophiles tend to alter the overall amino acid composition of enzymes to adapt to their environment, which can be observed in halophiles and thermophiles (Madern et al., 2000; Sterner and Liebl, 2001; Berezovsky and Shakhnovich, 2005; Radestock and Gohlke, 2011; Siglioccolo et al., 2011; Reed et al., 2013; Karan et al., 2020). The second reason is mainly ought to the very limited amount of characterized extremozymes. For example, alcohol dehydrogenases and γ-carbonic anhydrase (CA_D) discovered from uncharacterized archaea collected from brine pool at the bottom of the Red Sea showed sequence homology of about 30–37% to the nearest mesophilic homologs (Grötzinger et al., 2017; Akal et al., 2019; Vogler et al., 2020).

Error propagation further complicates the situation, especially in enzymes with few or highly diverged homologs. The function of the encoded protein can be proven experimentally only for a small and continuously decreasing fraction of gene sequences available from databases, which is currently 0.09% of the UniProt database (**Figure 4A**). Furthermore, more than 94% of all sequences originate from eukaryotes or bacteria (**Figure 4B**). Therefore, the challenge of a reliable annotation is even more pronounced for genes that originate from other organisms, such as archaea, including most extremophiles (Rampelotto, 2013). Hence, the more distinct a protein is, smaller the probability of being related to an experimentally described enzyme. At the same time, the more distinct an enzyme is, the higher is its impact on future annotations. For example, a gene from the newly discovered mesophilic microorganism can be compared to many homologs. The function of the majority of these homologs will be used as annotation of the new gene. In contrast, a gene from a newly discovered extremophilic microorganism might have far fewer homologs and therefore, the annotation of these homologs plays a significant role for future studies. Recent developments in annotation algorithms allow a higher degree of freedom between DNA sequences that will be annotated compared to the closest homologs. These methods include algorithms that focus on specific and short conserved sequences only, such as essential amino acids of the active center along with conserved regions such as catalytic sites and cofactor binding sites (Grötzinger et al., 2014). More elegant but substantially more complicated and computational resource-demanding are structure-based approaches. Here, the expression product of potential genes is modeled, and its function is assessed by its structure, not by the sequence. The biennial Critical Assessment of protein Structure Prediction (CASP) is a competition where their efficiency compares novel modeling algorithms to predict unusual and novel tertiary structures of proteins whose crystal and/or NMR structure was measured but not yet published. The overall *in silico* modeling success was limited until the rise of AI-based protein folding prediction. Last year, in 2020, a new version of a deep machine learning-based algorithm called AlphaFold 2 was released (Senior et al., 2020). This algorithm has substantially increased computational protein folding prediction potential and showed impressive results in the CASP assessment. AlphaFold 2 reached 92.4 out of 100 Global Distance Test (GDT), whereas the average crystal structure only gets 90. Thus, at least for the test samples, it was as good as a crystal structure. Such an impressive result gives hope for the future computational study of unknown proteins. In theory, it could mean that any given protein's structure and function could be predicted solely based on its amino acid sequence in the future. Still, today only a limited amount of algorithms have experimentally been shown to be efficient in the annotation of far distant related genes (Grötzinger et al., 2014; Akal et al., 2019; Vogler et al., 2020). Reliable annotation of the entire genome of an organism that is very distantly related to described organisms is not yet available.

SAGs reduce data complexity significantly compared to that of the metagenomics approach, but the analysis and management of next-generation whole-genome sequencing (NGS) data include

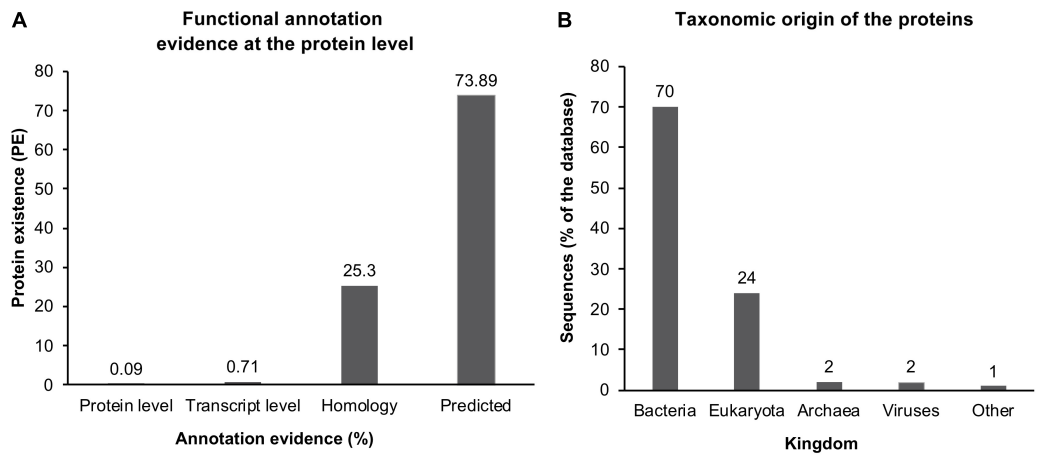


FIGURE 4 | Protein sequence entries in the UniProt database (2020_03 release, https://www.ebi.ac.uk/uniprot/TrEMBLstats). Number of proteins annotated using the listed evidence of their functional annotation prediction from 185 million sequence entries. Only 0.09% of all entries, corresponding to ∼169 thousand entries, show any functional annotation evidence at the protein level **(A)**. Taxonomic origin of the proteins in percent, separated into Kingdoms **(B)** (data from https://www.ebi.ac.uk/uniprot/TrEMBLstats).

an impressive number of various software applications. These applications are used for sequence read assembly, mapping to the reference genome, variant/SNP calling and annotation, transcript assembly and quantification, and small RNA identification (Horner et al., 2010; Garber et al., 2011; Pabinger et al., 2013). Compared to classical genomic sequencing, experimental characterization of SAG gene products requires gene synthesis, expression, purification, and functional characterization and, therefore, is several orders of magnitude more time-consuming and cost-intensive. Hence, false-positive results from flawed annotation are much more problematic than false-negative results (due to incomplete annotation) when genomic data are searched for the desired function, which is particularly true for genes from extremophilic organisms that require slow-growing expression systems (Grötzinger et al., 2014).

Therefore, the initial challenge of cultivating organisms can be addressed to a considerable degree by using novel culture-independent methods. However, these advance shifts the problem toward bioinformatics-based handling and interpretation of the data, where developments are rapidly progressing but still need time and data, to allow reliable, functional annotation of whole genomes. In the **Table 1**, a comprehensive list of extremozymes discovered with metagenomics (Berini et al., 2017) and SAGs are shown.

CONCLUSION/OUTLOOK

In the past decades, DNA sequence analysis has made tremendous progress with advances in sequencing technologies and the rapid development of data analysis algorithms. Through these developments, metagenomic studies of complex microbial communities have become more straightforward and more applicable. Given that most microbes cannot be cultivated in a laboratory, obtaining genome information directly from the environment is a massive step forward for genome profiling of unknown microbial communities. At the time, when read lengths are long enough to allow a confident assembly of genomes, SBM will become the method of choice for bioprospecting non-cultivable microorganisms. For now, the implementation of single-cell genomics has facilitated recovering proteins from uncultured microorganism, helped us understand the species diversity, and how enzymes adapt to harsh environments. Recent advances in developing new sequencing technologies with longer reads, higher throughput, and better cell sorting methods have certainly enhanced SCG as a tool for studying uncultivable microbes. Consequently, this relatively cost-efficient technology is essential to broaden the knowledge of extremophiles and extremozymes, which are very distantly related to the majority of described organisms. We believe that single-cell technologies will not only shed further light on microbial "dark matter" but also facilitate the development of *in silico*-designed and fine-tailored biocatalysts for specific reactions, thus bringing us closer to a sustainable future.

AUTHOR CONTRIBUTIONS

All authors contributed to and commented on this manuscript.

FUNDING

The research reported in this publication was supported by funding from the King Abdullah University of Science and Technology (KAUST).

ACKNOWLEDGMENTS

We are thankful to research interns Mayar A. Elbaz and Alvin Huang for their valuable inputs. **Figure 1** was created using resources from svgrepo.com and Flaticon.com.

REFERENCES

Abdel-Hamed, A. R., Abo-Elmatty, D. M., Wiegel, J., and Mesbah, N. M. (2016). Biochemical characterization of a halophilic, alkalithermophilic protease from *Alkalibacillus* sp. NM-Da2. *Extremophiles* 20, 885–894. doi: 10.1007/s00792-016-0879-x

Aitken, D. M., and Brown, A. D. (1969). Citrate and glyoxylate cycles in the halophil, Halobacterium salinarium. *Biochim. Biophys. Acta* 177, 351–354. doi: 10.1016/0304-4165(69)90148-2

Akal, A. L., Karan, R., Hohl, A., Alam, I., Vogler, M., Grötzinger, S. W., et al. (2019). A polyextremophilic alcohol dehydrogenase from the Atlantis II deep red sea brine pool. *FEBS Open Bio* 9, 194–205. doi: 10.1002/2211-5463.12557

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579

Al-Ghanayem, A. A., and Joseph, B. (2020). Current prospective in using cold-active enzymes as eco-friendly detergent additive. *Appl. Microbiol. Biotechnol.* 104, 2871–2882. doi: 10.1007/s00253-020-10429-x

Ali, I., Akbar, A., Yanwisetpakdee, B., Prasongsuk, S., Lotrakul, P., and Punnapayak, H. (2014). Purification, characterization, and potential of saline waste water remediation of a polyextremophilic α-amylase from an obligate halophilic *Aspergillus gracilis*. *BioMed Res. Int.* 2014, 106937.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. *Microbiol. Rev.* 59, 143–169. doi: 10.1128/mmbr.59.1.143-169.1995

Anastas, P., and Eghbali, N. (2010). Green chemistry: principles and practice. *Chem. Soc. Rev.* 39, 301–312. doi: 10.1039/b918763b

Anschutz, P., and Blanc, G. (1995). Geochemical dynamics of the Atlantis II Deep (Red Sea): silica behavior. *Mar. Geol.* 128, 25–36. doi: 10.1016/0025-3227(95)00085-d

Antunes, A., Ngugi, D. K., and Stingl, U. (2011). Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* 3, 416–433. doi: 10.1111/j.1758-2229.2011.00264.x

Atalah, J., Cáceres-Moreno, P., Espina, G., and Blamey, J. M. (2019). Thermophiles and the applications of their enzymes as new biocatalysts. *Bioresour. Technol.* 280, 478–488. doi: 10.1016/j.biortech.2019.02.008

Attrill, H., Gaudet, P., Huntley, R. P., Lovering, R. C., Engel, S. R., Poux, S., et al. (2019). Annotation of gene product function from high-throughput studies using the gene ontology. *Database* 2019, baz007.

Baker-Austin, C., and Dopson, M. (2007). Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* 15, 165–171. doi: 10.1016/j.tim.2007.02.005

Baliga, N. S., Goo, Y. A., Ng, W. V., Hood, L., Daniels, C. J., and DasSarma, S. (2000). Is gene expression in *Halobacterium* NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol. Microbiol.* 36, 1184–1185. doi: 10.1046/j.1365-2958.2000.01916.x

Berezovsky, I. N., and Shakhnovich, E. I. (2005). Physics and evolution of thermophilic adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12742–12747. doi: 10.1073/pnas.0503890102

Berini, F., Casciello, C., Marcone, G. L., and Marinelli, F. (2017). Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* 364, fnx211. doi: 10.1093/femsle/fnx211

Blainey, P. C. (2013). The future is now: Single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* 37, 407–427. doi: 10.1111/1574-6976.12015

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., et al. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29, 358–367. doi: 10.1016/j.tree.2014.04.003

Brock, T. D. (1997). The value of basic research: discovery of *Thermus aquaticus* and other extreme thermophiles. *Genetics* 146, 1207–1210. doi: 10.1093/genetics/146.4.1207

Bruins, M. E., Janssen, A. E., and Boom, R. M. (2001). Thermozymes and their applications: a review of recent literature and patents. *Appl. Biochem. Biotechnol.* 90, 155–186. doi: 10.1385/abab:90:2:155

Buerger, S., Spoering, A., Gavrish, E., Leslin, C., Ling, L., and Epstein, S. S. (2012). Microbial scout hypothesis and microbial discovery. *Appl. Environ. Microbiol.* 78, 3229–3233. doi: 10.1128/aem.07308-11

Carraro, L., Maifreni, M., Bartolomeoli, I., Martino, M. E., Novelli, E., Frigo, F., et al. (2011). Comparison of culture-dependent and-independent methods for bacterial community monitoring during Montasio cheese manufacturing. *Res. Microbiol.* 162, 231–239. doi: 10.1016/j.resmic.2011.01.002

Carrea, G., and Riva, S. (2000). Properties and synthetic applications of enzymes in organic solvents. *Angew. Chem. Int. Ed. Engl.* 39, 2226–2254. doi: 10.1002/1521-3773(20000703)39:13<2226::aid-anie2226>3.0.co;2-l

Cavicchioli, R., Charlton, T., Ertan, H., Omar, S. M., Siddiqui, K., and Williams, T. (2011). Biotechnological uses of enzymes from psychrophiles. *Microb. Biotechnol.* 4, 449–460. doi: 10.1111/j.1751-7915.2011.00258.x

Charlesworth, J. C., and Burns, B. P. (2015). Untapped resources: biotechnological potential of peptides and secondary metabolites in archaea. *Archaea* 2015, 282035. doi: 10.1155/2015/282035

Chen, G.-Q., and Jiang, X.-R. (2018). Next generation industrial biotechnology based on extremophilic bacteria. *Curr. Opin. Biotechnol.* 50, 94–100. doi: 10.1016/j.copbio.2017.11.016

Chen, G. Q., and Patel, M. K. (2012). Plastics derived from biological sources: present and future: a technical and environmental review. *Chem. Rev.* 112, 2082–2099. doi: 10.1021/cr200162d

Chi, K. R. (2014). Singled out for sequencing. *Nat. Methods* 11, 13–17. doi: 10.1038/nmeth.2768

Chinnathambi, A. (2015). Industrial important enzymes from alkaliphiles–an overview. *Biosci. Biotechnol. Res. Asia* 12, 2007–2016. doi: 10.13005/bbra/1868

Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.* 32, 1351–1359. doi: 10.1007/s10529-010-0306-9

Choi, Y. J., Hur, J. M., Lim, S., Jo, M., Kim, D. H., and Choi, J. I. (2014). Induction of apoptosis by deinoxanthin in human cancer cells. *Anticancer Res.* 34, 1829–1835.

Coker, J. A. (2016). Extremophiles and biotechnology: current uses and prospects. *F1000Res.* 5:F1000 Faculty Rev-396. doi: 10.12688/f1000research.7432.1

Damiano, V., Bocchini, D., Gomes, E., and Da Silva, R. (2003). Application of crude xylanase from *Bacillus licheniformis* 77-2 to the bleaching of eucalyptus Kraft pulp. *World J. Microbiol. Biotechnol.* 19, 139–144.

DasSarma, S., Capes, M. D., Karan, R., and DasSarma, P. (2013). Amino acid substitutions in cold-adapted proteins from *Halorubrum lacusprofundi*, an extremely halophilic microbe from Antarctica. *PLoS One* 8:e58587. doi: 10.1371/journal.pone.0058587

DasSarma, S., and DasSarma, P. (2015). Gas vesicle nanoparticles for antigen display. *Vaccines* 3, 686–702. doi: 10.3390/vaccines3030686

Davey, H. M., and Kell, D. B. (1996). Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiol. Rev.* 60, 641–696. doi: 10.1128/mr.60.4.641-696.1996

de la Coba, F., Aguilera, J., de Galvez, M. V., Alvarez, M., Gallego, E., Figueroa, F. L., et al. (2009). Prevention of the ultraviolet effects on clinical and histopathological changes, as well as the heat shock protein-70 expression in mouse skin by topical application of algal UV-absorbing compounds. *J. Dermatol. Sci.* 55, 161–169. doi: 10.1016/j.jdermsci.2009.06.004

de Lourdes Moreno, M., Pérez, D., García, M. T., and Mellado, E. (2013). Halophilic bacteria as a source of novel hydrolytic enzymes. *Life* 3, 38–51. doi: 10.3390/life3010038

De Santi, C., Altermark, B., Pierechod, M. M., Ambrosino, L., De Pascale, D., and Willassen, N. P. (2016a). Characterization of a cold-active and salt tolerant esterase identified by functional screening of Arctic metagenomic libraries. *BMC Biochem.* 17:1. doi: 10.1186/s12858-016-0057-x

De Santi, C., Ambrosino, L., Tedesco, P., de Pascale, D., Zhai, L., Zhou, C., et al. (2015). Identification and characterization of a novel salt-tolerant esterase from a *Tibetan glacier* metagenomic library. *Biotechnol. Prog.* 31, 890–899. doi: 10.1002/btpr.2096

De Santi, C., Tedesco, P., Ambrosino, L., Altermark, B., Willassen, N.-P., and de Pascale, D. (2014). A new alkaliphilic cold-active esterase from the psychrophilic marine bacterium *Rhodococcus* sp.: functional and structural studies and biotechnological potential. *Appl. Biochem. Biotechnol.* 172, 3054–3068. doi: 10.1007/s12010-013-0713-1

De Santi, C., Willassen, N. P., and Williamson, A. (2016b). Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine Arctic metagenome. *PLoS One* 11:e0159345. doi: 10.1371/journal.pone.0159345

Dhaulaniya, A. S., Balan, B., Agrawal, P. K., and Singh, D. K. (2019). Cold survival strategies for bacteria, recent advancement and potential industrial applications. *Arch. Microbiol.* 201, 1–16. doi: 10.1007/s00203-018-1602-3

Dopson, M., Ni, G., and Sleutels, T. H. (2015). Possibilities for extremophilic microorganisms in microbial electrochemical systems. *FEMS Microbiol. Rev.* 40, 164–181. doi: 10.1093/femsre/fuv044

Dumorne, K., Córdova, D. C., Astorga-Elo, M., and Renganathan, P. (2017). Extremozymes: a potential source for industrial applications. *J. Microbiol. Biotechnol.* 27, 649–659. doi: 10.4014/jmb.1611.11006

Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M. J., Alexander Richter, R., Valas, R., et al. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6, 1186–1199. doi: 10.1038/ismej.2011.189

Elleuche, S., Schroder, C., Sahm, K., and Antranikian, G. (2014). Extremozymes–biocatalysts with unique properties from extremophilic microorganisms. *Curr. Opin. Biotechnol.* 29, 116–123. doi: 10.1016/j.copbio.2014.04.003

Emery, K. O., Hunt, J. M., and Hays, E. E. (1969). "Summary of hot brines and heavy metal deposits in the red sea," in *Hot Brines and Recent Heavy Metal Deposits in the Red Sea*, eds E. Degens and D. Ross (Berlin: Springer), 557–571. doi: 10.1007/978-3-662-28603-6_50

Fan, J., Han, F., and Liu, H. (2014). Challenges of big data analysis. *Natl. Sci. Rev.* 1, 293–314. doi: 10.1093/nsr/nwt032

Ferrandi, E. E., Previdi, A., Bassanini, I., Riva, S., Peng, X., and Monti, D. (2017). Novel thermostable amine transferases from hot spring metagenomes. *Appl. Microbiol. Biotechnol.* 101, 4963–4979. doi: 10.1007/s00253-017-8228-2

Ferrandi, E. E., Sayer, C., Isupov, M. N., Annovazzi, C., Marchesi, C., Iacobone, G., et al. (2015). Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. *FEBS J.* 282, 2879–2894. doi: 10.1111/febs.13328

Ferrer, M., Golyshina, O., Beloqui, A., and Golyshin, P. N. (2007). Mining enzymes from extreme environments. *Curr. Opin. Microbiol.* 10, 207–214. doi: 10.1016/j.mib.2007.05.004

Frumkin, D., Wasserstrom, A., Itzkovitz, S., Harmelin, A., Rechavi, G., and Shapiro, E. (2008). Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnol.* 8:17. doi: 10.1186/1472-6750-8-17

Fu, L., He, Y., Xu, F., Ma, Q., Wang, F., and Xu, J. (2015). Characterization of a novel thermostable patatin-like protein from a Guaymas basin metagenomic library. *Extremophiles* 19, 829–840. doi: 10.1007/s00792-015-0758-x

Fuchsman, C. A., Collins, R. E., Rocap, G., and Brazelton, W. J. (2017). Effect of the environment on horizontal gene transfer between bacteria and archaea. *PeerJ* 5, e3865–e3865. doi: 10.7717/peerj.3865

Gabani, P., and Singh, O. V. (2013). Radiation-resistant extremophiles and their potential in biotechnology and therapeutics. *Appl. Microbiol. Biotechnol.* 97, 993–1004. doi: 10.1007/s00253-012-4642-7

Galinski, E. A., Pfeiffer, H. P., and Truper, H. G. (1985). 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus Ectothiorhodospira. *Eur. J. Biochem.* 149, 135–139. doi: 10.1111/j.1432-1033.1985.tb08903.x

Garber, M., Grabherr, M. G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–477. doi: 10.1038/nmeth.1613

Gatti-Lafranconi, P., Natalello, A., Rehm, S., Doglia, S. M., Pleiss, J., and Lotti, M. (2010). Evolution of stability in a cold-active enzyme elicits specificity relaxation and highlights substrate-related effects on temperature adaptation. *J. Mol. Biol.* 395, 155–166. doi: 10.1016/j.jmb.2009.10.026

Gerday, C., and Glansdorff, N. (2009). *EXTREMOPHILES*, Vol. I. Paris: Eolss Publishers.

Giddings, L.-A., and Newman, D. J. (2015). "Bioactive compounds from extremophiles," in *Bioactive Compounds from Extremophiles: Genomic Studies, Biosynthetic Gene Clusters, and New Dereplication Methods*, eds L.-A. Giddings and D. J. Newman (Cham: Springer International Publishing), 1–47. doi: 10.1007/978-3-319-14836-6_1

Gilbert, J. A., Jansson, J. K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biol.* 12:69. doi: 10.1186/s12915-014-0069-1

Grand View Research (2020). *Enzymes Market Size, Share, Global Industry Report, 2020-2027*. Available: https://www.grandviewresearch.com/industry-analysis/enzymes-industry (accessed January 2, 2021).

Grötzinger, S., Alam, I., Alawi, W., Bajic, V., Stingl, U., and Eppinger, J. (2014). Mining a database of single amplified genomes from Red Sea brine pool extremophiles—improving reliability of gene function prediction using a profile and pattern matching algorithm (PPMA). *Front. Microbiol.* 5:134. doi: 10.3389/fmicb.2014.00134

Grötzinger, S. W., Karan, R., Strillinger, E., Bader, S., Frank, A., Al Rowaihi, I. S., et al. (2017). Identification and experimental characterization of an extremophilic brine pool alcohol dehydrogenase from single amplified genomes. *ACS Chem. Biol.* 13, 161–170. doi: 10.1021/acschembio.7b00792

Hedlund, B. P., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter". *Extremophiles* 18, 865–875. doi: 10.1007/s00792-014-0664-7

Herbert, R. A. (1992). A perspective on the biotechnological potential of extremophiles. *Trends Biotechnol.* 10, 395–402. doi: 10.1016/0167-7799(92)90282-z

Horikoshi, K. (2016). "Alkaliphiles," in *Extremophiles: Where It All Began* (Tokyo: Springer Japan), 53–78.

Horner, D. S., Pavesi, G., Castrignanò, T., De Meo, P. D. O., Liuni, S., Sammeth, M., et al. (2010). Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief. Bioinform.* 11, 181–197. doi: 10.1093/bib/bbp046

Hua, Z.-S., Han, Y.-J., Chen, L.-X., Liu, J., Hu, M., Li, S.-J., et al. (2015). Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* 9, 1280–1294. doi: 10.1038/ismej.2014.212

Hugenholtz, P., and Kyrpides, N. C. (2009). A changing of the guard. *Environ. Microbiol.* 11, 551–553. doi: 10.1111/j.1462-2920.2009.01888.x

Hugenholtz, P., and Tyson, G. W. (2008). Metagenomics. *Nature* 455, 481–483.

Illanes, A., Cauerhff, A., Wilson, L., and Castro, G. R. (2012). Recent trends in biocatalysis engineering. *Bioresour. Technol.* 115, 48–57. doi: 10.1016/j.biortech.2011.12.050

Irwin, J. A. (2010). Extremophiles and their application to veterinary medicine. *Environ. Technol.* 31, 857–869. doi: 10.1080/09593330.2010.484073

Irwin, J. A., and Baird, A. W. (2004). Extremophiles and their application to veterinary medicine. *Ir. Vet. J.* 57, 348–354. doi: 10.1186/2046-0481-57-6-348

Jin, M., Gai, Y., Guo, X., Hou, Y., and Zeng, R. (2019). Properties and applications of extremozymes from deep-sea extremophilic microorganisms: a mini review. *Mar. Drugs* 17, 656. doi: 10.3390/md17120656

Johnson, D. B. (2014). Biomining-biotechnologies for extracting and recovering metals from ores and waste materials. *Curr. Opin. Biotechnol.* 30, 24–31. doi: 10.1016/j.copbio.2014.04.008

Jorquera, M. A., Graether, S. P., and Maruyama, F. (2019). Bioprospecting and biotechnology of extremophiles. *Front. Bioeng. Biotechnol.* 7:204. doi: 10.3389/fbioe.2019.00204

Kaeberlein, T., Lewis, K., and Epstein, S. S. (2002). Isolating "Uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296, 1127–1129. doi: 10.1126/science.1070633

Kamekura, M., Hamakawa, T., and Onishi, H. (1982). Application of halophilic nuclease H of *Micrococcus* varians subsp. halophilus to commercial production of flavoring agent 5'-GMP. *Appl. Environ. Microbiol.* 44, 994–995. doi: 10.1128/aem.44.4.994-995.1982

Karan, R., Capes, M. D., DasSarma, P., and DasSarma, S. (2013). Cloning, overexpression, purification, and characterization of a polyextremophilic β-galactosidase from the Antarctic haloarchaeon *Halorubrum lacusprofundi*. *BMC Biotechnol.* 13:3. doi: 10.1186/1472-6750-13-3

Karan, R., Capes, M. D., and DasSarma, S. (2012a). Function and biotechnology of extremophilic enzymes in low water activity. *Aquat. Biosyst.* 8, 4. doi: 10.1186/2046-9063-8-4

Karan, R., Grötzinger, S., Vogler, M., Akal, A., Renn, D., Vancea, A., et al. (2019). Single amplified genomes from Red Sea brine pool extremophiles-a source of novel enzymes. *Stud. Univ. Babes Bolyai Biol.* 64, 76.

Karan, R., and Khare, S. (2010). Purification and characterization of a solvent−stable protease from *Geomicrobium* sp. EMB2. *Environ. Technol.* 31, 1061–1072. doi: 10.1080/09593331003674556

Karan, R., and Khare, S. (2011). Stability of haloalkaliphilic *Geomicrobium* sp. protease modulated by salt. *Biochemistry (Moscow)* 76, 686. doi: 10.1134/s0006297911060095

Karan, R., Kumar, S., Sinha, R., and Khare, S. (2012b). "Halophilic microorganisms as sources of novel enzymes," in *Microorganisms in Sustainable Agriculture and Biotechnology*, eds T. Satyanarayana, B. Johri, and A. Prakash (Dordrecht: Springer), 555–579. doi: 10.1007/978-94-007-2214-9_25

Karan, R., Mathew, S., Muhammad, R., Bautista, D. B., Vogler, M., Eppinger, J., et al. (2020). Understanding high-salt and cold adaptation of a polyextremophilic enzyme. *Microorganisms* 8, 1594. doi: 10.3390/microorganisms8101594

Karan, R., Singh, S. P., Kapoor, S., and Khare, S. K. (2011). A novel organic solvent tolerant protease from a newly isolated *Geomicrobium* sp. EMB2 (MTCC 10310): production optimization by response surface methodology. *N. Biotechnol.* 28, 136–145. doi: 10.1016/j.nbt.2010.10.007

Kirk, R. G., and Ginzburg, M. (1972). Ultrastructure of two species of *Halobacterium*. *J. Ultrastruct. Res.* 41, 80–94. doi: 10.1016/s0022-5320(72)90040-8

Kristjánsson, J. K., and Hreggvidsson, G. O. (1995). Ecology and habitats of extremophiles. *World J. Microbiol. Biotechnol.* 11, 17–25. doi: 10.1007/BF00339134

Krüger, A., Schäfers, C., Schröder, C., and Antranikian, G. (2018). Towards a sustainable biobased industry – highlighting the impact of extremophiles. *N. Biotechnol.* 40, 144–153. doi: 10.1016/j.nbt.2017.05.002

Kumar, S., Grewal, J., Sadaf, A., Hemamalini, R., and Khare, S. K. (2016). Halophiles as a source of polyextremophilic α-amylase for industrial applications. *AIMS Microbiol.* 2, 1–26. doi: 10.3934/microbiol.2016.1.1

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., and Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* 72, 557–578.

Kvist, T., Ahring, B. K., Lasken, R. S., and Westermann, P. (2007). Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* 74, 926–935. doi: 10.1007/s00253-006-0725-7

Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., et al. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18, 1185–1193. doi: 10.1111/1469-0691.12023

Lagier, J.-C., Khelaifia, S., Alou, M. T., Ndongo, S., Dione, N., Hugon, P., et al. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* 1, 1–8. doi: 10.1007/978-1-4614-6418-1_540-3

Lamers, P., Searcy, E., Hess, J. R., and Stichnothe, H. (2016). *Developing the Global Bioeconomy: Technical, Market, and Environmental Lessons from Bioenergy*. Cambridge, MA: Academic Press.

Landry, Z. C., Giovanonni, S. J., Quake, S. R., and Blainey, P. C. (2013). Optofluidic cell selection from complex microbial communities for single-genome analysis. *Methods Enzymol.* 531, 61–90. doi: 10.1016/b978-0-12-407863-5.00004-6

Laurence, M., Hatzis, C., and Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One* 9:e97876. doi: 10.1371/journal.pone.0097876

Lebre, P. H., De Maayer, P., and Cowan, D. A. (2017). Xerotolerant bacteria: surviving through a dry spell. *Nat. Rev. Microbiol.* 15, 285–296. doi: 10.1038/nrmicro.2017.16

Leis, B., Angelov, A., Mientus, M., Li, H., Pham, V. T. T., Lauinger, B., et al. (2015).). Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front. Microbiol.* 6:275. doi: 10.3389/fmicb.2015.00275

Li, J., Yun, Y.-q, Xing, L., and Song, L. (2017). Novel bioflocculant produced by salt-tolerant, alkaliphilic strain *Oceanobacillus polygoni* HG6 and its application in tannery wastewater treatment. *Biosci. Biotechnol. Biochem.* 81, 1018–1025. doi: 10.1080/09168451.2016.1274635

Liu, Z., Zhao, C., Deng, Y., Huang, Y., and Liu, B. (2015). Characterization of a thermostable recombinant β-galactosidase from a thermophilic anaerobic bacterial consortium YTY-70. *Biotechnol. Biotechnol. Equip.* 29, 547–554. doi: 10.1080/13102818.2015.1015244

Locey, K. J., and Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5970–5975. doi: 10.1073/pnas.1521291113

López-López, O., Knapik, K., Cerdán, M. E., and González-Siso, M. I. (2015). Metagenomics of an alkaline hot spring in Galicia (Spain): microbial diversity analysis and screening for novel lipolytic enzymes. *Front. Microbiol.* 6:1291. doi: 10.3389/fmicb.2015.01291

Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles* 4, 91–98. doi: 10.1007/s007920050142

Marcy, Y., Ouverney, C., Bik, E. M., Losekann, T., Ivanova, N., Martin, H. G., et al. (2007). Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11889–11894. doi: 10.1073/pnas.0704662104

Margesin, R., and Feller, G. (2010). Biotechnological applications of psychrophiles. *Environ. Technol.* 31, 835–844. doi: 10.1080/09593331003663328

Martin, M., and Vandenbol, M. (2016). The hunt for original microbial enzymes: an initiatory review on the construction and functional screening of (meta) genomic libraries. *Biotechnol. Agron. Soc. Environ.* 20, 523–532.

McLean, J. S., Lombardo, M. J., Badger, J. H., Edlund, A., Novotny, M., Yee-Greenbaum, J., et al. (2013). Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2390–E2399. doi: 10.1073/pnas.1219809110

Mukhopadhyay, A., Dasgupta, A. K., and Chakrabarti, K. (2015). Enhanced functionality and stabilization of a cold active laccase using nanotechnology based activation-immobilization. *Bioresour. Technol.* 179, 573–584. doi: 10.1016/j.biortech.2014.12.070

Müller, S., and Nebe-Von-Caron, G. (2010). Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. *FEMS Microbiol. Rev.* 34, 554–587. doi: 10.1111/j.1574-6976.2010.00214.x

Nakamura, K., Iizuka, R., Nishi, S., Yoshida, T., Hatada, Y., Takaki, Y., et al. (2016). Culture-independent method for identification of microbial enzyme-encoding genes by activity-based single-cell sequencing using a water-in-oil microdroplet platform. *Sci. Rep.* 6, 22259. doi: 10.1038/srep22259

Nowrotek, M., Jałowiecki, Ł, Harnisz, M., and Płaza, G. A. (2019). Culturomics and metagenomics: in understanding of environmental resistome. *Front. Environ. Sci. Eng.* 13:40. doi: 10.1007/s11783-019-1121-8

Oren, A. (2010). Industrial and environmental applications of halophilic microorganisms. *Environ. Technol.* 31, 825–834. doi: 10.1080/09593330903370026

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., et al. (2013). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* 15, 256–278.

Petrovskaya, L. E., Novototskaya-Vlasova, K. A., Spirina, E. V., Durdenko, E. V., Lomakina, G. Y., Zavialova, M. G., et al. (2016). Expression and characterization of a new esterase with GCSAG motif from a permafrost metagenomic library. *FEMS Microbiol. Ecol.* 92, fiw046. doi: 10.1093/femsec/fiw046

Placido, A., Hai, T., Ferrer, M., Chernikova, T. N., Distaso, M., Armstrong, D., et al. (2015). Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (*Aeolian archipelago*) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Appl. Microbiol. Biotechnol.* 99, 10031–10046. doi: 10.1007/s00253-015-6873-x

Quillaguaman, J., Guzman, H., Van-Thuoc, D., and Hatti-Kaul, R. (2010). Synthesis and production of polyhydroxyalkanoates by halophiles: current potential and future prospects. *Appl. Microbiol. Biotechnol.* 85, 1687–1696. doi: 10.1007/s00253-009-2397-6

Raddadi, N., Cherif, A., Daffonchio, D., Neifar, M., and Fava, F. (2015). Biotechnological applications of extremophiles, extremozymes and extremolytes. *Appl. Microbiol. Biotechnol.* 99, 7907–7913. doi: 10.1007/s00253-015-6874-9

Radestock, S., and Gohlke, H. (2011). Protein rigidity and thermophilic adaptation. *Proteins* 79, 1089–1108. doi: 10.1002/prot.22946

Raghunathan, A., Ferguson, H. R. Jr., Bornarth, C. J., Song, W., Driscoll, M., and Lasken, R. S. (2005). Genomic DNA amplification from a single bacterium. *Appl. Environ. Microbiol.* 71, 3342–3347. doi: 10.1128/AEM.71.6.3342-3347.2005

Ramesh, A., Harani Devi, P., Chattopadhyay, S., and Kavitha, M. (2020). "Commercial applications of microbial enzymes," in *Microbial Enzymes: Roles and Applications in Industries*, eds N. K. Arora, J. Mishra, and V. Mishra (Singapore: Springer Singapore), 137–184. doi: 10.1007/978-981-15-1710-5_6

Rampelotto, P. H. (2013). Extremophiles and extreme environments. *Life (Basel)* 3, 482–485. doi: 10.3390/life3030482

Rashid, M., and Stingl, U. (2015). Contemporary molecular tools in microbial ecology and their application to advancing biotechnology. *Biotechnol. Adv.* 33, 1755–1773. doi: 10.1016/j.biotechadv.2015.09.005

Rasor, J. P., and Voss, E. (2001). Enzyme-catalyzed processes in pharmaceutical industry. *Appl. Catal. Gen.* 221, 145–158. doi: 10.1016/S0926-860x(01)00804-3

Reed, C. J., Lewis, H., Trejo, E., Winston, V., and Evilia, C. (2013). Protein adaptations in archaeal extremophiles. *Archaea* 2013, 373275. doi: 10.1155/2013/373275

Rego, A., Raio, F., Martins, T. P., Ribeiro, H., Sousa, A. G., Séneca, J., et al. (2019). Actinobacteria and cyanobacteria diversity in terrestrial antarctic microenvironments evaluated by culture-dependent and independent methods. *Front. Microbiol.* 10:1018. doi: 10.3389/fmicb.2019.01018

Rekadwad, B., and Khobragade, C. (2017). "Marine polyextremophiles and their biotechnological applications," in *Microbial Applications Vol.1: Bioremediation and Bioenergy*, eds V. C. Kalia and P. Kumar (Cham: Springer International Publishing), 319–331. doi: 10.1007/978-3-319-52666-9_15

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi: 10.1038/nprot.2014.067

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352

Rizk, M., Antranikian, G., and Elleuche, S. (2012). End-to-end gene fusions and their impact on the production of multifunctional biomass degrading enzymes. *Biochem. Biophys. Res. Commun.* 428, 1–5. doi: 10.1016/j.bbrc.2012.09.142

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays Biochem.* 59, 1–41. doi: 10.1042/bse0590001

Ross, D. A. (1972). Red Sea hot brine area: revisited. *Science* 175, 1455–1457. doi: 10.1126/science.175.4029.1455

Rothschild, L. J., and Mancinelli, R. L. (2001). Life in extreme environments. *Nature* 409, 1092–1101.

Sahoo, R. K., Kumar, M., Sukla, L. B., and Subudhi, E. (2017). Bioprospecting hot spring metagenome: lipase for the production of biodiesel. *Environ. Sci. Pollut. Res.* 24, 3802–3809. doi: 10.1007/s11356-016-8118-7

Sarhan, M. S., Hamza, M. A., Youssef, H. H., Patz, S., Becker, M., ElSawey, H., et al. (2019). Culturomics of the plant prokaryotic microbiome and the dawn of plant-based culture media–A review. *J. Adv. Res.* 19, 15–27. doi: 10.1016/j.jare.2019.04.002

Sarmiento, F., Peralta, R., and Blamey, J. M. (2015). Cold and hot extremozymes: industrial relevance and current trends. *Front. Bioeng. Biotechnol.* 3:148. doi: 10.3389/fbioe.2015.00148

Sayed, A., Ghazy, M. A., Ferreira, A. J. S., Setubal, J. C., Chambergo, F. S., Ouf, A., et al. (2014). A novel mercuric reductase from the unique deep brine environment of atlantis II in the red sea. *J. Biol. Chem.* 289, 1675–1687. doi: 10.1074/jbc.M113.493429

Schäfer, T., Borchert, T. W., Nielsen, V. S., Skagerlind, P., Gibson, K., Wenger, K., et al. (2007). "Industrial enzymes," in *White Biotechnology*, eds R. Ulber and D. Sell (Berlin: Springer), 59–131.

Schiraldi, C., Giuliano, M., and De Rosa, M. (2002). Perspectives on biotechnological applications of archaea. *Archaea* 1, 75–86. doi: 10.1155/2002/436561

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288. doi: 10.1371/journal.pone.0017288

Schröder, C., Elleuche, S., Blank, S., and Antranikian, G. (2014). Characterization of a heat-active archaeal β-glucosidase from a hydrothermal spring metagenome. *Enzyme Microb. Technol.* 57, 48–54. doi: 10.1016/j.enzmictec.2014.01.010

Sellek, G. A., and Chaudhuri, J. B. (1999). Biocatalysis in organic media using enzymes from extremophiles. *Enzyme Microb. Technol.* 25, 471–482. doi: 10.1016/S0141-0229(99)00075-7

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710. doi: 10.1038/s41586-019-1923-7

Shapiro, M. G., Ramirez, R. M., Sperling, L. J., Sun, G., Sun, J., Pines, A., et al. (2014). Genetically encoded reporters for hyperpolarized xenon magnetic resonance imaging. *Nat. Chem.* 6, 629. doi: 10.1038/nchem.1934

Sharma, A., Kawarabayasi, Y., and Satyanarayana, T. (2012). Acidophilic bacteria and archaea: acid stable biocatalysts and their potential applications. *Extremophiles* 16, 1–19. doi: 10.1007/s00792-011-0402-3

Sheldon, R. A. (2016). Green chemistry and resource efficiency: towards a green economy. *Green Chem.* 18, 3180–3183. doi: 10.1039/c6gc90040b

Siddiqui, K. S., Parkin, D. M., Curmi, P. M., De Francisci, D., Poljak, A., Barrow, K., et al. (2009). A novel approach for enhancing the catalytic efficiency of a protease at low temperature: reduction in substrate inhibition by chemical modification. *Biotechnol. Bioeng.* 103, 676–686. doi: 10.1002/bit.22300

Siglioccolo, A., Paiardini, A., Piscitelli, M., and Pascarella, S. (2011). Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct. Biol.* 11:50. doi: 10.1186/1472-6807-11-50

Singh, O. V., and Gabani, P. (2011). Extremophiles: radiation resistance microbial reserves and therapeutic implications. *J. Appl. Microbiol.* 110, 851–861. doi: 10.1111/j.1365-2672.2011.04971.x

Singh, R., Chopra, C., Gupta, V. K., Akhlaq, B., Verma, V., and Rasool, S. (2015). Purification and characterization of CHpro1, a thermotolerant, alkali-stable and oxidation-resisting protease of Chumathang hotspring. *Sci. Bull.* 60, 1252–1260. doi: 10.1007/s11434-015-0834-8

Singh, R., Kumar, M., Mittal, A., and Mehta, P. K. (2016). Microbial enzymes: industrial progress in 21st century. *3 Biotech* 6, 174.

Speda, J., Jonsson, B.-H., Carlsson, U., and Karlsson, M. (2017). Metaproteomics-guided selection of targeted enzymes for bioprospecting of mixed microbial communities. *Biotechnol. Biofuels* 10, 128.

Stefani, F. O., Bell, T. H., Marchand, C., de la Providencia, I. E., El Yassimi, A., St-Arnaud, M., et al. (2015). Culture-dependent and-independent methods capture different microbial community fractions in hydrocarbon-contaminated soils. *PLoS One* 10:e0128272. doi: 10.1371/journal.pone.0128272

Stepankova, V., Bidmanova, S., Koudelakova, T., Prokop, Z., Chaloupkova, R., and Damborsky, J. (2013). Strategies for stabilization of enzymes in organic solvents. *ACS Catal.* 3, 2823–2836. doi: 10.1021/cs400684x

Sterner, R.h, and Liebl, W. (2001). Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* 36, 39–106. doi: 10.1080/20014091074174

Steven, B., Briggs, G., McKay, C. P., Pollard, W. H., Greer, C. W., and Whyte, L. G. (2007). Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiol. Ecol.* 59, 513–523. doi: 10.1111/j.1574-6941.2006.00247.x

Stierle, A. A., Stierle, D. B., and Kelly, K. (2006). Berkelic acid, a novel spiroketal with selective anticancer activity from an acid mine waste fungal extremophile. *J. Org. Chem.* 71, 5357–5360. doi: 10.1021/jo060018d

Stuart, E. S., Morshed, F., Sremac, M., and DasSarma, S. (2001). Antigen presentation using novel particulate organelles from halophilic archaea. *J. Biotechnol.* 88, 119–128. doi: 10.1016/s0168-1656(01)00267-x

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., et al. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333, 1296–1300. doi: 10.1126/science.1203690

Sydlik, U., Gallitz, I., Albrecht, C., Abel, J., Krutmann, J., and Unfried, K. (2009). The compatible solute ectoine protects against nanoparticle-induced neutrophilic lung inflammation. *Am. J. Respir. Crit. Care Med.* 180, 29–35. doi: 10.1164/rccm.200812-1911OC

Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2, 3. doi: 10.1186/2042-5783-2-3

Thompson, A., Bench, S., Carter, B., and Zehr, J. (2013). Coupling FACS and genomic methods for the characterization of uncultivated symbionts. *Methods Enzymol.* 531, 45–60. doi: 10.1016/b978-0-12-407863-5.00003-4

Tripathy, S., Padhi, S. K., Mohanty, S., Samanta, M., and Maiti, N. K. (2016). Analysis of the metatranscriptome of microbial communities of an alkaline hot sulfur spring revealed different gene encoding pathway enzymes associated with energy metabolism. *Extremophiles* 20, 525–536. doi: 10.1007/s00792-016-0846-6

United Nations, Department of Economic and Social Affairs, Population Division (2017). *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*. New York, NY: United Nations.

Vieille, C., and Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* 65, 1–43. doi: 10.1128/MMBR.65.1.1-43.2001

Vogler, M., Karan, R., Renn, D., Vancea, A., Vielberg, M.-T., Grötzinger, S. W., et al. (2020). Crystal Structure and Active Site Engineering of a Halophilic γ-Carbonic Anhydrase. *Front. Microbiol.* 11:742. doi: 10.3389/fmicb.2020.00742

Wetterstrand, K. A. (2013). *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. Burlington, MA: ScienceOpen, Inc.

Wilson, M. C., Mori, T., Rückert, C., Uria, A. R., Helf, M. J., Takada, K., et al. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506, 58–62. doi: 10.1038/nature12959

Wiltschi, B., Cernava, T., Dennig, A., Casas, M. G., Geier, M., Gruber, S., et al. (2020). Enzymes revolutionize the bioproduction of value-added compounds: from enzyme discovery to special applications. *Biotechnol. Adv.* 40, 107520. doi: 10.1016/j.biotechadv.2020.107520

Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., et al. (2010). One bacterial cell, one complete genome. *PLoS One* 5:e10314. doi: 10.1371/journal.pone.0010314

Yamamura, S., Kishi, H., Tokimitsu, Y., Kondo, S., Honda, R., Ramachandra Rao, S., et al. (2005). Single-cell microarray for analyzing cellular response. *Anal. Chem.* 77, 8050–8056. doi: 10.1021/ac0515632

Yan, W., Li, F., Wang, L., Zhu, Y., Dong, Z., and Bai, L. (2017). Discovery and characterizaton of a novel lipase with transesterification activity from hot spring metagenomic library. *Biotechnol. Rep.* 14, 27–33. doi: 10.1016/j.btre.2016.12.007

Yoshimoto, N., Kida, A., Jie, X., Kurokawa, M., Iijima, M., Niimi, T., et al. (2013). An automated system for high-throughput single cell-based breeding. *Sci. Rep.* 3, 1191. doi: 10.1038/srep01191

Zarafeta, D., Moschidi, D., Ladoukakis, E., Gavrilov, S., Chrysina, E. D., Chatziioannou, A., et al. (2016). Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. *Sci. Rep.* 6:38886. doi: 10.1038/srep38886

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

Zhang, K., Martiny, A. C., Reppas, N. B., Barry, K. W., Malek, J., Chisholm, S. W., et al. (2006). Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24, 680–686. doi: 10.1038/nbt1214

Zhao, C., Chu, Y., Li, Y., Yang, C., Chen, Y., Wang, X., et al. (2017). High-throughput pyrosequencing used for the discovery of a novel cellulase from a thermophilic cellulose-degrading microbial consortium. *Biotechnol. Lett.* 39, 123–131. doi: 10.1007/s10529-016-2224-y

Zweerink, S., Kallnik, V., Ninck, S., Nickel, S., Verheyen, J., Blum, M., et al. (2017). Activity-based protein profiling as a robust method for enzyme identification and screening in extremophilic Archaea. *Nat. Commun.* 8, 15352. doi: 10.1038/ncomms15352

Conflict of Interest: 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 © 2021 Sysoev, Grötzinger, Renn, Eppinger, Rueping and Karan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 March 2021
doi: 10.3389/fmicb.2021.627200

Physicochemical and Microbial Diversity Analyses of Indian Hot Springs

Manik Prabhu Narsing Rao[1], Zhou-Yan Dong[1,2], Zhen-Hao Luo[1], Meng-Meng Li[1], Bing-Bing Liu[3], Shu-Xian Guo[3], Wael N. Hozzein[4,5], Min Xiao[1] and Wen-Jun Li[1,6]**

[1] State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Life Sciences, Sun Yat-sen University, Guangzhou, China, [2] Department of Pathogenic Biology, School of Basic Medical Sciences, Binzhou Medical University, Yantai, China, [3] Henan Key Laboratory of Industrial Microbial Resources and Fermentation Technology, College of Biological and Chemical Engineering, Nanyang Institute of Technology, Nanyang, China, [4] Bioproducts Research Chair, Zoology Department, College of Science, King Saud University, Riyadh, Saudi Arabia, [5] Botany and Microbiology Department, Faculty of Science, Beni-Suef University, Beni-Suef, Egypt, [6] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China

OPEN ACCESS

Edited by:
André Antunes,
Macau University of Science and Technology, China

Reviewed by:
Brandon Briggs,
University of Alaska Anchorage, United States
Shawn E. McGlynn,
Tokyo Institute of Technology, Japan

***Correspondence:**
Bing-Bing Liu
liubingbing200666@163.com;
454873856@qq.com
Wen-Jun Li
liwenjun3@mail.sysu.edu.cn

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

Received: 08 November 2020
Accepted: 08 February 2021
Published: 03 March 2021

Citation:
Narsing Rao MP, Dong Z-Y, Luo Z-H, Li M-M, Liu B-B, Guo S-X, Hozzein WN, Xiao M and Li W-J (2021) Physicochemical and Microbial Diversity Analyses of Indian Hot Springs. Front. Microbiol. 12:627200. doi: 10.3389/fmicb.2021.627200

In the present study, physicochemical and microbial diversity analyses of seven Indian hot springs were performed. The temperature at the sample sites ranged from 32 to 67°C, and pH remained neutral to slightly alkaline. pH and temperature influenced microbial diversity. Culture-independent microbial diversity analysis suggested bacteria as the dominant group (99.3%) when compared with the archaeal group (0.7%). Alpha diversity analysis showed that microbial richness decreased with the increase of temperature, and beta diversity analysis showed clustering based on location. A total of 131 strains (divided into 12 genera and four phyla) were isolated from the hot spring samples. Incubation temperatures of 37 and 45°C and T5 medium were more suitable for bacterial isolation. Some of the isolated strains shared low 16S rRNA gene sequence similarity, suggesting that they may be novel bacterial candidates. Some strains produced thermostable enzymes. Dominant microbial communities were found to be different depending on the culture-dependent and culture-independent methods. Such differences could be attributed to the fact that most microbes in the studied samples were not cultivable under laboratory conditions. Culture-dependent and culture-independent microbial diversities suggest that these springs not only harbor novel microbial candidates but also produce thermostable enzymes, and hence, appropriate methods should be developed to isolate the uncultivated microbial taxa.

Keywords: Indian hot springs, culture-dependent microbial diversity analysis, culture-independent microbial diversity analysis, alpha diversity, beta diversity

INTRODUCTION

Hot springs are the places where warm or hot groundwater comes out from the Earth (Narsing Rao et al., 2018). Hot springs were once considered a sterile area (Chaudhuri et al., 2017), but Thomas Brock's groundbreaking work in discovering *Thermus aquaticus* from a thermal environment (Brock and Freeze, 1969; Brock, 1997) has fully altered our perception of the microbial diversity of

hot springs. Researchers around the world have begun to study similar ecosystems using a culture-dependent approach to understand microbial diversity (Baker et al., 2001; Pathak and Rathod, 2014; Liu et al., 2016). For decades, microbial diversity analysis was carried out by the traditional culture-dependent method; however, this method has several disadvantages. Most of the microorganisms in this method remain hidden or difficult to grow (Kumar et al., 2004; Najar et al., 2018; Narsing Rao et al., 2018).

Next-generation sequencing allows for culture-free microbial diversity detection (Sabat et al., 2017), based on molecular phylogeny of the small-subunit ribosomal RNA gene (16S rRNA gene) (Myer et al., 2016). In the past few years, this method played an important role in understanding the microbial diversity of various ecological niches (Ghelani et al., 2015; De Mandal et al., 2017; Anguita-Maeso et al., 2020).

India harbors approximately 400 geothermal springs (Pednekar et al., 2011). People bathe in these hot springs to cure diseases (Pednekar et al., 2011; Narsing Rao et al., 2018). Culture-dependent microbial diversity analysis of Indian hot springs suggested that they harbor novel candidates and produce thermostable secondary metabolites. A novel genus, *Emticicia*, was reported from an Indian (Assam) warm water spring (Saha and Chakrabarti, 2006b). A moderately thermophilic, thiosulfate-oxidizing novel species *Thiomonas bhubaneswarensis* was reported from an Atri hot spring (Bhubaneswar, India) (Panda et al., 2009). The high arsenate-tolerant novel species *Pannonibacter indica* was reported from an Athamallik (Orissa, India) hot spring sediment (Bandyopadhyay et al., 2013). Antimicrobial ability of bacterial strains isolated from a Maharashtra (India) hot spring was also reported (Pednekar et al., 2011). A novel species, *Thermus parvatiensis,* isolated from the hot water spring of Manikaran, India, was reported to produce a thermostable enzyme (protease activity at 70°C) (Dwivedi et al., 2015).

Similarly, culture-independent microbial diversity analysis of Indian hot spring samples was also carried out suggesting that they have diverse microbial diversity and that many microbial communities are still unclassified (Ghelani et al., 2015; Chaudhuri et al., 2017). Although both culture-dependent and culture-independent microbial analyses have been carried out separately to understand the microbial diversity of Indian hot springs (Kumar et al., 2004; Ghelani et al., 2015; Chaudhuri et al., 2017), only a few studies performed both analyses together (Najar et al., 2018). In the present study, we performed both culture-dependent and culture-independent microbial analyses of seven Indian hot springs. We further analyzed the various physicochemical parameters that govern microbial diversity.

MATERIALS AND METHODS

Description of the Sample Sites and Physicochemical Analysis

Seven hot spring samples from three Indian provinces, namely, Karnataka, Maharashtra, and Telangana, were included in the present study (**Figure 1**). Four hot spring samples were collected from Maharashtra Province located at Vajreshwari (VAJ) (19°29′13.2″N, 73°01′40.8″E), Akaloli (AKA) (19°29′24.7″N, 73°02′20.8″E), Ganeshpuri (GAN) (19°30′05.0″N, 73°00′47.5″E), and Sativali (SATI) (19°37′51.0″N, 72°54′32.4″E). The hot spring located in Karnataka Province is called Bendru Theertha (BT) located in the village Irde in Puttur Taluka (12°45′53.3″N, 75°11′03.1″E). The hot spring of Tuwa (TUW) (22°47′58″N, 73°27′37″E) is situated in Panchmahal District, Gujarat. The hot spring at Bhadrachalam (BHA) (17°40′07.7″N, 80°53′37.0″E), Telangana Province in Gundala is located on the bed of the Godavari River.

Water temperature and pH were measured using a portable digital thermometer and pH meter. Samples for culture-dependent analysis were collected in sterilized tubes kept in the dark, while samples for culture-independent analysis were kept in dry ice until brought to the laboratory.

Culture-Independent Microbial Diversity and Statistical Analysis

Community genomic DNA from the sediment samples were extracted using the PowerSoil DNA isolation kit (MoBio) as per the manufacturer's instructions. PCR was carried out by the Phusion® High-Fidelity PCR Master Mix with GC Buffer (New England Biolabs) using specific primers (16S V4 region primers; 515F and 806R) (Caporaso et al., 2011) with a barcode (12 nt). The libraries were constructed using the TruSeq® DNA PCR-free sample preparation kit and sequenced using the HiSeq 2500 platform (Illumina). According to the barcode sequence, data samples were separated. The paired-end reads were assembled into single reads using USEARCH 11 (Edgar, 2010). The raw reads were filtered (Bokulich et al., 2013) referring to the quality control process of QIIME 2 (Bolyen et al., 2019). Chimeras were removed using UCHIME (Edgar et al., 2011). Zero radius operational taxonomic unit (ZOTU) clustering was performed using the UPARSE pipeline (Edgar, 2013), and the taxonomic assignment was performed against the SILVA database (release 132) (Quast et al., 2013). The multisequence alignment was performed using the MUSCLE (version 3.8.31) (Edgar, 2004) software to obtain a systematic relationship of all ZOTU representative sequences. The QIIME 2 software was used to calculate the observed species (Bolyen et al., 2019), and the R software (version 2.15.3)[1] was used to draw the species accumulation curve. The UniFrac distance was calculated using the QIIME 2 software (Bolyen et al., 2019) and the UPGMA sample clustering tree was constructed. The principal coordinate analysis (PCoA) map was drawn by the R software (see text footnote 1) using the WGCNA, stats, and ggplot2 software packages (Zhang and Horvath, 2005; Wickham, 2016). The correlation between ZOTUs was made using the sparse correlations for the compositional (SparCC) data algorithm implemented in the Python module (Friedman and Alm, 2012) with bootstraps and p-values of 1,000 and 0.01, respectively. The corresponding network was plotted using the R package igraph (Csardi and Nepusz, 2006).

[1] http://www.r-project.org/

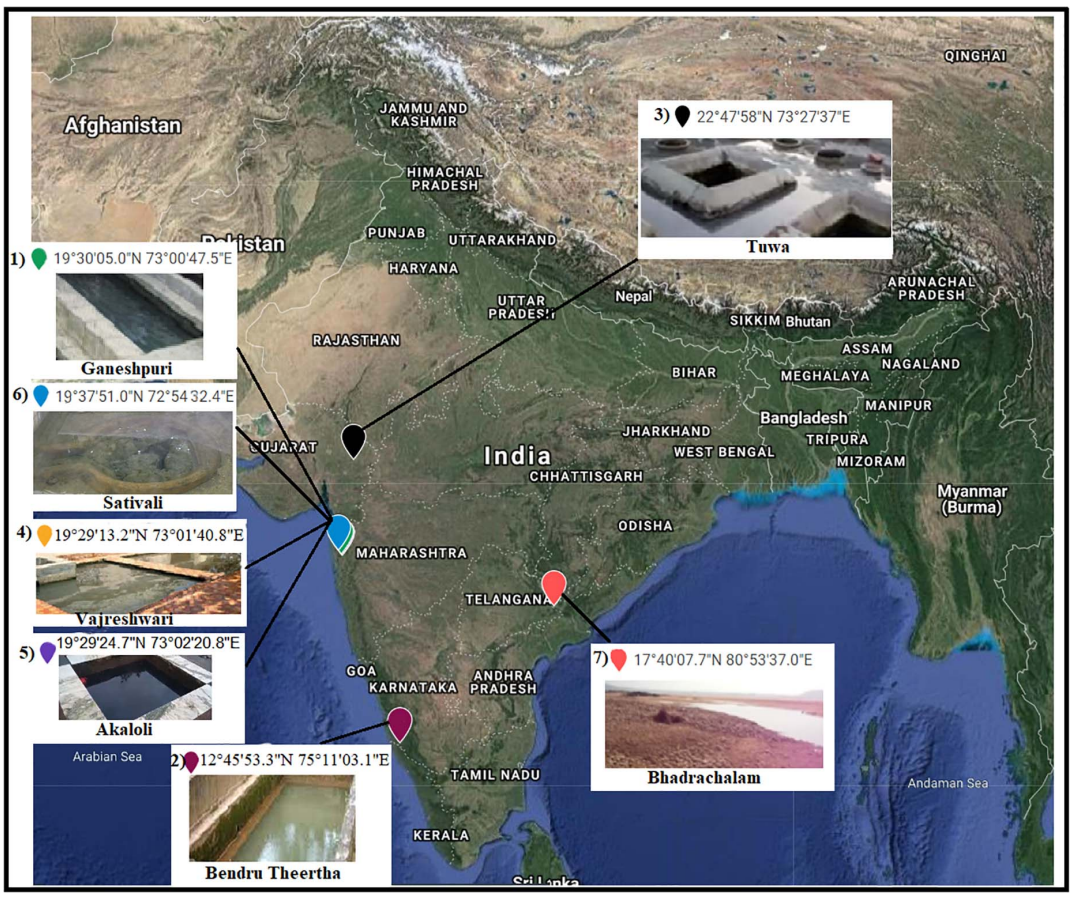


FIGURE 1 | Sample sites. 1, Ganeshpuri; 2, Bendru Theertha; 3, Tuwa; 4, Vajreshwari; 5, Akaloli; 6 Sativali; 7, Bhadrachalam.

Culture-Dependent Microbial Diversity Analysis

According to our pre-laboratory experience of cultivating high-temperature microorganisms (Liu et al., 2016; Khan et al., 2017a,b; Habib et al., 2018), T5, R2A, ISP5, CC, TSA, and TH media were used for the isolation. Their composition details are mentioned in **Supplementary Table S1**.

For culture-dependent microbial diversity analysis, the samples were serially diluted. About 100 μl of the serially diluted samples were spread on six different media and incubated at temperatures of 37, 45, 55, and 70°C. The isolates obtained were picked and re-streaked on the same isolation media until pure colonies were obtained. The isolated strains were identified using 16S rRNA gene sequencing. Genomic DNA extraction and PCR amplification of the 16S rRNA gene sequence was performed as described by Li et al. (2007). The obtained 16S rRNA gene sequence was compared with available sequences of cultured species from the EzTaxon database (Yoon et al., 2017).

Screening for Enzyme Production

The isolated organisms were screened for the presence of industrially important enzymes like amylase, protease, cellulase, and xylanase. Amylase activity was determined by inoculating bacterial strains on starch agar (Difco). The hydrolysis of starch was detected by flooding the plates with an iodine solution. The clear zone around the colony was positive for amylase activity. Protease activity was determined on skim milk agar (Sigma-Aldrich). The zone of hydrolysis around the colony was regarded as positive for protease activity.ı Cellulase activity was determined by incorporating 1% carboxymethylcellulose in tryptic soy agar (Difco). Congo red was used as an indicator for determining carboxymethylcellulose hydrolysis (Sazci et al., 1986).

Data Availability and Accession Numbers

Raw reads were submitted to the NCBI Sequence Read Archive (SRA) under accession numbers SRR10420990–SRR10420996.

RESULTS

Physicochemical Analysis of the Sample Sites

The temperature at the sample sites ranged from 32 to 67°C. The highest temperature was recorded at SATI (67°C), while the

lowest was at VAJ (32°C) (**Table 1**). The pH at all the sample sites remained neutral to slightly alkaline (**Table 1**).

Culture-Independent Microbial Diversity Analysis

A total of 504,226 high-quality clean reads from 536,302 raw reads from seven samples were obtained and grouped into 2,965 ZOTUs. About 99.3% of ZOTUs were assigned as bacteria and 0.7% of ZOTUs were assigned as archaea. **Figure 2** shows the relative abundance of the top 10 bacterial ZOTUs at different sample sites. Overall, in all the sample sites, phylum *Proteobacteria* was dominant. The other dominant phyla were *Firmicutes*, *Actinobacteria*, *Chloroflexi*, *Deinococcus–Thermus*, *Bacteroidetes*, *Cyanobacteria*, *Acidobacteria*, *Armatimonadetes*, and *Spirochaetes*, but their proportion in each sample varied (**Figure 2A**). The BHA hot spring sample had a high relative abundance of *Proteobacteria* and *Firmicutes* when compared with other samples. The GAN and SATI hot spring samples had a high relative abundance of *Deinococcus–Thermus*, while the VAJ and SATI samples had a high relative abundance of *Cyanobacteria* (**Figure 2A**). At the genus level, *Escherichia* and *Shigella* dominated the BHA sample site. *Meiothermus* dominated at the GAN and SATI sample sites. *Acinetobacter* dominated at the TUW and BT sample sites, while *Bacillus* dominated at the AKA and VAJ sample sites (**Figure 2B**).

Alpha diversity analysis showed that sample sites with low temperatures had a high taxonomic richness and the richness decreased with the increase of temperature (**Figure 3A**), with one exception (BHA sample site). Further, the petal diagram (**Figure 3B**) showed that the sample sites with low temperature had more unique ZOTUs when compared with high-temperature sample sites. Beta diversity analysis (**Figures 4A,B**) showed that clustering based on location (GAN, VAJ, and SATI) and their relative abundance at the phylum level were almost the same. Network analysis (**Figure 5**) showed that phyla *Chloroflexi*, *Spirochaetes*, and *Euryarchaeota* had a high correlation with other members.

Culture-Dependent Microbial Diversity Analysis

A total of 131 strains (**Supplementary Table S2**) were isolated from the hot spring samples which were divided into 12 genera and four phyla. At the phylum level, *Firmicutes* were dominant followed by *Actinobacteria*, *Proteobacteria*, and *Bacteroidetes*. At the genus level, *Bacillus* were dominant followed by *Brevibacillus*, *Paenibacillus*, *Oceanibaculum*, *Micromonospora*, *Microbacterium*, *Terrimonas*, *Aneurinibacillus*, *Actinocorallia*, *Sphingomonas*, *Micrococcus*, and *Rhodoligotrophos*. The highest number of strains was isolated using T5 media, followed by R2A, TH, ISP5, and CC media (**Supplementary Table S2**). Strains SYSU-M-14, SYSU-M-111, SYSU-M1-37, and SYSU-M1-3 shared < 97% 16S rRNA gene sequence similarity with *Rhodoligotrophos appendicifer*, *Bacillus alkalitolerans*, *Paenibacillus albidus*, and *Terrimonas lutea*, respectively. It was noticed that the incubation temperature of 37 and 45°C was better for isolation. It was further noticed that no strain was isolated above 55°C. Details about the number of strains, their 16S rRNA gene sequence similarity, sample location, and media used for isolation are mentioned in **Supplementary Table S2**.

The isolated strains were evaluated for their ability to produce enzymes. About 63, 35, 13, and 8% of strains were positive for protease, amylase, cellulase, and xylanase, respectively (**Supplementary Table S2**).

Comparison of Culture-Dependent and Culture-Independent Microbial Diversity Analysis

Figure 6 represents the culture-dependent and culture-independent microbial diversity analyses of seven hot spring samples. In the culture-dependent analysis, *Firmicutes* were dominant followed by *Proteobacteria*, whereas in the culture-independent analysis, *Proteobacteria* were dominant followed by *Firmicutes.* In both culture-dependent and culture-independent analyses, the microbial richness decreased with the increase of temperature. Culture-independent analysis suggests that seven hot spring samples harbored several bacterial phyla, while only four bacterial phyla were recovered using culture-dependent analysis.

DISCUSSION

Hot springs were once perceived to be a sterile environment, but the pioneering work in discovering Taq DNA polymerase from *Thermus aquaticus* not only improved the economy of the industry but also brought a revolution in the field of molecular biology (Chien et al., 1976). The hot spring atmosphere was postulated to be close to the early chemical environment on Earth, and hot springs have thus become a model ecosystem for research on the origin and evolution of life (Li et al., 2015).

In the present study, physicochemical and microbial diversity analyses of seven Indian hot springs were carried out. The temperature of the seven hot springs ranged between 32 and 67°C. The highest temperature was recorded at SATI (67°C) and the lowest at VAJ (32°C). The pH of the hot spring sample ranged

TABLE 1 | Physico-chemical properties of hot springs.

Physico-chemical properties	1	2	3	4	5	6	7
pH	7.6	7.2	7.0	7.4	7.3	7.5	7.0
Temperature°C	66	38	55	32	45	67	45

Sample site:1, Ganeshpuri; 2, Bendru theertha; 3, Tuwa; 4, Vajreshwari; 5, Akaloli; 6, Sativali; 7, Bhadrachalam.

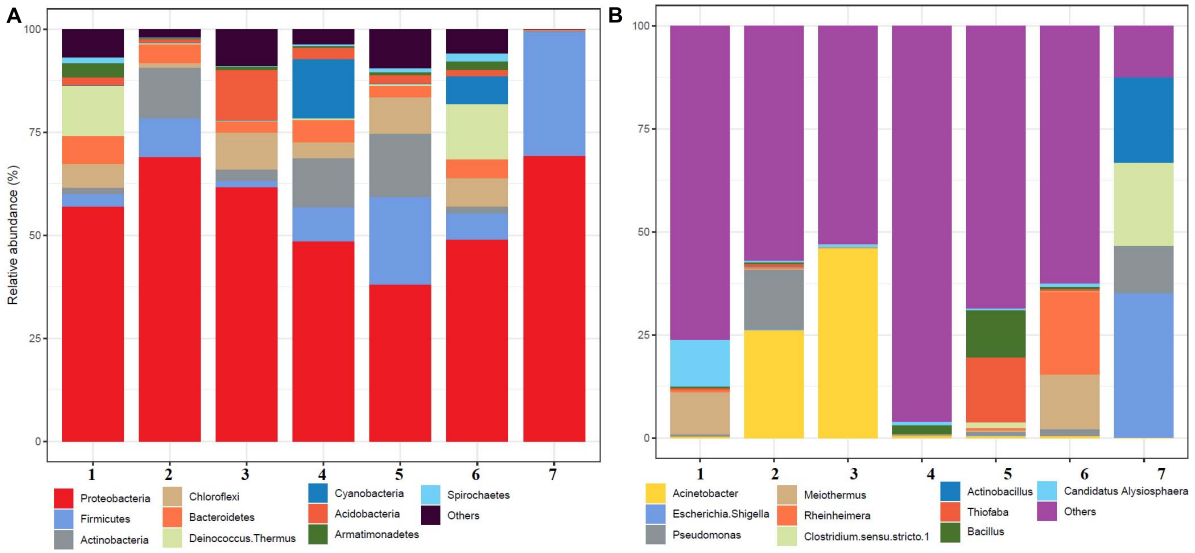


FIGURE 2 | Relative abundance of the top 10 microbial communities based on the culture-independent method: **(A)** phylum level and **(B)** genus level. 1, Ganeshpuri; 2, Bendru Theertha; 3, Tuwa; 4, Vajreshwari; 5, Akaloli; 6 Sativali; 7, Bhadrachalam.

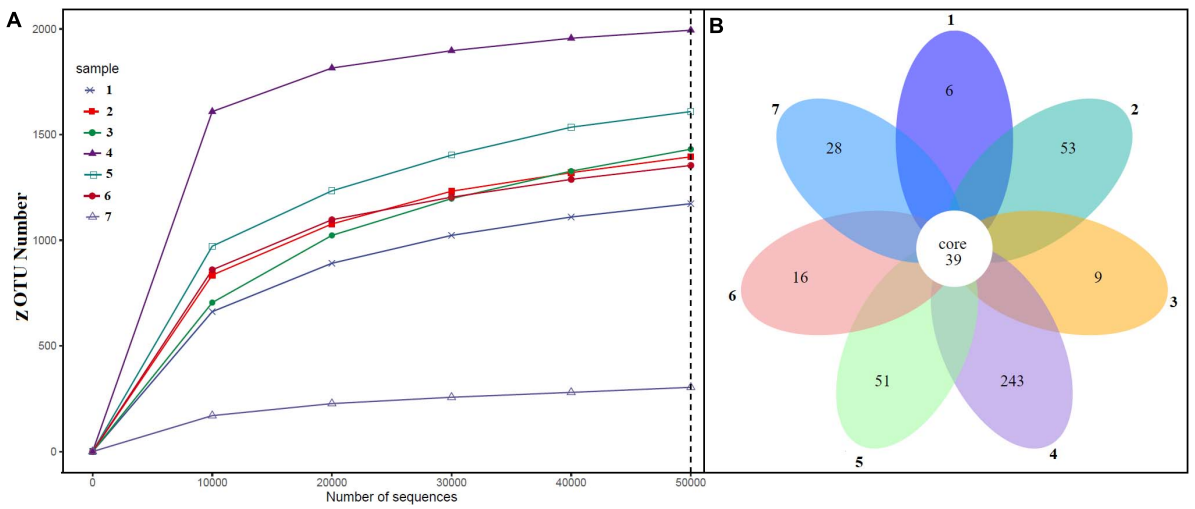


FIGURE 3 | Alpha diversity. **(A)** ZOTUs per sequence number; **(B)** petal graph showing unique and common ZOTUs present in different sample sites. 1, Ganeshpuri; 2, Bendru Theertha; 3, Tuwa; 4, Vajreshwari; 5, Akaloli; 6 Sativali; 7, Bhadrachalam.

from neutral to alkaline which was similar to other Indian hot springs (Craig et al., 2013; Badhai et al., 2015).

Proteobacteria and *Crenarchaeota* (**Figure 2A**) were the dominant bacterial and archaeal phyla found in the present study and also in other Indian hot springs (Badhai et al., 2015; Sahoo et al., 2015; Saxena et al., 2017). Photosynthetic bacteria (*Cyanobacteria*) abundance was observed at VAJ and SATI (**Figure 2A**). The presence of phototrophs in the hot spring suggests the nutritive interaction among the microorganisms (Narsing Rao et al., 2018).

The GAN and SATI sample site temperatures were relatively high (66–67°C) when compared with those of the other sample sites with high abundance of *Deinococcus–Thermus*. Similarly, Jain et al. (2014) noticed a high abundance of *Deinococcus–Thermus* at the Tantloi hot spring, India, with a similar temperature. The sample sites' temperature, ranging from 38 to 45°C (BT, AKA, and BHA), showed a high abundance of *Firmicutes* and their abundance decreased over this temperature range. At the genus level, *Bacillus* dominated at AKA and VAJ, while *Acinetobacter* dominated at the TUW and BT sample sites (**Figure 2B**). The abundance of the genus *Bacillus* has been reported in other Indian hot springs (Mangrola A. et al., 2015; Mangrola A. V. et al., 2015). Similarly, Saxena et al. (2017) noticed high *Acinetobacter* abundance at the Tattapani (India)

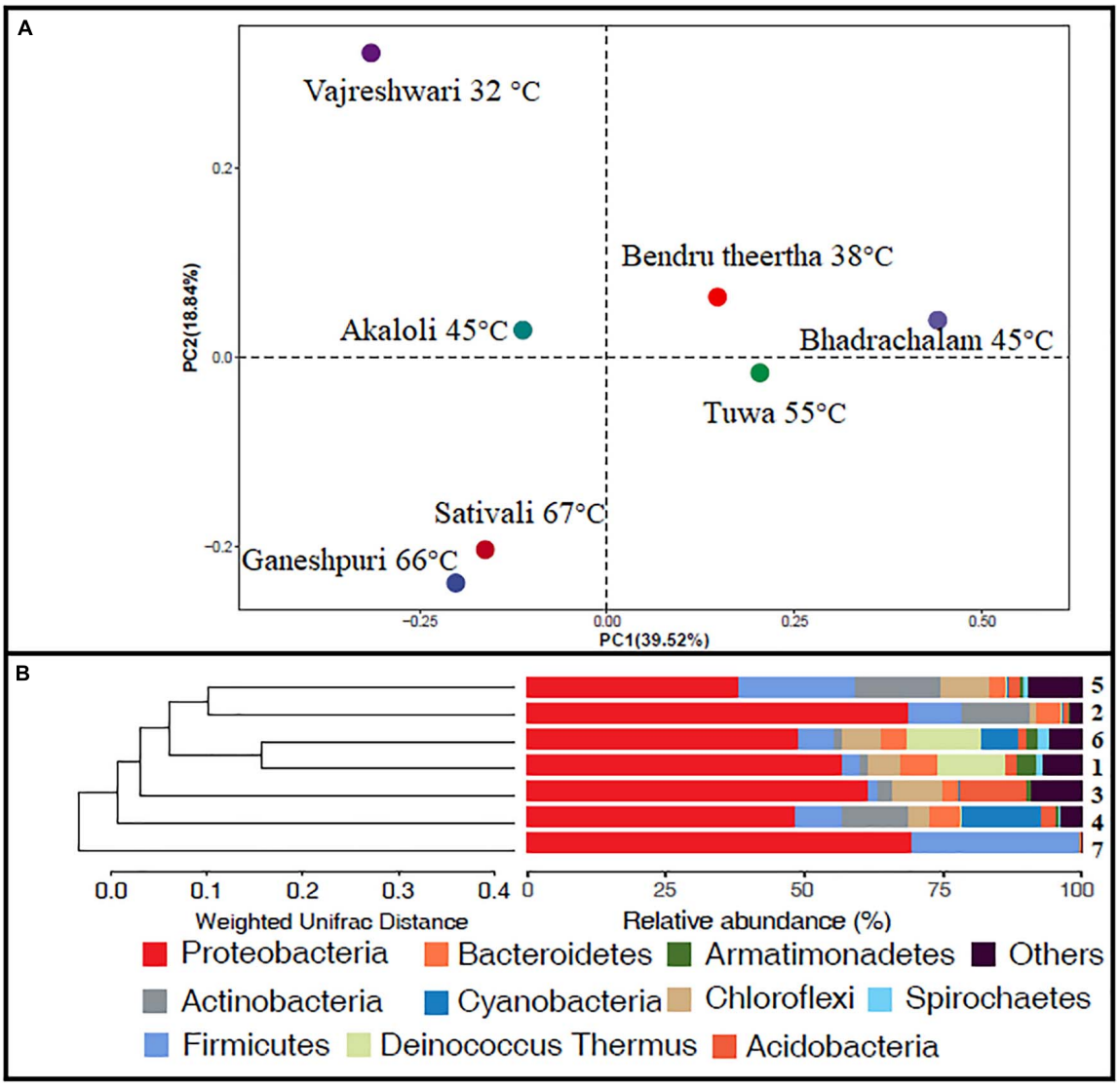


FIGURE 4 | Beta diversity. **(A)** Weighted UniFrac distances calculated using amplicon reads and plotted on a PCoA plot; **(B)** hierarchical clustering based on weighted UniFrac distance (UPGMA) algorithm (1, Ganeshpuri; 2, Bendru Theertha; 3, Tuwa; 4, Vajreshwari; 5, Akaloli; 6 Sativali; 7, Bhadrachalam).

hot spring. The genus *Meiothermus*, known for its thermophilic nature (Chen et al., 2002), dominated at sample sites (GAN and SATI) with a high temperature (66–67°C). The results suggest that temperature plays an important role in structuring hot spring microbial diversity.

In the BHA sample site, *Escherichia* and *Shigella* were dominant which was unusual when compared with the other sample sites in the present study. This might be due to environmental contamination as this sample site was located on the bed of a river.

Alpha diversity analysis showed that sample sites with low temperatures not only had high taxonomic richness but also had more unique ZOTUs (**Figures 3A,B**). In the beta diversity analysis, samples were clustered based on location (**Figures 4A,B**). Our results correlate with the study of Saxena et al. (2017), where the authors noticed high taxonomic richness at low temperatures and region-specific clustering.

Network analysis offers new insight into the structure of complex microbial communities, and such information is particularly valuable in environments, where the basic ecology and life history strategies of many microbial taxa remain unknown (Janssen, 2006). Further, it also helps to reveal the niche spaces shared by community members, direct symbioses between community members, potential biotic interactions, habitat affinities, or shared physiologies that could guide more focused studies or experimental settings (Proulx et al., 2005; Janssen, 2006; Barberán et al., 2012). In the present study, we noticed that phyla *Chloroflexi*, *Spirochaetes*, and *Euryarchaeota* had a high correlation

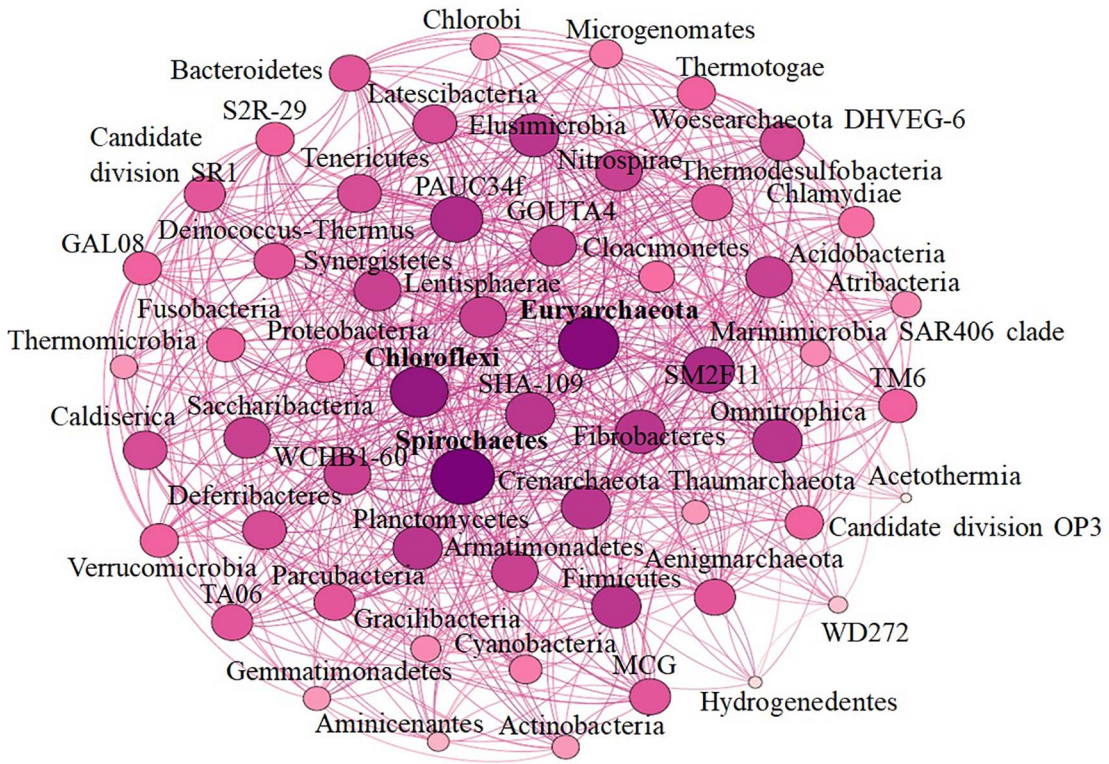


FIGURE 5 | Network-based analysis.

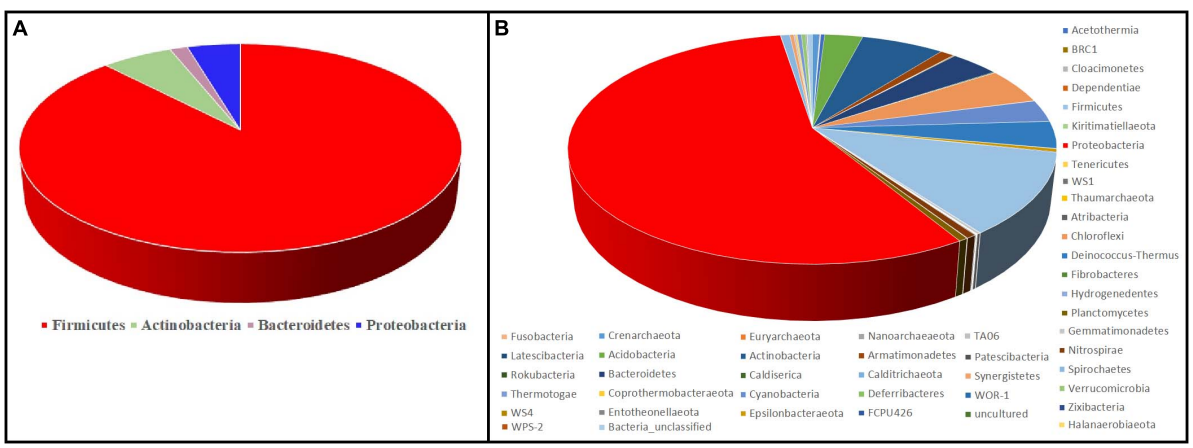


FIGURE 6 | Comparison of microbial diversity at the phylum level. **(A)** Culture-dependent method; **(B)** culture-independent method.

with other members. Similarly, Zamkovaya et al. (2021) reported a high correlation between *Chloroflexi* and other members in the hot spring samples. Although the relative abundance of *Proteobacteria* was high (**Figure 2A**) when compared with *Chloroflexi*, *Spirochaetes*, and *Euryarchaeota*, however, the correlation of *Proteobacteria* with other members was less when compared with *Chloroflexi*, *Spirochaetes*, and *Euryarchaeota* (**Figure 5**). This suggests that the microbe's interactions were specific irrespective of their abundance. However, further studies are needed to enlighten their interactions.

In the present study, *Firmicutes* at the phylum level and *Bacillus* at the genus level were dominant in the culture-dependent microbial diversity analysis. Similarly, Pathak and Rathod (2014) performed culture-dependent bacterial diversity analysis for the terrestrial thermal spring of Unkeshwar, India,

and noticed *Firmicutes* at the phylum level and *Bacillus* at the genus level as the dominant bacterial groups.

Isolation medium and incubation temperature play an important role in determining microbial diversity through culture-dependent analysis. Pathak and Rathod (2014) during the bacterial diversity analysis of the Unkeshwar thermal spring noticed luxuriant bacterial growth on nutrient agar when compared with the other isolation medium. Sen and Maiti (2014) noticed that the optimum growth temperature for strains isolated from Odisha (India) hot springs ranged between 37 and 50°C. In the present study, the temperature range of 37 and 45°C and T5 media were better for isolation. Some of the sample sites' temperature was above 60°C (**Table 1**), and culture-independent analysis showed the presence of microbial taxa (**Figures 2A,B**). However, we did not notice any microbial growth above 55°C using the culture-dependent method, suggesting that there are still many microbial taxa that remain uncultivated and there is still plenty of room for research.

Strains SYSU-M-14 and SYSU-M1-3 isolated from the TUW sample site and SYSU-M-111 and SYSU-M1-37 isolated from the BT sample site shared <97% 16S rRNA gene sequence similarity, suggesting that they could be novel candidates. Similarly, several novel bacterial candidates have been reported from Indian hot springs. Ruckmani et al. (2011) reported a novel genus *Calidifontibacter* from a hot spring sample of Irde (Mangalore, Karnataka State, India). Similarly, novel species *Aeromonas sharmana* and *Flavobacterium indicum* have been reported in an Indian (Assam) warm spring (Saha and Chakrabarti, 2006a,c). The result suggests that Indian hot springs harbor many novel bacterial candidates. We will further characterize strains SYSU-M-14, SYSU-M1-3, SYSU-M-111, and SYSU-M1-37 to determine their taxonomic positions in our future work. The culture-dependent analysis showed that about 13 and 8% of the isolated strains were positive for cellulase and xylanase, respectively (**Supplementary Table S2**). Many bacterial strains isolated from Indian hot springs were reported to produce various enzymes like protease, lipase, esterase, amidase, caseinase, urease, amylase, oxidase, and gelatinase (Tripathy et al., 2016). The results suggest that Indian hot springs could be a valuable source for the enzymes.

Although the culture-dependent method is a classical approach for determining microbial diversity, this method has several disadvantages as most of the microorganisms remain hidden or are difficult to grow due to the lack of essential nutrients and optimal environmental conditions such as temperature, pH, and essential mixtures of gases (Narsing Rao et al., 2018). Culture-independent analysis suggests that seven hot spring samples harbored several bacterial phyla, while only four bacterial phyla were recovered using culture-dependent analysis (**Figure 6**); hence, appropriate methods should be developed to isolate these microbial taxa.

CONCLUSION

Hot springs are an important source for novel strains and bioactive molecules. In the present study, physicochemical and microbial diversity analyses of seven Indian hot springs were performed. The result suggests that physicochemical parameters play an important role in shaping the microbial community of hot springs, and hence, these parameters should be considered during isolation. Culture-dependent and culture-independent microbial diversity analyses suggest that microbial richness decreases with an increase in temperature. Culture-dependent analysis suggests that these hot springs have industrially important thermostable enzymes producing bacterial strains. Culture-independent microbial diversity analysis suggests that the hot springs of the present study harbored diverse and novel microbial populations, but only a few phyla were recovered using culture-dependent analysis, and hence, appropriate methods should be developed to isolate the uncultivated microbial taxa.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

W-JL, MX, WH, and S-XG designed the research and project outline. B-BL and M-ML performed the DNA extraction and physicochemical analysis. MN, Z-HL, and Z-YD performed the culture-dependent and culture-independent analyses. MN drafted the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the Key Realm R&D Program of Guangdong Province (Grant No. 2018B020206001) and the National Natural Science Foundation of China (Grant Nos. 32061143043, 91951205, and 31800001). The Scientific Investigation Voyage supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai). The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through Vice Deanship of Scientific Research Chairs.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.627200/full#supplementary-material

REFERENCES

Anguita-Maeso, M., Olivares-García, C., Haro, C., Imperial, J., Navas-Cortés, J. A., and Landa, B. B. (2020). Culture-dependent and culture-independent characterization of the olive xylem microbiota: effect of sap extraction methods. *Front. Plant Sci.* 10:1708. doi: 10.3389/fpls.2019.01708

Badhai, J., Ghosh, T. S., and Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Front. Microbiol.* 6:1166. doi: 10.3389/fmicb.2015.01166

Baker, G. C., Gaffar, S., Cowan, D. A., and Suharto, A. R. (2001). Bacterial community analysis of Indonesian hot springs. *FEMS. Microbiol. Lett.* 200, 103–109. doi: 10.1111/j.1574-6968.2001.tb10700.x

Bandyopadhyay, S., Schumann, P., and Das, S. K. (2013). *Pannonibacter indica* sp. nov., a highly arsenate-tolerant bacterium isolated from a hot spring in India. *Arch. Microbiol.* 195, 1–8. doi: 10.1007/s00203-012-0840-z

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly juice diversity factors from illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Brock, T. D. (1997). The value of basic research: discovery of *Thermus aquaticus* and other extreme thermophiles. *Genetics* 146, 1207–1210. doi: 10.1093/genetics/146.4.1207

Brock, T. D., and Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J. Bacteriol.* 98, 289–297. doi: 10.1128/JB.98.1.289-297.1969

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

Chaudhuri, B., Chowdhury, T., and Chattopadhyay, B. (2017). Comparative analysis of microbial diversity in two hot springs of Bakreshwar, West Bengal, India. *Genom. Data* 12, 122–129. doi: 10.1016/j.gdata.2017.04.001

Chen, M. Y., Lin, G. H., Lin, Y. T., and Tsay, S. S. (2002). *Meiothermus taiwanensis* sp. nov., a novel filamentous, thermophilic species isolated in Taiwan. *Int. J. Syst. Evol. Microbiol.* 52, 1647–1654. doi: 10.1099/00207713-52-5-1647

Chien, A., Edgar, D. B., and Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* 127, 1550–1557. doi: 10.1128/jb.127.3.1550-1557.1976

Craig, J., Absar, A., Bhat, G., Cadel, G., Hafiz, M., Hakhoo, N., et al. (2013). Hot springs and the geothermal energy potential of Jammu & Kashmir State, N.W. Himalaya, India. *Earth Sci. Rev.* 126, 156–177. doi: 10.1016/j.earscirev.2013.05.004

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Int. J. Complex Syst.* 1695, 1–9.

De Mandal, S., Chatterjee, R., and Kumar, N. S. (2017). Dominant bacterial phyla in caves and their predicted functional roles in C and N cycle. *BMC Microbiol.* 17:90. doi: 10.1186/s12866-017-1002-x

Dwivedi, V., Kumari, K., Gupta, S. K., Kumari, R., Tripathi, C., Lata, P., et al. (2015). *Thermus parvatiensis* RL^T sp. nov., isolated from a hot water spring, located a top the Himalayan ranges at Manikaran, India. *Indian. J. Microbiol.* 55, 357–365. doi: 10.1007/s12088-015-0538-4

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

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME Improves Speed and Sensitivity of Chimera Detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8:e1002687. doi: 10.1371/journal.pcbi.1002687

Ghelani, A., Patel, R., Mangrola, A., and Dudhagara, P. (2015). Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam hot springs, India. *Genom. Data* 4, 54–56. doi: 10.1016/j.gdata.2015.03.003

Habib, N., Khan, I. U., Salam, N., Xiao, M., Ahmed, I., Zhi, X. Y., et al. (2018). *Tepidimonas sediminis* sp. nov. and *Tepidimonas alkaliphilus* sp. nov., two novel moderately thermophilic species isolated from a hot spring. *Antonie Van Leeuwenhoek* 111, 1023–1031. doi: 10.1007/s10482-017-1002-8

Jain, P., Reza, H. M., and Pa, S. (2014). Molecular phylogenetic analysis of bacterial community and characterization of Cr(VI) reducers from the sediments of Tantloi hot spring, India. *Aquat. Biosyst.* 10:7. doi: 10.1186/2046-9063-10-7

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72, 1719–1728. doi: 10.1128/AEM.72.3.1719-1728.2006

Khan, I. U., Habib, N., Hussain, F., Xian, W. D., Amin, A., Zhou, E. M., et al. (2017a). *Thermus caldifontis* sp. nov., a thermophilic bacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 67, 2868–2872. doi: 10.1099/ijsem.0.002037

Khan, I. U., Hussain, F., Tian, Y., Habib, N., Xian, W. D., Jiang, Z., et al. (2017b). *Tibeticola sediminis* gen. nov., sp. nov., a thermophilic bacterium isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 67, 1133–1139. doi: 10.1099/ijsem.0.001777

Kumar, B., Trivedi, P., Mishra, A. K., Pandey, A., and Palni, L. M. (2004). Microbial diversity of soil from two hot springs in Uttaranchal Himalaya. *Microbiol. Res.* 159, 141–146. doi: 10.1016/j.micres.2004.01.004

Li, H., Yang, Q., Li, J., Gao, H., Li, P., and Zhou, H. (2015). The impact of temperature on microbial diversity and AOA activity in the Tengchong Geothermal Field, China. *Sci. Rep.* 5:17056. doi: 10.1038/srep17056

Li, W. J., Xu, P., Schumann, P., Zhang, Y. Q., Pukall, R., Xu, L. H., et al. (2007). Georgenia ruanii sp. nov., a novel actinobacterium isolated from forest soil in Yunnan (China), and emended description of the genus *Georgenia*. *Int. J. Syst. Evol. Microbiol.* 57, 1424–1428. doi: 10.1099/ijs.0.64749-0

Liu, L., Salam, N., Jiao, J. Y., Jiang, H. C., Zhou, E. M., Yin, Y. R., et al. (2016). Diversity of culturable thermophilic actinobacteria in hot springs in Tengchong, China and studies of their biosynthetic gene profiles. *Microb. Ecol.* 72, 150–162. doi: 10.1007/s00248-016-0756-2

Mangrola, A., Dudhagara, P., Koringa, P., Joshi, C. G., Parmar, M., and Patel, R. (2015). Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genom. Data* 4, 153–155. doi: 10.1016/j.gdata.2015.04.014

Mangrola, A. V., Dudhagara, P., Koringa, P., Joshi, C. G., and Patel, R. K. (2015). Shotgun metagenomic sequencing based microbial diversity assessment of Lasundra hot spring, India. *Genom. Data* 4, 73–75. doi: 10.1016/j.gdata.2015.03.005

Myer, P. R., Kim, M., Freetly, H. C., and Smith, T. P. L. (2016). Evaluation of 16S rRNA amplicon sequencing using two next-generation sequencing technologies for phylogenetic analysis of the rumen bacterial community in steers. *J. Microbiol. Methods* 127, 132–140. doi: 10.1016/j.mimet.2016.06.004

Najar, I. N., Sherpa, M. T., Das, S., Das, S., and Thakur, N. (2018). Microbial ecology of two hot springs of Sikkim: predominate population and geochemistry. *Sci. Total Environ.* 63, 730–745. doi: 10.1016/j.scitotenv.2018.05.037

Narsing Rao, M. P., Liu, L., Jiao, J. Y., Xiao, M., and Li, W. J. (2018). "Hot springs of india: occurrence and microbial diversity," in *Extremophiles in Eurasian Ecosystems: Ecology, Diversity, and Applications. Microorganisms for Sustainability*, Vol. 8, eds D. Egamberdieva, N. K. Birkeland, H. Panosyan, and W. J. Li (Singapore: Springer), 29–55. doi: 10.1007/978-981-13-0329-6_2

Panda, S. K., Jyoti, V., Bhadra, B., Nayak, K. C., Shivaji, S., Rainey, F. A., et al. (2009). *Thiomonas bhubaneswarensis* sp. nov., an obligately mixotrophic, moderately thermophilic, thiosulfate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* 59, 2171–2175. doi: 10.1099/ijs.0.007120-0

Pathak, A. P., and Rathod, M. G. (2014). Cultivable bacterial diversity of terrestrial thermal spring of Unkeshwar, India. *J. Biochem. Technol.* 5, 814–818.

Pednekar, P., Jain, R., and Mahajan, G. (2011). Anti-infective potential of hotspring bacteria. *J. Glob. Infect. Dis.* 3, 241–245. doi: 10.4103/0974-777X.83529

Proulx, S. R., Promislow, D. E., and Phillips, P. C. (2005). Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20, 345–353. doi: 10.1016/j.tree.2005.04.004

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

Ruckmani, A., Kaur, I., Schumann, P., Klenk, H. P., and Mayilraj, S. (2011). Calidifontibacter indicus gen. nov., sp. nov., a member of the family *Dermacoccaceae* isolated from a hot spring, and emended description of the family *Dermacoccaceae*. *Int. J. Syst. Evol Microbiol.* 61, 2419–2424. doi: 10.1099/ijs.0.025593-0

Sabat, A. J., van Zanten, E., Akkerboom, V., Wisselink, G., van Slochteren, K., de Boer, R. F., et al. (2017). Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification - increased discrimination of closely related species. *Sci. Rep.* 7:3434. doi: 10.1038/s41598-017-03458-6

Saha, P., and Chakrabarti, T. (2006a). *Aeromonas sharmana* sp. nov., isolated from a warm spring. *Int. J. Syst. Evol. Microbiol.* 56, 1905–1909. doi: 10.1099/ijs.0.63972-0

Saha, P., and Chakrabarti, T. (2006b). *Emticicia oligotrophica* gen. nov., sp. nov., a new member of the family '*Flexibacteraceae*', phylum Bacteroidetes. *Int. J. Syst. Evol. Microbiol.* 56, 991–995. doi: 10.1099/ijs.0.64086-0

Saha, P., and Chakrabarti, T. (2006c). *Flavobacterium indicum* sp. nov., isolated from warm spring water in Assam. India. *Int. J. Syst. Evol. Microbiol.* 56, 2617–2621. doi: 10.1099/ijs.0.64309-0

Sahoo, R. K., Subudhi, E., and Kumar, M. (2015). Investigation of bacterial diversity of hot springs of Odisha, India. *Genom. Data* 6, 188–190. doi: 10.1016/j.gdata.2015.09.018

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., et al. (2017). Metagenomic analysis of hot springs in Central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.* 7:2123. doi: 10.3389/fmicb.2016.02123

Sazci, A., Radford, A., and Erenle, K. (1986). Detection of cellulolytic fungi by using Congo red as an indicator: a comparative study with the dinitrosalicyclic acid reagent method. *J. Appl. Bacteriol.* 61, 559–562. doi: 10.1111/j.1365-2672.1986.tb01729.x

Sen, R., and Maiti, N. K. (2014). Genomic and functional diversity of bacteria isolated from hot springs in Odisha. India. *Genomicrobiol. J.* 31, 541–550. doi: 10.1080/01490451.2013.850560

Tripathy, S., Padhi, S. K., Sen, R., Maji, U., Samanta, M., Mohanty, S., et al. (2016). Draft genome sequence of *Brevibacillus borstelensis* cifa_chp40, a thermophilic strain having biotechnological importance. *J. Genom.* 4, 4–6. doi: 10.7150/jgen.14036

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBio-Cloud: 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

Zamkovaya, T., Foster, J. S., de Crécy-Lagard, V., and Conesa, A. (2021). A network approach to elucidate and prioritize microbial dark matter in microbial communities. *ISME J.* 15, 228–244. doi: 10.1038/s41396-020-00777-x

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:16646834. doi: 10.2202/1544-6115.1128

Conflict of Interest: 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 © 2021 Narsing Rao, Dong, Luo, Li, Liu, Guo, Hozzein, Xiao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) 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: 11 March 2021
doi: 10.3389/fmicb.2021.603791

Taxogenomic and Comparative Genomic Analysis of the Genus *Saccharomonospora* Focused on the Identification of Biosynthetic Clusters PKS and NRPS

Ninfa Ramírez-Durán[1,2†], Rafael R. de la Haba[2†], Blanca Vera-Gargallo[2], Cristina Sánchez-Porro[2], Scarlett Alonso-Carmona[1], Horacio Sandoval-Trujillo[3] and Antonio Ventosa[2]*

[1] Faculty of Medicine, Autonomous University of the State of Mexico, Toluca, Mexico, [2] Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Seville, Spain, [3] Department of Biological Systems, Metropolitan Autonomous University-Xochimilco, Mexico City, Mexico

OPEN ACCESS

Edited by:
Jesse G. Dillon,
California State University, Long Beach, United States

Reviewed by:
David Edward Whitworth,
Aberystwyth University, United Kingdom
Lorena Carro,
Newcastle University, United Kingdom
Hilal Ay,
Ondokuz Mayıs University, Turkey

***Correspondence:**
Antonio Ventosa
ventosa@us.es

[†] These authors have contributed equally to this work

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

Received: 07 September 2020
Accepted: 17 February 2021
Published: 11 March 2021

Citation:
Ramírez-Durán N, de la Haba RR, Vera-Gargallo B, Sánchez-Porro C, Alonso-Carmona S, Sandoval-Trujillo H and Ventosa A (2021) Taxogenomic and Comparative Genomic Analysis of the Genus Saccharomonospora Focused on the Identification of Biosynthetic Clusters PKS and NRPS. Front. Microbiol. 12:603791. doi: 10.3389/fmicb.2021.603791

Actinobacteria are prokaryotes with a large biotechnological interest due to their ability to produce secondary metabolites, produced by two main biosynthetic gene clusters (BGCs): polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS). Most studies on bioactive products have been carried out on actinobacteria isolated from soil, freshwater or marine habitats, while very few have been focused on halophilic actinobacteria isolated from extreme environments. In this study we have carried out a comparative genomic analysis of the actinobacterial genus *Saccharomonospora*, which includes species isolated from soils, lake sediments, marine or hypersaline habitats. A total of 19 genome sequences of members of *Saccharomonospora* were retrieved and analyzed. We compared the 16S rRNA gene-based phylogeny of this genus with evolutionary relationships inferred using a phylogenomic approach obtaining almost identical topologies between both strategies. This method allowed us to unequivocally assign strains into species and to identify some taxonomic relationships that need to be revised. Our study supports a recent speciation event occurring between *Saccharomonospora halophila* and *Saccharomonospora iraqiensis*. Concerning the identification of BGCs, a total of 18 different types of BGCs were detected in the analyzed genomes of *Saccharomonospora*, including PKS, NRPS and hybrid clusters which might be able to synthetize 40 different putative products. In comparison to other genera of the *Actinobacteria*, members of the genus *Saccharomonospora* showed a high degree of novelty and diversity of BGCs.

Keywords: actinobacteria, *Saccharomonospora*, biosynthetic gene clusters, secondary metabolites, polyketide synthase, non-ribosomal peptide synthetase, comparative genomic analysis, taxophylogenomic analysis

INTRODUCTION

Actinobacteria are a diverse group of clinical, industrial and ecologically important bacteria known for their capacity to cause diseases as well as their potential to produce secondary metabolites widely used in a variety of fields, such as medicine, pharmacy, industrial microbiology and biotechnology, among others (Bérdy, 2012; Doroghazi and Metcalf, 2013). Synthesis of bacterial secondary metabolites might have been originated due to necessary adaptations to diverse environments with the objective to help the producer microorganism to compete for survival resources (Doroghazi and Metcalf, 2013).

Genes involved in secondary metabolite production are commonly grouped within the genome and are usually known as biosynthetic gene clusters (BGCs) (Osbourn, 2010). *Actinobacteria* represent one of the most important sources for discovering new biologically active metabolites (Kamjam et al., 2017; Matsumoto and Takahashi, 2017) due to the presence of two main biosynthetic clusters, polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS), involved in the synthesis of bioactive molecules through multifunctional pathways (Fischbach and Walsh, 2006). Biosynthetic gene clusters for secondary metabolites have been studied in actinobacteria isolated from different environments, such as soil (Sharma et al., 2016; Wang et al., 2018), caves (Gosse et al., 2019), freshwater (Zothanpuia et al., 2018), marine (Zhang et al., 2019; Ishaque et al., 2020) and as endophytes from plants (Passari et al., 2016, 2017). Halophilic actinobacteria from extreme ecosystems have also been reported, but only biosynthetic gene clusters from *Actinopolyspora erythraea* YIM 90600^T have been located and identified so far (Chen et al., 2014; Corral et al., 2020).

Although a large number of secondary metabolites has been discovered, new bioactive products are still required to cover the current needs of clinical, veterinary and agricultural areas, given the frequent development of microbial resistance of pathogens (Rangseekaew and Pathom-aree, 2019), the appearance of novel diseases, the existence of naturally resistant bacteria and the toxicity of several compounds used in agriculture and cattle industry (Fischbach and Walsh, 2009). The rapid advance of complete genome sequencing and comparative genomic techniques have made it possible to apply bioinformatic tools, such as genomic mining, to carry out a comparative analysis of genomes available in public databases (Challis, 2008). This approach allows one, on the one hand, to comprehend the evolutionary relationships among microbial species and, on the other hand, to identify clusters of genes involved in biosynthetic pathways of relevant biomolecules with potential use for disease treatments.

The actinobacterial genus *Saccharomonospora*, belonging to the family *Pseudonocardiaceae*, was proposed in 1971 (Nonomura and Ohara, 1971) and currently comprises 14 validly published species names (Nonomura and Ohara, 1971; Challis, 2008; Syed et al., 2008; Fischbach and Walsh, 2009; Veyisoglu et al., 2013; Chen et al., 2014; Passari et al., 2016, 2017; Tseng et al., 2018; Zothanpuia et al., 2018; Rangseekaew and Pathom-aree, 2019; Zhang et al., 2019; Corral et al., 2020; Ishaque et al., 2020). The species name "*Saccharomonospora internatus*" has been effectively but not validly published (Greiner-Mai et al., 1987) and has been recently reclassified as an additional strain of the species *Saccharomonospora viridis* (Nouioui et al., 2018). Finally, the species initially described as *Saccharomonospora paurometabolica* (Li et al., 2003) was proved to be a subspecies of *Saccharomonospora iraqiensis* (Nouioui et al., 2018). Members of this genus have been isolated from diverse environments, such as soil, lake sediments, marsh soil, peat, manure, compost, and overheated fodder, with some of them being classified as halophilic microorganisms (Kim, 2015).

In this study we have carried out phylogenomic and comparative genomic analyses of the genus *Saccharomonospora* aimed at reviewing the current taxonomy of this genus and locating and identifying PKS and NRPS biosynthetic clusters present in the genomes with potential for future development of novel active biomolecules.

MATERIALS AND METHODS

Genome Sequence Retrieval

All genome sequences affiliated to members of the genus *Saccharomonospora* and publicly available at the National Center for Biotechnology Information (NCBI) and at the Joint Genome Institute (JGI) databases as of May 31st, 2019 were retrieved. A total of 19 complete or draft genomes were recovered, including type and other reference strains (**Table 1**). Genome sequences for the type strains of the species *S. colocasiae*, *S. oceani*, and *S. xiaoerkulensis* were not included in the analyses due to the absence of data in public databases. Prediction of Open Reading Frames and genome annotation, including protein-coding genes as well as other functional genome units such as structural RNAs, tRNAs, was performed following the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Haft et al., 2018).

Phylogenetic and Phylogenomic Analyses

Phylogenetic trees based on the 16S rRNA gene sequences extracted from the annotated genomes or retrieved from GenBank/EMBL/DDBJ databases were constructed using the neighbor-joining (Saitou and Nei, 1987), maximum-parsimony (Fitch, 1971) and maximum-likelihood (Felsenstein, 1981) algorithms in MEGA X software (Kumar et al., 2018). The distance matrix was corrected using the Jukes-Cantor model of DNA evolution (Jukes and Cantor, 1969). Branch support was assessed by 1,000 bootstrap *pseudo*-replicates (Felsenstein, 1985). For phylogenomic tree construction only amino acidic sequences of single-copy core genes shared by all the analyzed genomes were used, as described elsewhere (de la Haba et al., 2019). Selection of those common orthologous genes (OGs) was done by finding the reciprocal best matches between pairs of genome sequences after an all-versus-all Blast search as implemented in the Enveomics collection toolbox (Rodriguez-R and Konstantinidis, 2016). Subsequently, selected OGs were aligned by using Muscle (Edgar, 2004), concatenated and employed for approximately maximum-likelihood phylogenomic tree reconstruction with FastTreeMP v.2.1.8 (Price et al., 2010)

TABLE 1 | Main features of genome sequences of strains of the genus *Saccharomonospora* used in this study.

Strain	Accession No.	Assembly	Level	Size (Mb)	GC%	Scaffolds	Contigs	CDS	N50	L50	Group
Saccharomonospora amisosensis DSM 45685^T	PRJNA546941 [1]	GCF_011761185.1	Scaffold	5.53	68.6	3	3	5,395	4,804,314	1	Marine/Lake
Saccharomonospora azurea NA-128^T	CM001466.1	GCF_000231055.3	Chromosome	4.76	70.0	1	96	4,406	4,763,852	1	Terrestrial
Saccharomonospora azurea SZMC 14600	AHBX00000000.1	GCF_000236985.1	Contig	4.97	70.3	216	218	4,624	55,033	29	Terrestrial
Saccharomonospora cyanea NA-134^T	CM001440.1	GCF_000244975.1	Chromosome	5.41	69.7	1	5	5,086	5,408,301	1	Terrestrial
Saccharomonospora glauca K62^T	CM001484.1 (chromosome), CM001485.1 (plasmid)	GCF_000243395.3	Chromosome	4.56	69.1	2	11	4,269	4,538,746	1	Terrestrial
Saccharomonospora halophila 8^T	AICX00000000.1	GCF_000383775.1	Scaffold	3.69	70.4	116	644	3,656	69,647	16	Moderately halophilic terrestrial
Saccharomonospora iraqiensis subsp. *iraqiensis* IQ-H1^T	AICW00000000.1	GCF_000430445.1	Scaffold	3.90	71.0	139	732	4,014	87,719	16	Moderately halophilic terrestrial
Saccharomonospora iraqiensis subsp. *paurometabolica* YIM 90007^T	AGIT00000000.2	GCF_000231035.3	Scaffold	4.67	71.3	82	1,031	4,537	171,692	9	Moderately halophilic terrestrial
Saccharomonospora marina XMU15^T	CM001439.1	GCF_000244955.1	Chromosome	5.97	68.9	1	8	5,664	5,965,593	1	Marine/Lake
Saccharomonospora piscinae KCTC 19743^T	VCEK00000000.1	GCF_005862235.1	Scaffold	4.90	71.0	11	21	4,508	1,086,926	3	Marine/Lake
Saccharomonospora saliphila YIM 90502^T	AICY00000000.1	GCF_000383795.1	Scaffold	4.03	70.6	145	679	3,916	96,274	11	Moderately halophilic terrestrial
Saccharomonospora viridis DSM 43017^T	CP001683.1	GCF_000023865.1	Complete	4.31	67.3	1	1	3,870	4,308,349	1	Clinical
Saccharomonospora viridis JCM 3315	JRZE00000000.1	GCF_000787535.1	Contig	4.30	67.4	13	13	3,876	1,283,650	2	Clinical
Saccharomonospora viridis ATCC 33517	FOWS00000000.1	GCF_900115515.1	Scaffold	4.31	67.4	12	17	3,879	1,283,650	2	Clinical
Saccharomonospora xinjiangensis XJ-54^T	AICV00000000.1	GCF_000258175.1	Scaffold	4.78	68.9	2	16	4,366	4,773,543	1	Terrestrial
Saccharomonospora sp. 31sw	CP038101.1	GCF_004519465.1	Complete	4.71	69.1	1	1	4,350	4,713,652	1	Terrestrial
Saccharomonospora sp. LRS4.154	MWIH00000000.1	GCF_002077655.1	Scaffold	4.86	71.0	13	17	4,568	731,563	3	Marine/Lake
Saccharomonospora sp. CNQ-490	AZUM00000000.1	GCF_000527075.1	Contig	4.94	71.1	25	25	4,518	613,253	3	Marine/Lake
Saccharomonospora sp. CUA-673	MKKE00000000.1	GCF_001942305.1	Contig	5.42	70.0	85	85	5,219	122,768	14	Marine/Lake

[1] *NCBI BioProject ID.*

TABLE 2 | Grouping of *Saccharomonospora* strains under study according to their isolation source and NaCl requirements for growing.

Strain	Isolation source	NaCl range (optimum),% (w/v)	Group
Saccharomonospora amisosensis DSM 45685^T	Deep marine sediment in Black Sea (Turkey)	0–10	Marine/Lake
Saccharomonospora marina XMU15^T	Ocean sediment of the East China Sea	0–5 (0–3)	
Saccharomonospora piscinae KCTC 19743^T	Fishpond sediment in Taiwan	0–8 (5)	
Saccharomonospora sp. LRS4.154	Laguna del Rosario in Oaxaca (Mexico)	0–8 (5)	
Saccharomonospora sp. CNQ-490	Marine sediment in San Diego (United States)	∼2.5	
Saccharomonospora sp. CUA-673	Marine sponge in San Diego (United States)	∼2.5	
Saccharomonospora halophila 8^T	Marsh soil in Kuwait	10–30 (10)	Moderately halophilic terrestrial
Saccharomonospora iraqiensis subsp. *iraqiensis* IQ-H1^T	Extremely saline soil in Iraq	5–20 (10–15)	
Saccharomonospora iraqiensis subsp. *paurometabolica* YIM 90007^T	Soil sample from the Xinjiang Province (China)	5–20 (10)	
Saccharomonospora saliphila YIM 90502^T	Muddy soil Karnataka Province (India)	0–20 (10)	
Saccharomonospora azurea NA-128^T	Soil sample from Sichuan (China)	≤ 7	Terrestrial
Saccharomonospora azurea SZMC 14600	Soil from China	≤ 7	
Saccharomonospora cyanea NA-134^T	Soil samples from Sichuan (China)	≤ 10	
Saccharomonospora glauca K62^T	Moldy hay, soil, compost, and manure from Germany	≤ 7	
Saccharomonospora xinjiangensis XJ-54^T	Soil in Xinjiang (China)	0	
Saccharomonospora sp. 31sw	Soil from Iran	3	
Saccharomonospora viridis DSM 43017^T	Manure, compost, overheated fodder, soil, lake sediments, peat	≤ 3	Clinical
Saccharomonospora viridis JCM 3315			
Saccharomonospora viridis ATCC 33517			

where branch confidence was evaluated using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999).

Comparative Genomic Analyses

Measures of similarity between genome sequences were achieved by using different algorithms to calculate Overall Genome Relatedness Indexes (OGRI) values, i.e., Orthologous Average Nucleotide Identity (OrthoANI), estimated using the OrthoANI-usearch Tool (Yoon et al., 2017), digital DNA-DNA hybridization (dDDH), determined using the Genome-to-Genome Distance Calculator (GGDC) (formula 2) (Meier-Kolthoff et al., 2013), and Average Amino-acid Identity (AAI), calculated using AAI calculator (Rodriguez-R and Konstantinidis, 2016).

Pan-genome analysis was carried out using the Enveomics tool (Rodriguez-R and Konstantinidis, 2016) to cluster CDSs in orthologous and singleton gene clusters, as indicated in section "Phylogenetic and Phylogenomic Analyses." Pan-genome visualization was performed using Anvi'o (Eren et al., 2015). Flower plot displaying the core, accessory, and strain-specific genes was drawn in R using the 'plotrix' package. The characteristic curves showing the pan-genome and the core-genome evolution at the genus level were depicted using the Pan-Genome Profile Analyze Tool (PanGP) (Zhao et al., 2014) with distance guide (DG) sampling algorithm.

Synteny among the genomes of strains of the genus *Saccharomonospora* was studied after rearranging the draft genome scaffolds to a related reference genome using the Mauve Contig Mover functionality (Rissman et al., 2009). Pairwise comparison between ordered genomes was achieved by Blastn search (*e*-value $\leq 10{-}^3$) and synteny plots were displayed using Easyfig v.2.2.3 (Sullivan et al., 2011). Pan- and core-genome calculations were performed with the Enveomics package (Rodriguez-R and Konstantinidis, 2016).

Biosynthetic Gene Cluster (BGC) Analysis

Location and identification of BGCs within genomes of *Saccharomonospora* was accomplished with the help of antiSMASH 5.0 web server (Blin et al., 2019), which uses Hidden Markov Models and rules based-detection to identify a broad array of BGCs, including those encoding PKS and NRPS. AntiSMASH detection strictness level was set to "relaxed." In order to detect if the identified BGCs were enriched in the core- or accessory-genomes, we performed a nucleotide Blast search of the reverse translated protein pertaining to the core- and dispensable-genomes, respectively, against the BGCs from each of the analyzed genomes, and the corresponding Blast alignment length was used to calculate the fraction of BGCs spanned by core and accessory genes, respectively.

Since large BGCs (such as PKS, NRPS and hybrid clusters) may be split across several scaffolds/contigs when analyzing incomplete genomes leading to miscalculation of the predicted number of BGCs by antiSMASH, we used the NaPDoS (Natural Product Domain Seeker) tool (Ziemert et al., 2012) to connect PKS, NRPS or hybrid clusters divided into several scaffolds/contigs. BGC domains harbor the genetic signature of their historical relative and thus scaffolds/contigs containing

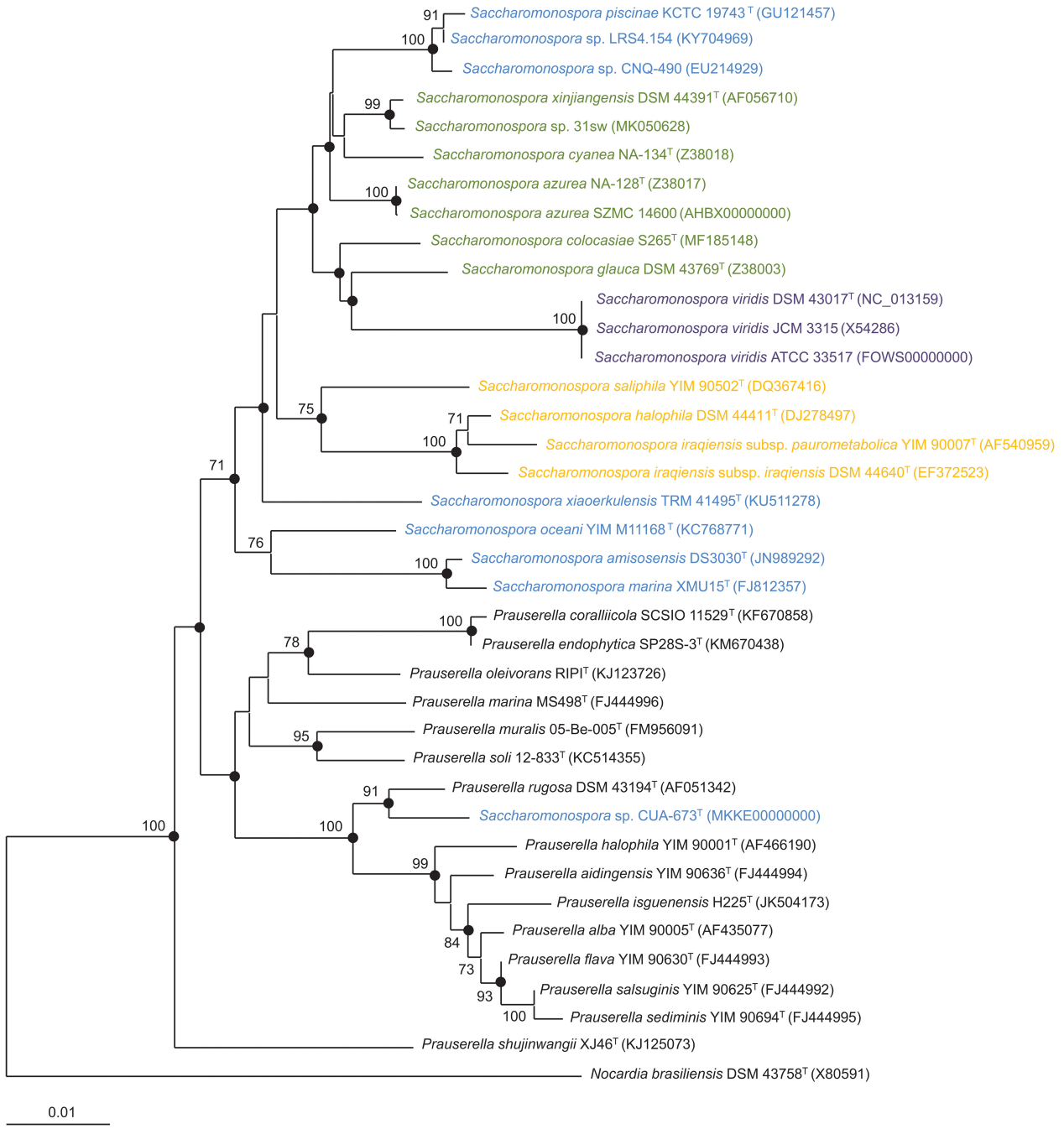


FIGURE 1 | Neighbor-joining phylogenetic tree based on the 16S rRNA gene sequence comparison of members of the genus *Saccharomonospora* and the most closely related genus *Prauserella*. Bootstrap values ≥ 70% (based on 1,000 *pseudo*-replicates) are shown above the branches. Strains of *Saccharomonospora* were colored according to the group they belong to (blue for marine/lake, orange for moderately halophilic terrestrial, green for terrestrial and purple for clinical). Filled circles indicate clusters that were also recovered using maximum-parsimony and maximum-likelihood algorithms. *Nocardia brasiliensis* DSM 43758^T was used as an outgroup. Bar, 0.01 changes per nucleotide position.

pieces of one gene cluster are likely to phylogenetically clade together. Therefore, detection of PKS-derived ketosynthase domains (KS) and NRPS-derived condensation domains (C) with NaPDoS software, followed by alignment of the identified domain sequences with Muscle (Edgar, 2004) and phylogenetic reconstruction with Mega X software (Kumar et al., 2018) allow us to narrow down a more accurate number of PKS/NRPS/hybrid clusters present in fragmented next-generation sequencing assemblies. If a BGC was in the middle of a scaffold/contig (i.e., has sequence before and after the region antiSMASH identified),

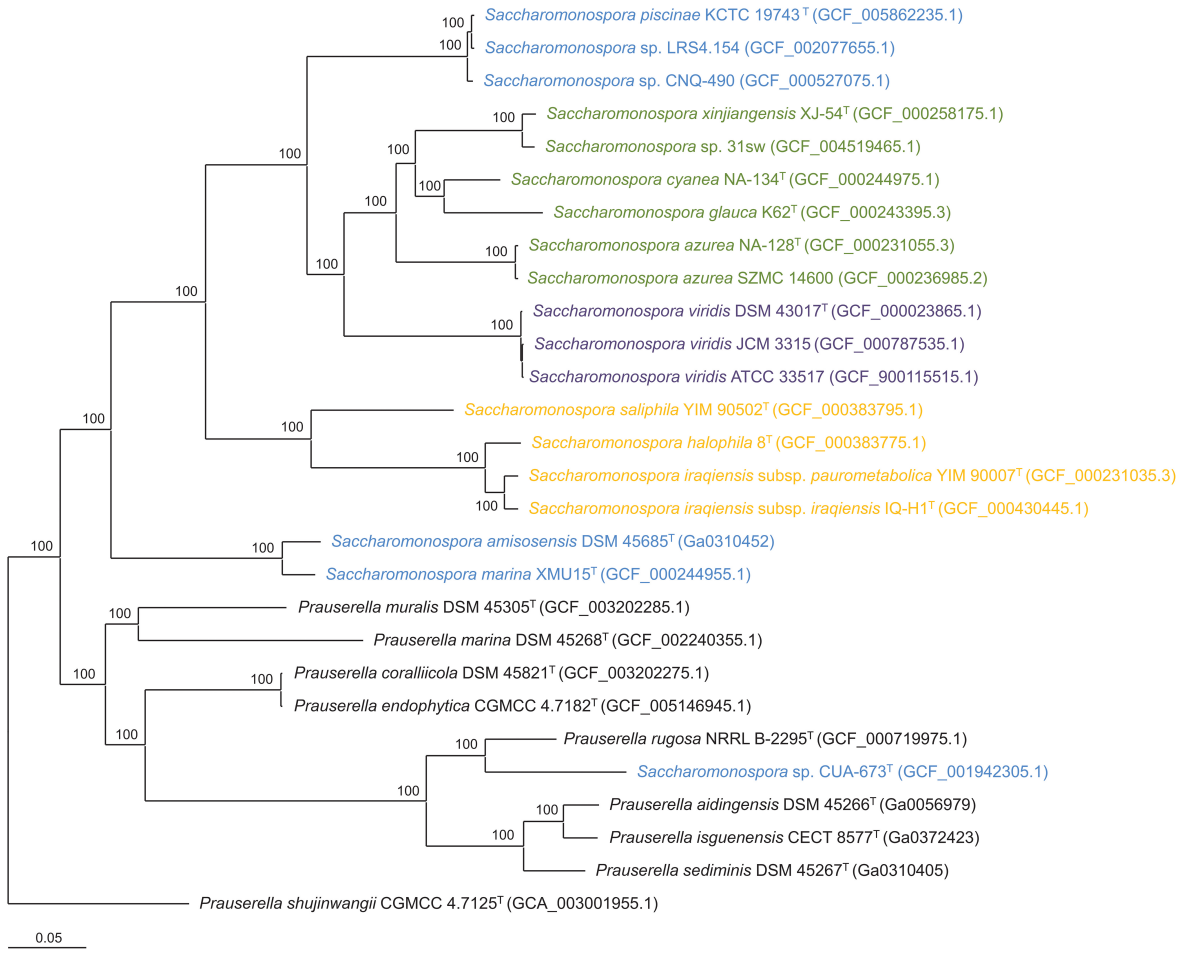


FIGURE 2 | Approximately maximum-likelihood phylogenomic tree based on the concatenation of the translated sequence of the 876 single-copy genes shared by the members of the genera *Saccharomonospora* and *Prauserella* under study. Bootstrap values ≥ 70% (based on Shimodaira-Hasegawa-like local support) are shown above the branches. Strains of *Saccharomonospora* were colored according to the group they belong to (blue for marine/lake, orange for moderately halophilic terrestrial, green for terrestrial and purple for clinical). Bar, 0.05 changes per nucleotide position.

it is considered complete; otherwise it was marked as a partial cluster. If KS or C domains on different scaffolds/contigs were sister taxa in the domain-based trees (**Supplementary Figure 1**), the BGCs on the two or more scaffolds/contigs were considered part of one cluster and the corresponding scaffolds/contigs were joined together.

The total length of the prospective BGC was also taken into consideration to determine the most probably number of PKS and NRPS clusters. For each genome, the sum of the lengths of all clusters (partial and complete) within a category (PKS or NRPS) was divided by the average length of all the complete clusters into this category. The resulting measure is the expected number of PKS or NRPS clusters based on an average length specific to each genome (**Supplementary Table 1**). These estimates support the joining of clusters using NaPDoS and domain-based phylogenetic trees (**Supplementary Figure 1**).

Circular display of the corrected antiSMASH results was carried out using Circos Table Viewer v.0.63-9 visualization software (Krzywinski et al., 2009).

In order to predict the association of the detected BGCs with the production of known or novel bioactive compounds, the PKS, NRPS and hybrid clusters matches localized with antiSMASH were further screened with Blast search within the MIBiG database (Kautsar et al., 2020).

Phenotypic Characterization

Comparative phenotypic features of strains *S. piscinae* KCTC 19743^T and *Saccharomonospora* sp. LRS4.154 were determined as described elsewhere (Tseng et al., 2018).

RESULTS

Saccharomonospora Strain Grouping

A total of 19 genome sequences affiliated to members of the genus *Saccharomonospora* were retrieved from NCBI GenBank (18 genomes) and from JGI (1 genome) databases, 12 of which belong to type strains of species or subspecies within this genus, and the

remaining seven were not affiliated to any particular species or subspecies (**Table 1**). The draft genome of *Saccharomonospora iraqiensis* subsp. *iraqiensis* IQ-H1^T was named as *Actinopolyspora iraqinesis* IQ-H1^T in GenBank database but it was also included in this study since both names are homotypic synonyms. Genome assembly varied between contig, scaffold, chromosome and complete genome level, with six out of the genomes assembled as a single scaffold or contig (**Table 1**). *Saccharomonospora* genomes have a size ranging from 3.69 Mb to 5.97 Mb, with a DNA G + C content of 67.3–71.3 mol% and contain between 3,656 and 5,664 protein-coding genes (**Table 1**).

The recovered genomes were grouped into four different sets according to their isolation source and to their NaCl requirements for growing (**Table 2**):

- Marine/lake: strains isolated from marine environments or lakes and with optimal growth at or lower than 5% (w/v) NaCl.
- Moderately halophilic terrestrial: strains recovered from terrestrial hypersaline habitats and with optimal growth around 10% (w/v) NaCl.
- Terrestrial: strains from soil samples and manure growing between 0 and 10% (w/v) NaCl.
- Clinical: strains of the species *S. viridis*, the sole taxon in the genus associated to diseases.

Taxophylogenomic Assessment

The 16S rRNA gene sequence analysis was employed to infer the evolutionary relationships among the members of the currently described species of the genus *Saccharomonospora*, as well as other *Saccharomonospora* strains unassigned to any of the existing species (**Figure 1**). The closest related genus *Prauserella* was also included as a reference. The two genera were clearly separated according to this 16S rRNA gene-based phylogeny. All the unassigned *Saccharomonospora* strains were grouped into one of the described species of *Saccharamonospora*, with the exception of the strain *Saccharomonospora* sp. CUA-673, which clustered with the type strain of *Prauserella rugosa* and, therefore, seems to be misnamed. Sequence similarities for reference non-type or unnamed strains of *Saccharomonospora* with respect to the most closely related type strain were always above 99.5%, that is, beyond the threshold value for circumscribing prokaryotic species (98.65%) (Kim et al., 2014). Again, the only exception was *Saccharomonospora* sp. CUA-673, which shared less than 96.0% 16S rRNA gene sequence similarity with respect to the *Saccharomonospora* type strains but showed 98.7% sequence similarity to the type strain of *Prauserella rugosa*, indicating that it is closely related to that species. Therefore, no putative novel species of *Saccharomonospora* might be proposed after 16S rRNA gene sequence phylogenetic analysis. On the other hand, some type strains that clustered together based on 16S rRNA gene phylogeny showed a sequence similarity above the mentioned threshold value, particularly, *S. amisosensis* vs *S. marina* and the group constituted by *S. iraqiensis* subsp. *iraqiensis*, *S. iraqiensis* subsp. *paurometabolica* and *S. halophila*. So, according to the 16S rRNA gene sequence results, the species within each of those clusters might be merged into a single taxon, but a more confident evolutionary study is required to confirm the 16S rRNA gene-based hypothesis.

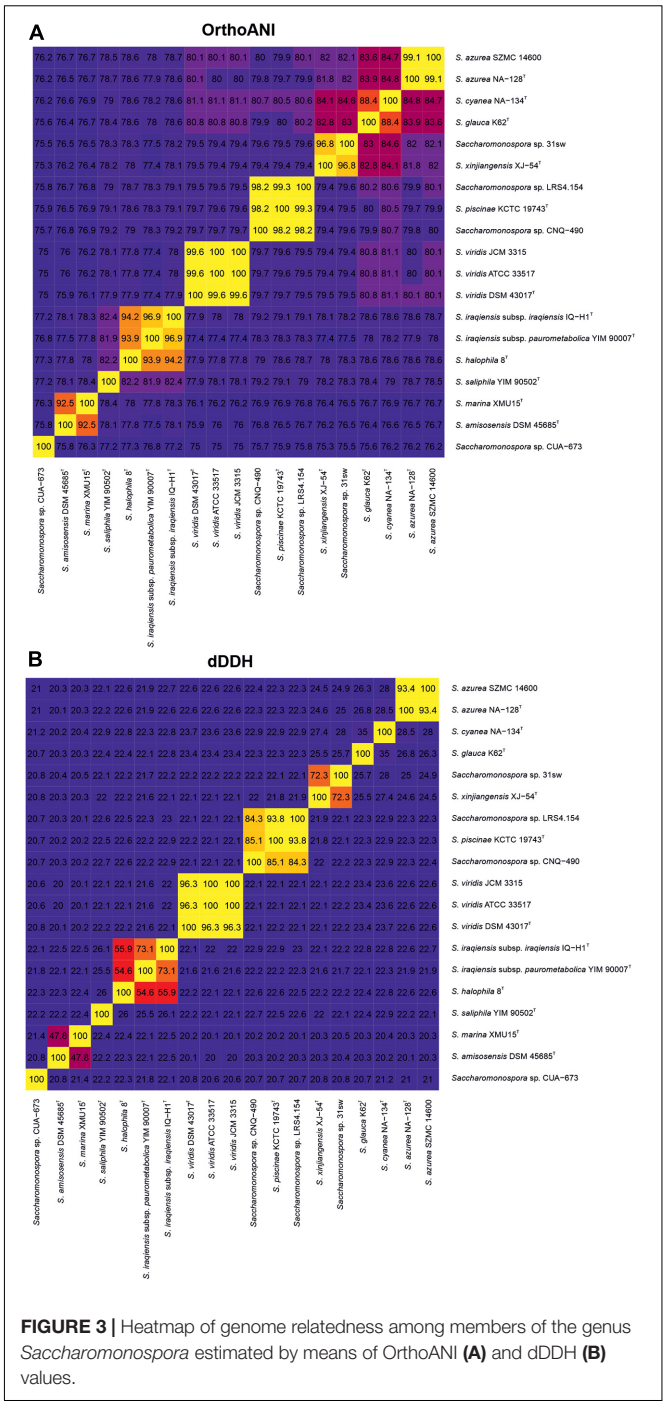


FIGURE 3 | Heatmap of genome relatedness among members of the genus *Saccharomonospora* estimated by means of OrthoANI **(A)** and dDDH **(B)** values.

Phylogenomic trees that more reliably infer the evolutionary relationships among taxa can be obtained using the genome sequence information. In this particular case, we employed 876 concatenated amino acid sequences of the orthologous single-copy genes present in all the genomes under study (**Figure 2**). Although the clusters obtained in both trees

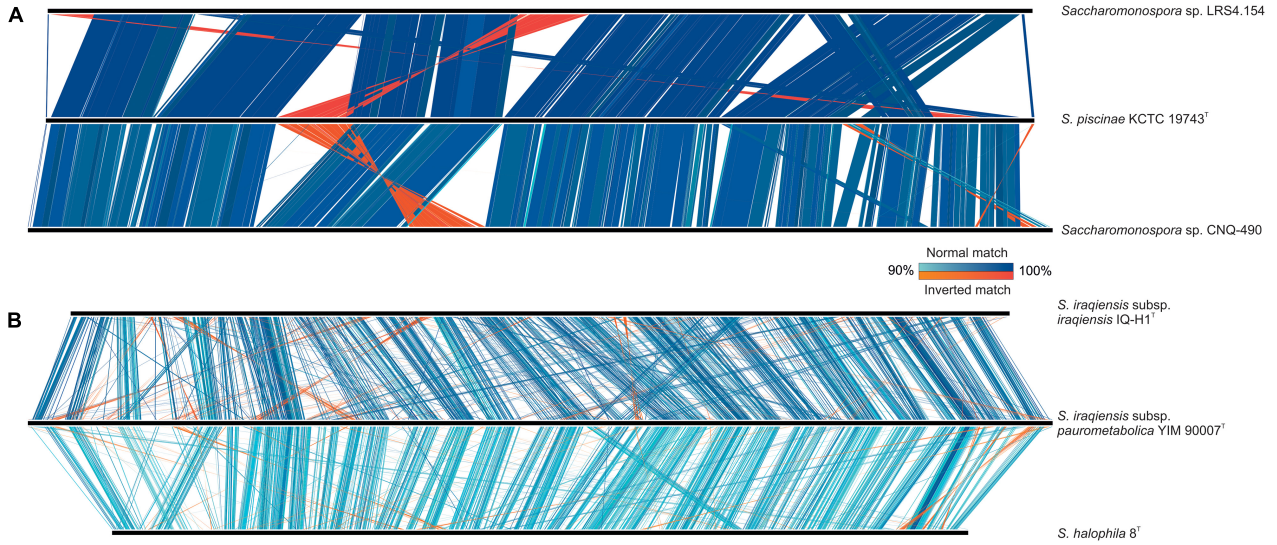


FIGURE 4 | Synteny plot among highly conserved **(A)** and partially conserved **(B)** genomes of *Saccharomonospora*. Only matches with ≥ 500 bp alignment length and ≥ 90% identity are shown.

were mostly in agreement, the phylogenomic tree showed a branch support of 100% in all bifurcations. This tree also points to a clear separation between *Saccharomonospora* and *Prauserella* genera, with probably no cryptic species of *Saccharomonospora* and with the need of transfer the strain *Saccharomonospora* sp. CUA-673 to the genus *Prauserella* (most likely as a new species given the long branch between this strain and the type strain of *Prauserella rugosa*). Similar to the 16S rRNA gene tree, the phylogenomic tree shows evidence for merging the species *S. amisosensis* – *S. marina*, on the one hand, and the species *S. iraqiensis* subsp. *iraqiensis* – *S. iraqiensis* subsp. *paurometabolica* – *S. halophila* on the other hand. Finally, according to the phylogenomic study, the species *Prauserella coralliicola* and *Prauserella endophytica* clustered very closely and most probably they constitute a single species, but this proposal is beyond the scope of the present manuscript.

We further analyzed the congruence between the clustering of *Saccharomonospora* strains in the phylogenetic/phylogenomic trees and the groups previously established based on the strain isolation source and NaCl requirements. Only clinical and moderately halophilic terrestrial isolates formed independent groups in the inferred phylogenies (**Figures 1**, **2**). In the case of the clinical strains clustering was expected since the three clinical strains belonged to the same species and, in fact, two of them, namely *S. viridis* ATCC 33517 and *S. viridis* JCM 3315, are equivalent strains with different culture collection number. More interesting seems the grouping of moderately halophilic strains under the same clade, which suggests an environment-driven evolution of those strains to thrive in such extreme conditions. Although terrestrial strains were not clustered as a monophyletic branch according to the 16S rRNA gene phylogeny, they formed a monophyletic group in the phylogenomic tree (excepting the genome sequence of *S. colocasiae*), hinting at a common origin of those strains from soils, contaminants, and manure sources.

Re-evaluation of the Species of *Saccharomonospora* Using Genome Sequence Distances

Similarity among genome sequences can be estimated by means of OGRI (Chun and Rainey, 2014). Although there are many algorithms to calculate OGRI values, the two most widely used for taxonomic purposes at species level are OrthoANI (Lee et al., 2016) and dDDH (Meier-Kolthoff et al., 2013). Gold standard for prokaryotic species delineation is still the wet-lab DNA-DNA hybridization (DDH), a widely considered tedious, laborious and error-prone method (Rosselló-Mora, 2006). A DDH cutoff value to unequivocally discriminate if two microbial genomes should be considered as the same or different species has been established at 70% (Wayne et al., 1987; Stackebrandt and Goebel, 1994). Nevertheless, OrthoANI and dDDH have been proposed as surrogates for DDH (Meier-Kolthoff et al., 2013; Lee et al., 2016), with well accepted threshold values for species delineation of approximately 95% (Konstantinidis and Tiedje, 2005a; Goris et al., 2007; Richter and Rossello-Mora, 2009; Chun and Rainey, 2014) and 70% (Auch et al., 2010), respectively. With regard to genus delineation, the most used OGRI value is AAI (Konstantinidis and Tiedje, 2005b), with a cutoff value about 65% (Konstantinidis et al., 2017). OrthoANI and dDDH values were calculated for all pairs of *Saccharomonospora* genomes (**Figure 3**) and for the pair *Saccharomonospora* sp. CUA-673 vs. *Prauserella rugosa* DSM 43194^T (OrthoANI 84.2%, dDDH 27.8%). Additionally, AAI values between members of *Saccharomonospora* and *Prauserella* were calculated.

As supported by phylogenetic/phylogenomic trees, AAI values indicated a good delineation of the genera *Saccharomonospora*

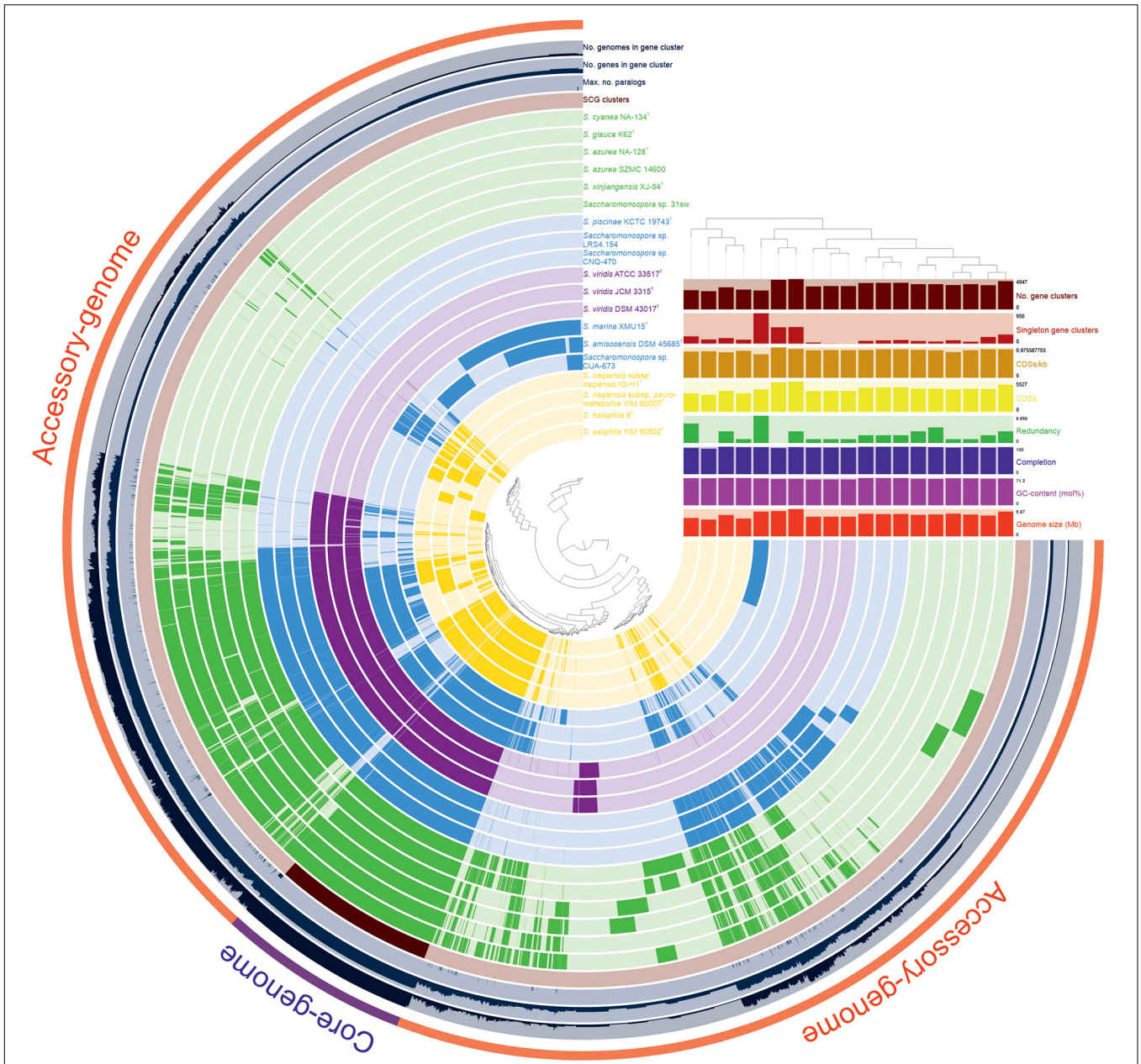


FIGURE 5 | Comparative pangenome display of the 19 *Saccharomonospora* strains analyzed in this study. The inner layers represent individual genomes organized regarding their isolation source and their NaCl requirements for growing. In the layers, dark colors indicate the presence of a gene cluster and light color its absence. The core- (978 genes) and the accessory- (11,297 genes) genomes are indicated in purple and read, respectively, in the outmost layer.

and *Prauserella* with only a small overlap. AAI values within the *Saccharomonospora* genomes (excluding that of strain CUA-673) ranged from 71 to 100%, while values between the two genera varied from 67 to 76%. According to AAI values, strain CUA-673 was clearly a member of *Prauserella*, showing 66–82% AAI values with members of this genus in comparison to 65-66% AAI values with member of *Saccharomonospora*. On the other hand, OrthoANI and dDDH values demonstrated that no putative new species exist among the studied genomes of *Saccharomonospora* besides the already mentioned strain *Saccharomonospora* sp. CUA-673, which should be renamed as *Praseurella* sp. CUA-673 and constitute a novel species within this genus that may be proposed after a detailed taxonomic polyphasic approach (Tindall et al., 2010). The reference non-type strains *S. azurea* SZMC 14600 and *S. viridis* ATCC 33517 (= JCM 3315) are properly named, according to the OGRI values obtained with respect to the type strains of the corresponding species. Furthermore, the strains *Saccharomonospora* sp. LRS4.154 and *Saccharomonospora* sp. CNQ-490 are clearly additional strains of *S. piscinae* and should be renamed accordingly. Similarly,

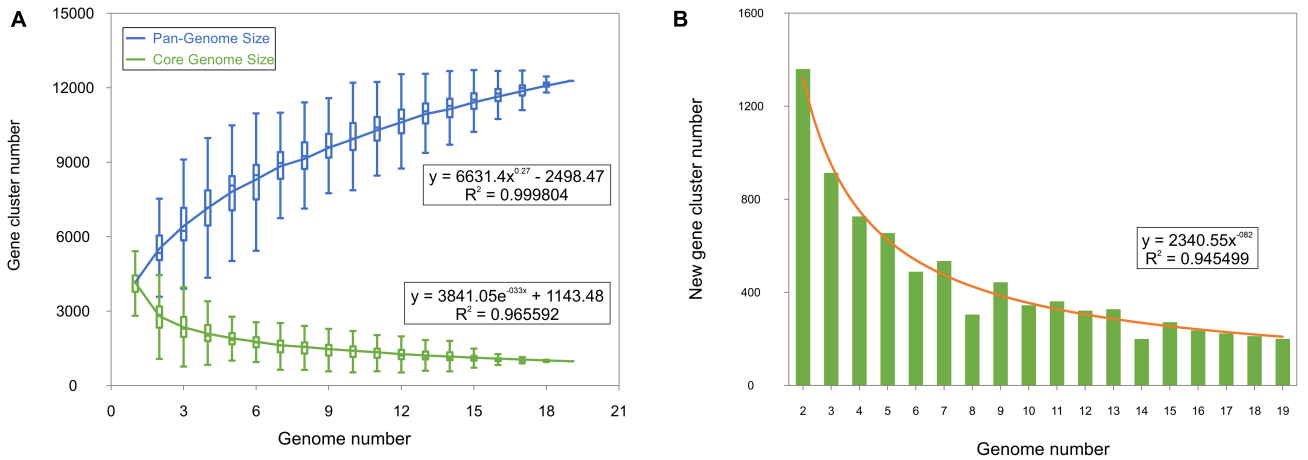


FIGURE 6 | Pan- and core-genome size evolution within the genus *Saccharomonospora*. **(A)** Gene accumulation curves of the pan-genome (blue) and the core-genome (green). The curve is the least squares fit of the power law for the average values. **(B)** Number of new genes and fit curve (orange) with an increase in the number of *Saccharomonospora* genomes.

the strain *Saccharomonospora* sp. 31sw can be considered unequivocally a member of *S. xinjiangensis*, but the dDDH value slightly above the 70% cutoff between that strain (*Saccharomonospora* sp. 31sw) and the type strain of the species (*S. xinjiangensis* XJ-54^T) and the 79–80% dDDH accepted value for subspecies boundaries (Meier-Kolthoff et al., 2014) suggest that strain *Saccharomonospora* sp. 31sw may constitute a novel separate subspecies within *S. xinjiangensis* taxon. Core-genome and 16S rRNA gene trees pointed to the need to merge the species *S. amisosensis* and *S. marina* into a single taxon, but OGRI analyses clearly rejected that idea. On the other hand, the taxonomic status of the group formed by *S. iraqiensis* subsp. *iraqiensis*, *S. iraqiensis* subsp. *paurometabolica* and *S. halophila* was not so clear since their OGRI results were close to the threshold values. *S. iraqiensis* subsp. *iraqiensis* and *S. iraqiensis* subsp. *paurometabolica* are doubtlessly different subspecies of the same species (supported by dDDH values of 73.1%, below the aforementioned subspecies cutoff). However, *S. halophila* fell within or close to the fuzzy zone [93–96% for OrthoANI (Rosselló-Móra and Amann, 2015), 60–70% for dDDH (Rosselló-Mora, 2006)] where the boundary of a species may not be clear. Therefore, *S. halophila* might be considered as a separate species or an additional novel subspecies of the species *S. iraqiensis*.

Spatial distribution of locally collinear blocks between two or more genomes (synteny) analysis can provide some clues about the evolutionary processes that lead to diversity, chromosomal dynamics, and rearrangement rates between species (Bhutkar et al., 2006). An approach to gain insight into the evolutionary distance between two species is to inspect the synteny of the genome sequences (Borriss et al., 2011). Here, we have applied the synteny study to elucidate suspicious clusters in the phylogenomic tree, particularly, the one formed by the two subspecies of *S. iraqiensis* and *S. halophila* 8^T, using the cluster including *S. piscinae* KCTC 19743^T, *Saccharomonospora* sp. LRS4.154 and *Saccharomonospora* sp. CNQ-490 as a reference for comparative purposes. The analysis revealed that the latter cluster possesses a very high level of synteny while the former was not so high (**Figure 4**). This might support the hypothesis that members of the second cluster are, indeed, members of the same species, while members of the first can be distinguished at species or subspecies level.

Going in depth in the *S. iraqiensis* subsp. *iraqiensis* – *S. iraqiensis* subsp. *paurometabolica* – *S. halophila* synteny analysis, although the genome coverage of locally collinear blocks is approximately the same between *S. iraqiensis* subsp. *iraqiensis* vs. *S. iraqiensis* subsp. *paurometabolica* and *S. iraqiensis* subsp. *iraqiensis* vs. *S. halophila*, the pairwise similarity between blocks is significantly higher for the pair *S. iraqiensis* subsp. *iraqiensis* – *S. iraqiensis* subsp. *paurometabolica* (**Figure 4**), endorsing the proposal to keep *S. halophila* as a different species while *S. iraqiensis* subsp. *iraqiensis* and *S. iraqiensis* subsp. *paurometabolica* would remain as subspecies of the same species.

Almost Closed Pan- and Core-genomes of *Saccharomonospora*

The pan-genome defines the entire genomic repertoire of a given group of microorganisms and encodes for all possible lifestyles of its organisms, including the core genome, dispensable genome and strain-specific genes (Vernikos et al., 2015). The 79,484 protein CDSs detected in the 19 analyzed genomes of *Saccharomonospora* were grouped into 8,467 orthologous gene clusters (978 core genes and 7,489 dispensable genes) and 3,808 singletons (strain-specific) gene clusters, with a pan-genome constituted of 12,275 gene clusters (**Figure 5**, **Supplementary Figure 2**). This wide range difference between the core- and pan-genomes might be due to the large genome size of actinobacteria and the existence of a large number of accessory genes in the *Saccharomonospora* genomes, some of them related to BGCs. In fact, BGCs were enriched in the accessory-genome in comparison to the core-genome, with an average of 71.1% of

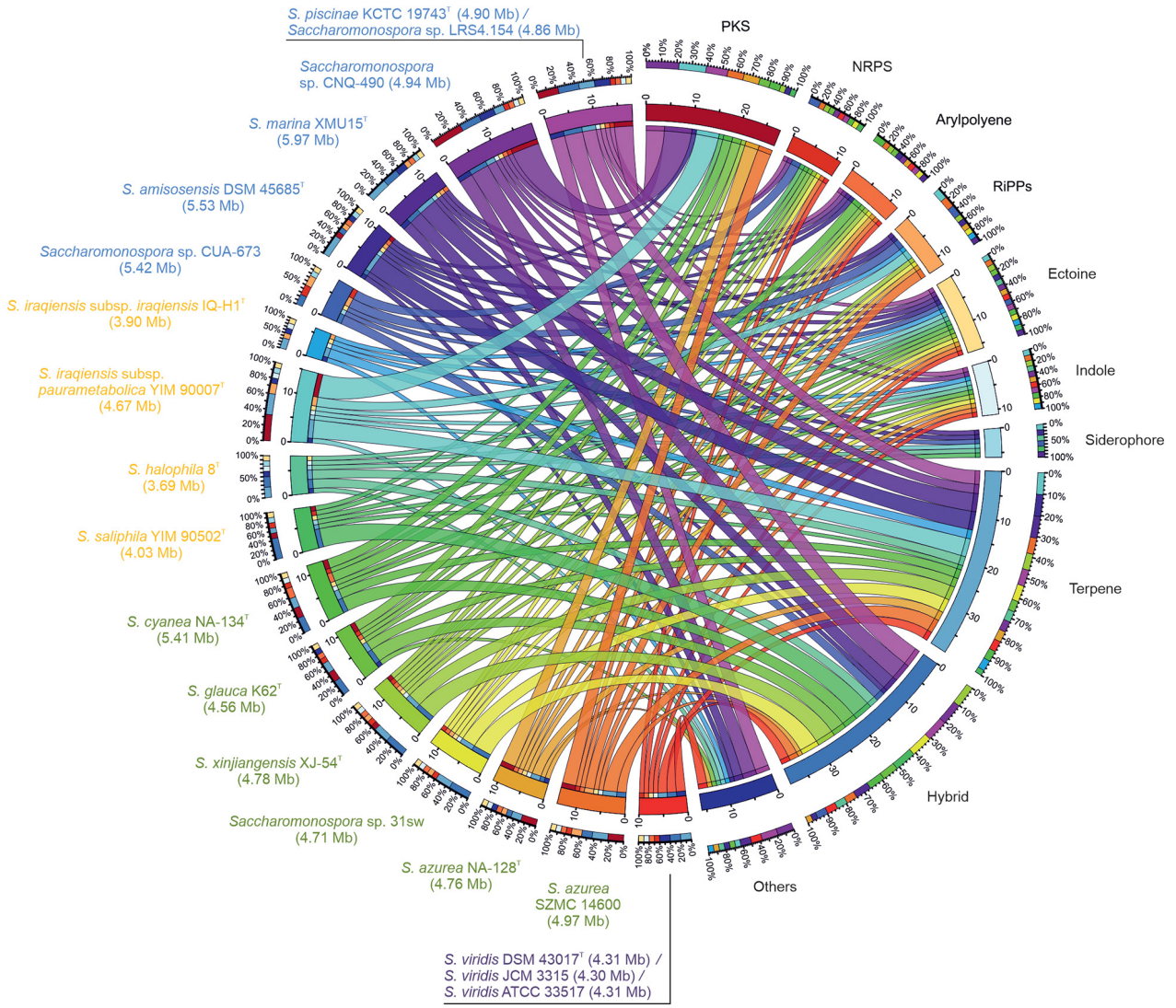


FIGURE 7 | Circular diagram of *Saccharomonospora* BGCs diversity. Genomes of strains with identical BGCs pattern are collapsed in a single band. Each genome is represented by a different colored band (left half of the circle) that can be traced from the organism to the types of BGCs found in that genome (right half of the circle). The width of these bands indicates the number of BGCs of that type. The cluster types are also assigned colors that make up the outer ring next to each genome to easily see what portion of each genome is assigned to a specific BGC. Conversely, the outer ring next to the BGC categories show the proportion of that BGC attributed to each genome represented by the genome color. The RiPP category includes clusters identified as bacteriocin, lanthipeptide or lassopeptide. The Hybrid category includes gene clusters constituted by two or more BGC types. The "Others" category includes uncommon BGCs found in less than five genomes, i.e., betalactone, homoserine lactone, ladderane, linaridin and oligosaccharide. Strains of *Saccharomonospora* were colored according to the group they belong to (blue for marine/lake, orange for moderately halophilic terrestrial, green for terrestrial and purple for clinical) and the genome size is indicated in parentheses.

BGC length spanned by accessory genes versus an average of 6.5% spanned by core genes.

Compared to other actinobacteria, studies focused on the genus *Streptomyces* identified 15,404 (Wu et al., 2017), 23,672 (Tidjani et al., 2019), and 34,592 (Kim et al., 2015) gene clusters conforming the pan-genome of 9, 11, and 17 members, respectively, in this genus, while the core-genome decreased from 5,047 (when considering 9 streptomycetes) to 2,018 (for 17 streptomycetes) gene clusters, suggesting an open pan- and core-genomes. Nevertheless, the evolution of the pan- and core-genomes of the 19 *Saccharomonospora* genomes rapidly reaches a plateau (**Figure 6**), indicating almost closed pan- and core-genomes, which means that most of the natural variation has been captured.

Secondary Metabolite Profiling

The genomes of members of the genus *Saccharomonospora* were analyzed for the presence of a broad array of secondary metabolites BGCs, including those encoding PKS, NRPS, terpenes, ectoine, indole, arylpolyene, siderophore and ribosomally synthesized and post-translationally modified peptides (RiPPs). The BGCs identified were further curated to

TABLE 3 | Putative synthesis of known bioactive compounds by PKS, NRPS or hybrid BGCs of strains of *Saccharomonospora* based on Blast search against MIBiG database.

Most similar known cluster	MIBiG accession number	*S. piscinae* KCTC 19743^T	*Saccharomonospora* sp. LRS4.154	*Saccharomonospora* sp. CNQ-490	*S. marina* XMU15^T	*S. amisosensis* DSM 45685^T	*Saccharomonospora* sp. CUA-673	*S. iraqiensis* subsp. *paurometabolica* YIM 90007^T	*S. halophila* 8^T	*S. saliphila* YIM 90502^T	*S. cyanea* NA-134^T	*S. glauca* K62^T	*S. xinjiangensis* XJ-54^T	*Saccharomonospora* sp. 31sw	*S. azurea* NA-128^T	*S. azurea* SZMC 14600	*S. viridis* DSM 43017^T	*S. viridis* JCM 3315	*S. viridis* ATCC 33517
Acarviostatin I03/Acarviostatin II03/Acarviostatin III03/Acarviostatin IV03	BGC0000804															18			
Alkil-O-dihydrogeranyl-methoxyhydro quinones	BGC0001077	57	57	57							57	42			57	28			
Allocyclinone	BGC0001500								10		13	13							
Amycolamycin A/Amycolamycin B	BGC0001503	10	10	10															
Anthracimycin	BGC0001300										13								
Apoptolidin	BGC0000021		33																
Arsenopolyketides	BGC0001283									45									
Clavunalic acid	BGC0000845																58	58	58
Collismycin A	BGC0000973	7	7	7															
Concanamycin A	BGC0000040	42												46					
Curamycin	BGC0000215	71	71	71												85			
Desosamine	BGC0000898											44							
Ebelactone	BGC0001580							11											
ECO-02301	BGC0000052							39											
Enduracidin	BGC0000341											4							
Ficellomycin	BGC0001593	14	14																
Friulimicin	BGC0000354				21														
Herboxidiene	BGC0001065			8								8							

(Continued)

TABLE 3 | Continued

Most similar known cluster	MIBiG accession number	*S. piscinae* KCTC 19743^T	*Saccharomonospora* sp. LRS4.154	*Saccharomonospora* sp. CNQ-490	*S. marina* XMU15^T	*S. amisosensis* DSM 45685^T	*Saccharomonospora* sp. CUA-673	*S. iraqiensis* subsp. *paurometabolica* YIM 90007^T	*S. halophila* 8^T	*S. saliphila* YIM 90502^T	*S. cyanea* NA-134^T	*S. glauca* K62^T	*S. xinjiangensis* XJ-54^T	*Saccharomonospora* sp. 31sw	*S. azurea* NA-128^T	*S. azurea* SZMC 14600	*S. viridis* DSM 43017^T	*S. viridis* JCM 3315	*S. viridis* ATCC 33517
Hygrocin A/Hygrocin B	BGC0000075			35											19				
Indanomycin	BGC0000079										13								
JBIR-100	BGC0001348												66						
Kedarcidin	BGC0000081													12					
Mediomycin A	BGC0001662	28	28																
Mirubactin	BGC0000392						71				71	71			71		71	71	71
Octacosamicin	BGC0001714							16											
Oxazolepoxidomycin A	BGC0001885															86			
Oxazolomycin	BGC0001106												21	31					
PM100117/PM100118	BGC0001359												31						
Primycin	BGC0001447														94	97			
Rabelomycin	BGC0000262									27									
Ralsolamycin	BGC0001754										40								
Retimycin	BGC0001228				10														
Salinichelins	BGC0001767												53	53					
Sanglifehrin A	BGC0001042							11											
Saprolmycin	BGC0001384						52												
Selvamicin	BGC0001773			18															
Skyllamycin	BGC0000429									24									
Sporolide	BGC0000150	36	36	36	51				44										
Taromycin	BGC0000439			96													50	46	50
Tiancimycin	BGC0001804					38													

Cell values indicate the percentage of genes within the MIBiG cluster showing similarity to the query sequence (color code: light orange, < 50% genes; green, 50–69% genes; dark green, 70–100% genes).

determine which PKS, NRPS and hybrid partial clusters most likely belonged to the same gene cluster. In total numbers, the most abundant non-hybrid BGCs predicted in the studied actinobacterial genomes were terpene, followed by PKS and ectoine, although the hybrid BGCs were even more plentiful (**Figure 7**, **Supplementary Tables 2, 3**). Marine/lake and terrestrial strains possess on average the higher amount of putative BGCs (14.5 and 12.2, respectively), while clinic and moderately halophilic terrestrial ones contain slightly less (10 and 9.5, respectively). *Saccharomonospora* sp. CNQ-490 followed by *S. piscinae* KCTC 19743^T and *Saccharomonospora* sp. LRS4.154 (all three belonging to the same species as aforementioned) were, theoretically, the most prolific secondary metabolites producers with the higher number of identified BGCs (**Figure 7**, **Supplementary Tables 2, 3**). There were some types of secondary metabolites able to be produced only by specific strains/species, i.e., homoserine lactone (in the closely related strains *S. piscinae* KCTC 19743^T, *Saccharomonospora* sp. LRS4.154 and *Saccharomonospora* sp. CNQ-490), lassopeptide (in the probably misnamed strain *Saccharomonospora* sp. CUA-673) and oligosaccharide (in the moderately halophilic strain *S. iraqiensis* subsp. *paurometabolica* YIM 90007^T). It is also remarkable that no PKS or NRPS clusters were predicted for *S. marina* XMU15^T, *S. iraqiensis* subsp. *iraqiensis* IQ-H1^T and *S. halophila* 8^T (**Figure 7**, **Supplementary Tables 1–3**).

Polyketide synthase, NRPS and hybrid PKS-NRPS clusters found by antiSMASH (Blin et al., 2019) were further searched using Blast tool against MIBiG database (Kautsar et al., 2020) in order to predict known and novel biomolecules that might putatively be synthesized by members of *Saccharomonospora*. In theory, BGCs for production of alkil-O-dihydrogeranyl-methoxyhydroquinones, mirubactin, sporolide, curamycin and taromycin are the most widely spread among the members of this genus, being present in four to seven genomes with at least 50% of genes showing similarity to the corresponding MIBiG cluster (**Table 3**). On the contrary, BGCs for the synthesis of JBIR-100, oxazolepoxidomycin A and saprolmycin could only be detected in one genome with more than 50% of genes showing similarity to a known MIBiG cluster. Other BGCs related to the production of acarviostatin I03/acarviostatin II03/acarviostatin III03/acarviostatin IV03, anthracimycin, apoptolidin, arsenopolyketides, desosamine, ebelactone, ECO-02301, enduracidin, friulimicin, indanomycin, kedarcidin, octacosamicin, PM100117/PM100118, rabelomycin, ralsolamycin, retimycin, sanglifehrin A, selvamicin, skyllamycin and tiancimycin were also present only in one genome, but sharing less than 50% of the genes with the registered MIBiG cluster (**Table 3**), which means that might be involved in the synthesis of a novel compound. Noticeably, for a specific type 1 PKS cluster identified in *Saccharomonospora* sp. CNQ-490, two NRPS and hybrid PKS-NRPS clusters in *Saccharomonospora* sp. CUA-673, and a type 3 PKS cluster in *S. iraqiensis* subsp. *paurometabolica* YIM 90007^T no matches to any predicted known product were found by cluster Blast search and, therefore, a more detailed study of these clusters as potential producers of new biomolecules might be interesting.

DISCUSSION

Comparative 16S rRNA gene sequence analysis has been widely used to infer relationships within the genus *Saccharomonospora* (Kim, 2015), but despite its success in providing the phylogenetic backbone for the classification of *Actinobacteria* in the past, this gene possesses a low evolutionary rate, which means only a limited number of differential characters and, thus, may yield trees with many statistically unsupported branches (Klenk and Göker, 2010; Breider et al., 2014). Furthermore, this gene might be present in multiple divergent copies within a single genome and might have suffered horizontal gene transfer yielding to mistaken phylogenies (Větrovský and Baldrian, 2013; Papke et al., 2015). For those reasons, the use of the concatenated sequence of several single-copy housekeeping genes (an approach called Multilocus Sequence Analysis, MLSA) with a higher resolution was proposed to overcome these limitations (Gevers et al., 2005), but the produced phylogenies were limited to a certain number of genes. With the decreasing cost of genome sequencing and the ever-increasing prokaryotic genomes in public databases, phylogenetic trees may be substituted by phylogenomic trees inferred based on a large number of core single-copy protein-coding genes (Eisen and Fraser, 2003). In this study we have compared the 16S rRNA gene-based phylogeny of the genus *Saccharomonospora* with the evolutionary relationships obtained by using a phylogenomic approach. A previous study inferring trees from genome-scale data within the phylum *Actinobacteria* has been carried out (Nouioui et al., 2018) but, on the one hand, it included only 11 genomes of *Saccharomonospora* and, secondly, the phylogenomic tree was constructed using whole proteomes (instead of translated single-copy core genes) and distance-based clustering methods, which are not as robust as approximately maximum-likelihood-based methods (Price et al., 2010). In any case, the topology of the phylogenomic tree regarding the genus *Saccharomonospora* were consistent between the study of Nouioui et al. (2018) and the present research (**Figure 2**).

Here, we have studied in more detail all the 19 available genomes of *Saccharomonospora* using an approximately maximum-likelihood-based approach. Our results showed that 16S rRNA gene-based phylogenies reliably reflect the relationships among members of *Saccharomonospora*, with almost identical topologies between the phylogenetic and the phylogenomic trees. Therefore, despite all the drawbacks that have been described for 16S rRNA single gene trees, this approach is still quite accurate for the actinobacterial genus *Saccharomonospora* besides the modest tree branch support. The use of genomic data allowed us to unequivocally assign the strains *Saccharomonospora* sp. LRS4.154 and *Saccharomonospora* sp. CNQ-490 to the species *S. piscinae* (which makes it necessary to emend the description of the species as indicated below), the strain *Saccharomonospora* sp. 31sw to the species *S. xinjiangensis* and the strain *Saccharomonospora* sp. CUA-673 to the closest genus *Prauserella* as a putative new species which might be formally proposed after a detailed polyphasic study (that could not be conducted here since the strain CUA-673 is not publicly available in any culture collection). We have also confirmed that the strains *S. azurea* SZMC 14600 and *S. viridis* ATCC

33517 (= JCM 3315) are properly named. The taxonomic status of *S. iraqiensis* subsp. *iraqiensis*, *S. iraqiensis* subsp. *paurometabolica* and *S. halophila* remains unclear after several studies. Tang et al. proposed the transference of *S. iraqiensis* subsp. *iraqiensis* IQ-H1^T (at that time *Actinopolyspora iraqiensis* IQ-H1^T) as a heterotypic synonym of *S. halophila* 8^T based on wet-lab DDH data (Tang et al., 2011). Later, using dDDH values and distance-based phylogenomic inference it was found that the strain IQ-H1^T was more closely related to *S. paurometabolica* YIM 90007^T and it was proposed the merger of both taxa as a single species, but separated as two different subspecies (Nouioui et al., 2018). Authors of this study considered that the name *Actinopolyspora iraqiensis* had priority over *S. paurometabolica*, which is true taking into account that the former name was validly published before the latter, and therefore they proposed the name *S. iraqiensis* to accommodate the two subspecies. Beyond nomenclature issues, the relationships among the three above mentioned taxa needed revision. In this study, we have hypothesized that *S. halophila* might constitute another subspecies of *S. iraqiensis* because of OrthoANI and dDDH values close to the transition zone of the species boundary. Nevertheless, synteny plots among those taxa demonstrated that locally collinear blocks possess higher similarity between *S. iraqiensis* subsp. *iraqiensis* and *S. iraqiensis* subsp. *paurometabolica* compared to those shared by *S. iraqiensis* subsp. *iraqiensis* and *S. halophila*. Therefore, our study supports a recent speciation event occurring in *S. halophila* or *S. iraqiensis* which may explain that the obtained OGRI values between both taxa fall close to the species threshold value. Finally, our analysis endorses the current status of the remaining species of *Saccharomonospora* as independent taxa.

In this study we also focused on the identification of biosynthetic gene clusters (BGCs) in the genome of the analyzed strains of *Saccharomonospora*. Only a few previous reports mining BGCs in this genus have been published so far (Yamanaka et al., 2014; Schorn et al., 2016; Reynolds et al., 2018; Kovács et al., 2020), which investigated the genomes of *Saccharomonospora* sp. CNQ-490, *Saccharomonospora* sp. CUA-673 (hereby putatively relocated to the genus *Prauserella*), and two strains of *S. azurea* in search of their secondary metabolic capacity. Compared to other genera of *Actinobacteria*, the genus *Saccharomonospora* showed a high degree of novelty and diversity of BGCs (Schorn et al., 2016). Our study, considering all the available *Saccharomonospora* genomes, supports the previous findings, with a wide BGC abundance in each genome ranging from 6 to 19 identified clusters corresponding to 18 different categories. Although it has been stated that the number and variety of BGC pathways generally increases as the size of the genome increases (Schorn et al., 2016), we have observed that this pattern does not properly fit to the genus *Saccharomonospora*, where the Pearson's correlation value between number of BGCs and genome size is rather weak (0.38). Moreover, strains belonging to the same species can contain large differences, such as *S. iraqiensis* subsp. *paurometabolica* YIM 90007^T which almost triples *S. iraqiensis* subsp. *iraqiensis* IQ-H1^T in amount of BGCs. The three strains pertaining to the species *S. piscinae* showed the highest abundance and variety of BGCs, with a total of 18 to 19 clusters corresponding to 11 different categories. However, attention must be paid not only to the abundance and diversity, but also to particular BGCs contained only in some strains. That is the case of the homoserine lactone cluster, occurring only in *S. piscinae* strains, or lassopeptide and oligosaccharide clusters, found exclusively in *Saccharomonospora* (*Prauserella*) sp. CUA-673 and *S. iraqiensis* subsp. *paurometabolica*, respectively.

Polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) are multi-domain megasynthases involved in the biosynthesis of a remarkable amount of biological active compounds clinically valuable as anti-microbial, anti-fungal, anti-parasitic, anti-tumor and immunosuppressive agents (Cane and Walsh, 1999) and, therefore, require special attention. Detailed analysis of PKS, NRPS and hybrid clusters, and their comparison to available databases may provide clues to predict the product that a microorganism can synthetize. BGC mining conducted in this research demonstrated that a total of 40 different putative bioactive products might be produced by PKS, NRPS and hybrid clusters of *Saccharomonospora* (**Table 3**). Some compounds can theoretically be made by several strains of this genus, while others are specific to a particular strain. Most important, several BGCs showed no match to any already discovered compound or had a low similarity to known molecules, in particular clusters related to the synthesis of acarviostatin I03/acarviostatin II03/acarviostatin III03/acarviostatin IV03, allocyclinone, amycolamycin A/amycolamycin B, anthracimycin, apoptolidin, arsenopolyketides, collismycin A, concanamycin A, desosamine, ebelactone, ECO-02301, enduracidin, ficellomycin, friulimicin, herboxidiene, hygrocin A/hygrocin B, indanomycin, kedarcidin, mediomycin A, octacosamicin, oxazolomycin, PM100117/PM100118, rabelomycin, ralsolamycin, retimycin, sanglifehrin A, selvamicin, skyllamycin and tiancimycin, which shared less than 50% of the genes with the BGCs in the MIBiG database (**Table 3**), and they are probably involved in the synthesis of new bioactive compounds with promising clinical and biotechnological applications. In spite of the large genomic potential of *Saccharomonospora* BGCs to synthetize bioactive molecules, a literature search shows that only a few bioactive compounds (primycin, taromycin, and saccharomonopyrones A-C) have been isolated so far from members of the genus *Saccharomonospora* (Yamanaka et al., 2014; Yim et al., 2017; Reynolds et al., 2018; Kovács et al., 2020) and, therefore, a more comprehensive screening to isolate more biomolecules that can be synthetized by members of this genus should be carried out. Primycin is known to be produced only by strains of the species *S. azurea* (Kovács et al., 2020), which coincides with our genomic data (**Table 3**). Taromycin is reported to be isolated from the marine *Saccharomonospora* sp. CNQ-490 (Yamanaka et al., 2014; Reynolds et al., 2018), although our genomic analysis indicated that members of the species *S. viridis* might also produce taromycin-like compounds (**Table 3**). Besides, the strain *Saccharomonospora* sp. CNQ-490 can yield three α-pyrones (namely saccharomonopyrones A-C) (Yim et al., 2017), but those biomolecules were not detected by the *in silico* analysis because they are not included into the MIBiG database.

CONCLUSION

Genome-scale data of *Saccharomonospora* strains can be used to elucidate evolutionary relationships among microorganisms of this genus and to unequivocally assign strains into species. Based on the results of this manuscript we, hereby, propose the emended description of the species *S. piscinae* as indicated below. This genus has been revealed as an important source of a wide variety of biosynthetic gene clusters, which may be responsible for the production of novel non-yet-discovered bioactive molecules with potential application to biotechnology and biomedicine.

Emended Description of *Saccharomonospora piscinae* Tseng et al. (2018)

The description is as given before (Tseng et al., 2018), with the following amendment. The aerial mycelium is pale green to greenish gray or black. Oval spores have smooth or rugose surface. Olive green soluble pigments may be produced. Halotolerant or moderate halophile. Growth occurs between 0 to 20% (w/v) NaCl and 20 to 55°C. Gelatin may be liquefied. Menaquinone MK-8(H_2) may be present as minor component. DNA G + C content is 70.6–71.0 mol% (genome).

Strains LRS4.154 (= CECT 9353 = DSM 105201) and CNQ-490, isolated from saline soil of Laguna El Rosario, Oaxaca, Mexico, and from a deep-sea sediment sample 2 km west of the Scripps pier, La Jolla, CA, United States, respectively, are additional reference strains of this species.

DATA AVAILABILITY STATEMENT

The genome sequences analyzed in this study can be found in the GenBank and Joint Genome Institute databases under the accession numbers shown in **Table 1**.

AUTHOR CONTRIBUTIONS

NR-D, RH, BV-G, CS-P, HS-T, and AV did the conceptualization. NR-D and RH did the methodology, formal analysis, data curation, writing-original draft preparation, and visualization. BV-G, CS-P, and SA-C did the validation. NR-D, RH, and CS-P did the investigation. NR-D, RH, and AV did the resources. NR-D, RH, BV-G, CS-P, SA-C, HS-T, and AV did the writing – review and editing. HS-T and AV did the supervision. CS-P and AV did the project administration. NR-D, CS-P, and AV did the funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by FEDER/Spanish Ministry of Science and Innovation-State Research Agency (project CGL2017-83385-P), the National Council of Science and Technology with a scholarship granted to N. Ramírez-Durán (application 2018-000007-01EXTV-00289) for a sabbatical stay at the University of Sevilla, the Secretary for Research and Advanced Studies of the Autonomous University of the State of Mexico (project 5009/2020CIB), and Junta de Andalucía (grants BIO-213 and US-1263771), Spain, which included FEDER funds.

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.2021.603791/full#supplementary-material

REFERENCES

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

Bérdy, J. (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. *J. Antibiot.* 65, 385–395. doi: 10.1038/ja.2012.27

Bhutkar, A., Russo, S., Smith, T. F., and Gelbart, W. M. (2006). Techniques for multi-genome synteny analysis to overcome assembly limitations. *Genome Inform.* 17, 152–161. doi: 10.11234/gi1990.17.2_152

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: 10.1093/nar/gkz310

Borriss, R., Rueckert, C., Blom, J., Bezuidt, O., Reva, O., and Klenk, H.-P. (2011). "Whole genome sequence comparisons in taxonomy," in *Methods in Microbiology. Taxonomy of Prokaryotes*, eds F. Rainey, and A. Oren, (London: Academic Press), 409–436. doi: 10.1016/B978-0-12-387730-7.00018-8

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

Cane, D. E., and Walsh, C. T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem. Biol.* 6, R319–R325. doi: 10.1016/s1074-5521(00)80001-0

Challis, G. L. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* 154, 1555–1569. doi: 10.1099/mic.0.2008/018523-0

Chen, D., Feng, J., Huang, L., Zhang, Q., Wu, J., Zhu, X., et al. (2014). Identification and characterization of a new erythromycin biosynthetic gene cluster in *Actinopolyspora erythraea* YIM90600, a novel erythronolide-producing halophilic actinomycete isolated from salt field. *PLoS One* 9:108129. doi: 10.1371/journal.pone.0108129

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

Corral, P., Amoozegar, M. A., and Ventosa, A. (2020). Halophiles and their biomolecules: recent advances and future applications in biomedicine. *Mar. Drugs* 18:E33. doi: 10.3390/md18010033

de la Haba, R. R., López-Hermoso, C., Sánchez-Porro, C., Konstantinidis, K. T., and Ventosa, A. (2019). Comparative genomics and phylogenomic analysis of the genus *Salinivibrio*. *Front. Microbiol.* 10:2104. doi: 10.3389/fmicb.2019.02104

Doroghazi, J. R., and Metcalf, W. W. (2013). Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* 14:611. doi: 10.1186/1471-2164-14-611

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

Eisen, J. A., and Fraser, C. M. (2003). Phylogenomics: intersection of evolution and genomics. *Science* 300, 1706–1707. doi: 10.1126/science.1086292

Eren, A. M., Esen, ÖC., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ.* 3:e1319. doi: 10.7717/peerj.1319

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). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Fischbach, M. A., and Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* 106, 3468–3496. doi: 10.1021/cr0503097

Fischbach, M. A., and Walsh, C. T. (2009). Antibiotics for emerging pathogens. *Science (80-.).* 325, 1089–1093. doi: 10.1126/science.1176667

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Biol.* 20, 406–416. doi: 10.1093/sysbio/20.4.406

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

Gosse, J. T., Ghosh, S., Sproule, A., Overy, D., Cheeptham, N., and Boddy, C. N. (2019). Whole genome sequencing and metabolomic study of cave *Streptomyces* isolates ICC1 and ICC4. *Front. Microbiol.* 10:1020. doi: 10.3389/fmicb.2019.01020

Greiner-Mai, E., Kroppenstedt, R. M., Korn-Wendisch, F., and Kutzner, H. J. (1987). Morphological and biochemical characterization and emended descriptions of thermophilic actinomycetes species. *Syst. Appl. Microbiol.* 9, 97–109. doi: 10.1016/S0723-2020(87)80062-0

Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., et al. (2018). RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860. doi: 10.1093/nar/gkx1068

Ishaque, N. M., Burgsdorf, I., Limlingan Malit, J. J., Saha, S., Teta, R., Ewe, D., et al. (2020). Isolation, genomic and metabolomic characterization of *Streptomyces tendae* VITAKN with quorum sensing inhibitory activity from Southern India. *Microorganisms* 8:121. doi: 10.3390/microorganisms8010121

Jukes, T. H., and Cantor, C. R. (1969). "Evolution of protein molecules," in *Mammalian Protein Metabolism*, Vol. III, ed. H. N. Munro, (New York, NY: Academic Press), 21–132. doi: 10.1016/B978-1-4832-3211-9.50009-7

Kamjam, M., Sivalingam, P., Deng, Z., and Hong, K. (2017). Deep sea actinomycetes and their secondary metabolites. *Front. Microbiol.* 8:760. doi: 10.3389/fmicb.2017.00760

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., et al. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48, D454–D458. doi: 10.1093/nar/gkz882

Kim, J.-N., Kim, Y., Jeong, Y., Roe, J.-H., Kim, B.-G., and Cho, B.-K. (2015). Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. *J. Microbiol. Biotechnol.* 25, 1599–1605. doi: 10.4014/jmb.1504.04008

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0

Kim, S. B. (2015). *Saccharomonospora* in *Bergey's Manual of Systematics of Archaea and Bacteria*. Chichester, UK: John Wiley & Sons, Ltd, 1–11. doi: 10.1002/9781118960608.gbm00185

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

Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi: 10.1038/ismej.2017.113

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

Kovács, M., Seffer, D., Pénzes-Hűvös, Á, Juhász, Á, Kerepesi, I., Csepregi, K., et al. (2020). Structural and functional comparison of *Saccharomonospora azurea* strains in terms of primycin producing ability. *World J. Microbiol. Biotechnol.* 36:160. doi: 10.1007/s11274-020-02935-x

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

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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.-J., Tang, S.-K., Stackebrandt, E., Kroppenstedt, R. M., Schumann, P., Xu, L.-H., et al. (2003). *Saccharomonospora paurometabolica* sp. nov., a moderately halophilic actinomycete isolated from soil in China. *Int. J. Syst. Evol. Microbiol.* 53, 1591–1594. doi: 10.1099/ijs.0.02633-0

Matsumoto, A., and Takahashi, Y. (2017). Endophytic actinomycetes: promising source of novel bioactive compounds. *J. Antibiot.* 70, 514–519. doi: 10.1038/ja.2017.20

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., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., et al. (2014). Complete genome sequence of DSM 30083^{T}, the type strain (U5/41^{T}) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand. Genomic Sci.* 9:2. doi: 10.1186/1944-3277-9-2

Nonomura, H., and Ohara, Y. (1971). Distribution of actinomycetes in soil. X. New genus and species of monosporic actinomycetes. *J. Ferment. Technol.* 49, 895–903.

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., et al. (2018). Genome-based taxonomic classification of the phylum *Actinobacteria*. *Front. Microbiol.* 9:2007. doi: 10.3389/fmicb.2018.02007

Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* 26, 449–457. doi: 10.1016/j.tig.2010.07.001

Papke, R. T., Corral, P., Ram-Mohan, N., de la Haba, R. R., Sánchez-Porro, C., Makkay, A., et al. (2015). Horizontal gene transfer, dispersal and haloarchaeal speciation. *Life* 5, 1405–1426. doi: 10.3390/life5021405

Passari, A. K., Chandra, P., Zothanpuia, Mishra, V. K., Leo, V. V., Gupta, V. K., et al. (2016). Detection of biosynthetic gene and phytohormone production by endophytic actinobacteria associated with *Solanum lycopersicum* and their plant-growth-promoting effect. *Res. Microbiol.* 167, 692–705. doi: 10.1016/j.resmic.2016.07.001

Passari, A. K., Mishra, V. K., Singh, G., Singh, P., Kumar, B., Gupta, V. K., et al. (2017). Insights into the functionality of endophytic actinobacteria with a focus on their biosynthetic potential and secondary metabolites production. *Sci. Rep.* 7:11809. doi: 10.1038/s41598-017-12235-4

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

Rangseekaew, P., and Pathom-aree, W. (2019). Cave actinobacteria as producers of bioactive metabolites. *Front. Microbiol.* 10:387. doi: 10.3389/fmicb.2019.00387

Reynolds, K. A., Luhavaya, H., Li, J., Dahesh, S., Nizet, V., Yamanaka, K., et al. (2018). Isolation and structure elucidation of lipopeptide antibiotic taromycin B from the activated taromycin biosynthetic gene cluster. *J. Antibiot.* 71, 333–338. doi: 10.1038/ja.2017.146

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

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

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

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*. Heidelberg: Springer, 23–50. doi: 10.1007/978-3-540-31292-5_2

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. doi: 10.1093/oxfordjournals.molbev.a040454

Schorn, M. A., Alanjary, M. M., Aguinaldo, K., Korobeynikov, A., Podell, S., Patin, N., et al. (2016). Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology* 162, 2075–2086. doi: 10.1099/mic.0.000386

Sharma, P., Kalita, M. C., and Thakur, D. (2016). Broad spectrum antimicrobial activity of forest-derived soil actinomycete, *Nocardia* sp. PB-52. *Front. Microbiol.* 7:347. doi: 10.3389/fmicb.2016.00347

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

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. Evol. Microbiol.* 44, 846–849. doi: 10.1099/00207713-44-4-846

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Syed, D. G., Tang, S.-K., Cai, M., Zhi, X.-Y., Agasar, D., Lee, J.-C., et al. (2008). *Saccharomonospora saliphila* sp. nov., a halophilic actinomycete from an Indian soil. *Int. J. Syst. Evol. Microbiol.* 58, 570–573. doi: 10.1099/ijs.0.65449-0

Tang, S.-K., Wang, Y., Klenk, H.-P., Shi, R., Lou, K., Zhang, Y.-J., et al. (2011). *Actinopolyspora alba* sp. nov. and *Actinopolyspora erythraea* sp. nov., isolated from a salt field, and reclassification of *Actinopolyspora iraqiensis* Ruan et al. 1994 as a heterotypic synonym of *Saccharomonospora halophila*. *Int. J. Syst. Evol. Microbiol.* 61, 1693–1698. doi: 10.1099/ijs.0.022319-0

Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., et al. (2019). Massive gene flux drives genome diversity between sympatric *Streptomyces* conspecifics. *mBio* 10, e1533–e1519. doi: 10.1128/mBio.01533-19

Tindall, B. J., Rosselló-Móra, 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

Tseng, M., Chiang, W.-P., Liao, H.-C., Hsieh, S.-Y., and Yuan, G.-F. (2018). *Saccharomonospora piscinae* sp. nov., a novel actinobacterium from fishpond sediment in Taiwan. *Int. J. Syst. Evol. Microbiol.* 68, 1418–1422. doi: 10.1099/ijsem.0.002653

Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. doi: 10.1016/j.mib.2014.11.016

Větrovský, T., and Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8:e57923. doi: 10.1371/journal.pone.0057923

Veyisoglu, A., Sazak, A., Cetin, D., Guven, K., and Sahin, N. (2013). *Saccharomonospora amisosensis* sp. nov., isolated from deep marine sediment. *Int. J. Syst. Evol. Microbiol.* 63, 3782–3786. doi: 10.1099/ijs.0.051516-0

Wang, P., Liu, Z., and Huang, Y. (2018). Complete genome sequence of soil actinobacteria *Streptomyces cavourensis* TJ430. *J. Basic Microbiol.* 58, 1083–1090. doi: 10.1002/jobm.201800181

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

Wu, H., Liu, W., Shi, L., Si, K., Liu, T., Dong, D., et al. (2017). Comparative genomic and regulatory analyses of natamycin production of *Streptomyces lydicus* A02. *Sci. Rep.* 7:9114. doi: 10.1038/s41598-017-09532-3

Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S., Gonzalez, D. J., Nizet, V., et al. (2014). Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1957–1962. doi: 10.1073/pnas.1319584111

Yim, C.-Y., Le, T. C., Lee, T. G., Yang, I., Choi, H., Lee, J., et al. (2017). Saccharomonopyrones A-C, new α-pyrones from a marine sediment-derived bacterium *Saccharomonospora* sp. CNQ-490. *Mar. Drugs* 15:239. doi: 10.3390/md15080239

Yoon, S.-H., Ha, S.-M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Zhang, C., Ding, W., Qin, X., and Ju, J. (2019). Genome sequencing of *Streptomyces olivaceus* SCSIO T05 and activated production of lobophorin CR4 via metabolic engineering and genome mining. *Mar. Drugs* 17:593. doi: 10.3390/md17100593

Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., et al. (2014). PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* 30, 1297–1299. doi: 10.1093/bioinformatics/btu017

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7:34064. doi: 10.1371/journal.pone.0034064

Zothanpuia, Passari, A. K., Leo, V. V., Chandra, P., Kumar, B., Nayak, C., et al. (2018). Bioprospection of actinobacteria derived from freshwater sediments for their potential to produce antimicrobial compounds. *Microb. Cell Fact.* 17:68. doi: 10.1186/s12934-018-0912-0

Conflict of Interest: 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 © 2021 Ramírez-Durán, de la Haba, Vera-Gargallo, Sánchez-Porro, Alonso-Carmona, Sandoval-Trujillo 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: 15 March 2021
doi: 10.3389/fmicb.2021.639396

Methylobacterium ajmalii sp. nov., Isolated From the International Space Station

Swati Bijlani[1], Nitin K. Singh[2], V. V. Ramprasad Eedara[3], Appa Rao Podile[3], Christopher E. Mason[4], Clay C. C. Wang[1] and Kasthuri Venkateswaran[2]**

[1] Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, CA, United States, [2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, [3] Department of Plant Science, School of Life Sciences, University of Hyderabad, Hyderabad, India, [4] WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, New York, NY, United States

Four strains belonging to the family of *Methylobacteriaceae* were isolated from different locations on the International Space Station (ISS) across two consecutive flights. Of these, three were identified as Gram-negative, rod-shaped, catalase-positive, oxidase-positive, motile bacteria, designated as IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5, whereas the fourth was identified as *Methylorubrum rhodesianum*. The sequence similarity of these three ISS strains, designated as IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5, was <99.4% for 16S rRNA genes and <97.3% for *gyrB* gene, with the closest being *Methylobacterium indicum* SE2.11^T. Furthermore, the multi-locus sequence analysis placed these three ISS strains in the same clade of *M. indicum.* The average nucleotide identity (ANI) values of these three ISS strains were <93% and digital DNA-DNA hybridization (dDDH) values were <46.4% with any described *Methylobacterium* species. Based on the ANI and dDDH analyses, these three ISS strains were considered as novel species belonging to the genus *Methylobacterium.* The three ISS strains showed 100% ANI similarity and dDDH values with each other, indicating that these three ISS strains, isolated during various flights and from different locations, belong to the same species. These three ISS strains were found to grow optimally at temperatures from 25 to 30°C, pH 6.0 to 8.0, and NaCl 0 to 1%. Phenotypically, these three ISS strains resemble *M. aquaticum* and *M. terrae* since they assimilate similar sugars as sole carbon substrate when compared to other *Methylobacterium* species. Fatty acid analysis showed that the major fatty acid produced by the ISS strains are $C_{18:1}-\omega7c$ and $C_{18:1}-\omega6c$. The predominant quinone was ubiquinone 10, and the major polar lipids were diphosphatidylglycerol, phosphatidylcholine, phosphatidylethanolamine, phosphatidylglycerol, and an unidentified lipid. Therefore, based on genomic, phylogenetic, biochemical, and fatty acid analyses, strains IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5, are assigned to a novel species within the genus *Methylobacterium,* and the name *Methylobacterium ajmalii* sp. nov. is proposed. The type strain is IF7SW-B2^T (NRRL B-65601^T and LMG 32165^T).

Keywords: *Methylobacterium*, polyphasic taxomony, ANI, international space station (ISS), whole genome sequencing

OPEN ACCESS

Edited by:
Brian P. Hedlund, University of Nevada, Las Vegas, United States

Reviewed by:
Jian-Yu Jiao, Sun Yat-sen University, China
En-Min Zhou, Sun Yat-sen University, China

****Correspondence:***
Clay C. C. Wang
clayw@usc.edu
Kasthuri Venkateswaran
kjvenkat@jpl.nasa.gov

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

Received: *08 December 2020*
Accepted: *22 February 2021*
Published: *15 March 2021*

Citation:
Bijlani S, Singh NK, Eedara VVR, Podile AR, Mason CE, Wang CCC and Venkateswaran K (2021) Methylobacterium ajmalii sp. nov., Isolated From the International Space Station. Front. Microbiol. 12:639396. doi: 10.3389/fmicb.2021.639396

INTRODUCTION

The genus *Methylobacterium* contains more species than any other genera within the family *Methylobacteriaceae*, order *Rhizobiales*, and class *Alphaproteobacteria* (Kelly et al., 2014). *Methylobacterium* species are Gram-negative, rod-shaped bacteria. The genus was first proposed by Patt et al. (1976) with *Methylobacterium organophilum* as the type species. The genus *Methylobacterium* was first emended to include facultative methylotrophs that have the ability to grow on methane or methanol as the source of carbon and energy, in addition to sugars and organic acids (Patt et al., 1976). Another taxonomic study classified all other previously known pink-pigmented facultative methylotrophic bacteria under the genus *Methylobacterium* (Green and Bousfield, 1982). Thereafter, 11 species from the genus *Methylobacterium* were redefined into a new genus proposed as *Methylorubrum*, based on 16S rRNA gene sequence, multi-locus sequence analysis (MLSA), genomic, and phenotypic data (Green and Ardley, 2018).

The genus *Methylobacterium* consists of 45 recognized species, which are ubiquitously present in a wide variety of habitats including air, soil, freshwater, and sediments, and can exist either in free-form or associated with plant tissues (Gallego et al., 2005a,b; Kang et al., 2007; Veyisoglu et al., 2013; Kelly et al., 2014; Kwak et al., 2014; Chaudhry et al., 2016; Green and Ardley, 2018; Park et al., 2018). *Methylobacterium* species are involved in nitrogen fixation, phosphate solubilization, abiotic stress tolerance, plant growth promotion, and biocontrol activity against plant pathogens (Madhaiyan et al., 2006; Kumar M. et al., 2016; Parasuraman et al., 2019; Grossi et al., 2020; Krug et al., 2020). For instance, a novel *Methylobacterium* sp. 2A was observed to result in higher density of lateral roots in inoculated potato crops, even under salt stress conditions, compared with control plants that were not inoculated with the bacteria; it was also found to exhibit biocontrol activity against several plant pathogens (Grossi et al., 2020). Furthermore, genomic analysis of *Methylobacterium* sp. 2A revealed the presence of metabolic pathways involved in plant growth promotion, including the genes for producing an auxin, 3-indole acetic acid (Grossi et al., 2020).

In an ongoing Microbial Tracking experiment on the International Space Station (ISS), four strains belonging to the family *Methylobacteriaceae* were isolated (Checinska Sielaff et al., 2019). Some of the *Methylobacterium* species that are phylogenetically related to these ISS strains have been isolated from plant sources (Kang et al., 2007; Chaudhry et al., 2016), indicating that the ISS strains might also display properties related to plant growth promotion. The objectives of this study were to generate whole genome sequences (WGS) and define the phylogenetic novelty of the ISS *Methylobacterium* strains using polyphasic taxonomic analyses. The WGS generated and annotated in this study was used to predict biotechnologically useful genetic determinants.

MATERIALS AND METHODS

Sample Collection and Isolation of Bacteria

Several surface samples (1 m^2) were collected from the ISS during Microbial Tracking–1 flight experiments from 2015 to 2016. Sample collection, processing, and isolation of cultivable microorganisms were published elsewhere (Checinska Sielaff et al., 2019). Briefly, the polyester wipes used to collect samples and particulates associated with the sampling devices were transported to Earth before being disassociated into sterile phosphate-buffered saline (pH 7.4) solution and plated onto R2A agar medium (Checinska et al., 2015; Checinska Sielaff et al., 2019). The microbial cultures that were grown at 25°C for 7 days were picked from the R2A plates, purified, and stored for further analyses. Distinct colonies (n = 4) isolated from three different locations and from a high-efficiency particulate arrestance (HEPA) filter were characterized during this study. These colonies exhibited unique coloration and differential genomic phylogeny. The type strain IF7SW-B2^T was isolated during Flight 1 (March 2015) at Location #7, the Overhead-3 panel surface of the Materials Science Research Rack 1, which is used for basic materials research in the microgravity environment of the ISS. The second strain, IIF1SW-B5, was isolated during Flight 2 (May 2015) at Location #1, the Port panel of the Cupola. The Cupola is a small module devoted to the observation of operations outside the ISS, such as robotic activities, spacecraft approaches, and extravehicular activities. The third strain, IIF4SW-B5, was isolated during Flight 2 (May 2015) at Location #4, the surface of the dining table. Even though the main function of the table was for dining, crewmembers also used the table for experimental work. The fourth strain was I1-R3, isolated from the ISS HEPA filter that was returned aboard STS-134/ULF6 in May 2011 and archived as reported earlier (Checinska et al., 2015).

DNA Extraction and Whole Genome Sequencing Analysis

A biomass of approximately 1 μg wet weight was collected for DNA extraction from each strain after growing on R2A medium at 25°C for 3 days. Total nucleic acid extraction was carried out using ZymoBIOMICS 96 MagBead DNA kit (lysis tubes) (Zymo Research, United States) after bead beating with a Bertin Precellys homogenizer. This was followed by library preparation using the Illumina Nextera Flex Protocol as per Illumina document number 1000000025416 v07. The initial amount of DNA for library preparation was quantified, and 5 to 12 cycles of polymerase chain reaction (PCR) amplification were carried out to normalize the output depending on the input DNA concentration. The amplified genomic DNA fragments were indexed and pooled in 384-plex configuration. Whole-genome shotgun sequencing was performed on a NovaSeq 6000 S4 flowcell PE 2 × 150 platform with a paired-end module. The data were filtered with NGS QC Toolkit v2.3 (Patel and Jain, 2012) for high-quality (HQ) vector- and adaptor-free reads for genome

assembly (cutoff read length for HQ, 80%; cutoff quality score, 20). The number of filtered reads obtained were used for assembly with SPAdes 3.14.0 (Bankevich et al., 2012) genome assembler (*k*-mer size- 32 to 72 bases) using default parameters. The genome was annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline 4.11 (Tatusova et al., 2016; Haft et al., 2018). In addition, functional annotation of genome and seed categories were assigned to the genome by implementing the Rapid Annotations using Subsystems Technology (RAST) tool (Aziz et al., 2008).

Genomes of all other strains used in this study were downloaded from NCBI, and the genomic relatedness of ISS strains was identified based on average nucleotide identity (ANI; FastANI) calculations (Jain et al., 2018) and digital DNA-DNA hybridization (dDDH) analysis (Meier-Kolthoff et al., 2013). FastANI was run on all the genomes using the default parameters: Mashmap identity cutoff I_0 = 80%, non-overlapping fragments of size l = *3Kb,* and minimum count of reciprocal mappings τ = 50.

Phylogenetic Analysis

Phylogenetic analysis was carried out based on 16S rRNA gene sequencing, and MLSA using six housekeeping genes: ATP synthase F1 beta subunit (*atpD*), DNA strand exchange and recombination gene (*recA*), chaperone gene (*dnaK*), DNA-directed RNA polymerase subunit beta (*rpoB*), glutamine synthetase type I (*glnI*), and DNA gyrase subunit B (*gyrB*), for differentiating *Methylobacterium* species (Green and Ardley, 2018). The 16S rRNA gene sequences of type strains of all 45 *Methylobacterium* species were included in the phylogenetic analysis. In addition, representative species of genus *Methylorubrum*, *Enterovirga*, *Microvirga*, and *Neomegalonema* from family *Methylobacteriaceae*, *Rhizobium* from order *Rhizobiales*, *Caulobacter* from order *Caulobacterales*, in class *Alphaproteobacteria* were included. *Pseudomonas aeruginosa* was selected as the outgroup.

The 16S rRNA gene sequences of all strains were retrieved from NCBI except for the four ISS strains, which were recovered from their respective WGS. Phylogenetic analysis based on housekeeping genes and MLSA was carried out with type strains of 24 *Methylobacterium* species and representative species of other genera. All the gene sequences were retrieved from the genome sequences using RAST v2.0[1] (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015). The individual gene sequences for all strains were aligned separately using ClustalW, and then the maximum likelihood tree was generated using MEGA 7.0.26 (Kumar S. et al., 2016). For MLSA, six housekeeping gene sequences for each strain were concatenated manually and aligned using ClustalW, and then the maximum likelihood tree was generated using MEGA 7.0.26 (Kumar S. et al., 2016).

The genome-based tree for the *Methylobacterium* species, including ISS strains and representative species of other genus with available WGS, was constructed using GToTree (Lee, 2019). This tool takes the complete/draft genomes as input and creates a phylogenomic tree based on the prespecified single-copy gene set using a hidden Markov model (HMM); the tool currently has 2,044 unique HMM set as identifiers to cover all three domains of microbial life.

Phenotypic Characterization of ISS Strains

Phenotypic characterization was performed according to standard protocols (Jones, 1981). Growth of the ISS strains at different temperatures (7, 25, 30, 37, and 45°C) was assessed after incubation on nutrient agar (Sigma, United States) for 7 days. Growth at different pH (4.0–10.0 at intervals of 1.0) was assessed after incubation in nutrient broth (Sigma, United States) at 30°C for 7 days. The pH of the nutrient medium was adjusted using citrate/NaH_2PO_4 buffer (pH 4.0–5.0), phosphate buffer (pH 6.0–8.0), and tris buffer (pH 9.0–10.0) (Kim et al., 2019). Salt tolerance was tested by streaking the strains on R2A supplemented with NaCl (0–10% at intervals of 1%) and incubating the plates at 30°C for 7 days. Motility was assessed via the "hanging drop" method by observing the culture under a light microscope (Tindall et al., 2007). Catalase activity was tested by adding 3% hydrogen peroxide to culture grown on R2A at 30°C for 7 days, and effervescence was monitored (Tindall et al., 2007). An oxidase test was carried out in a filter paper soaked with the substrate tetramethyl-p-phenylenediamine dihydrochloride, and coloration was documented (JurtshukJr., and McQuitty, 1976). All other physiological and biochemical tests were carried out using API 20 NE, API 50 CH, and API ZYM kits as per manufacturer's procedures (bioMérieux, France).

Chemotaxonomic Analysis

All strains grown in the R2A broth were harvested when growth of the cultures reached around 70% of the maximal optical density (exponential growth phase), and then the cultures were used for analyses of cellular fatty acids, polar lipids, and quinones, which were carried out as described previously (Ramaprasad et al., 2015). Briefly, for cellular fatty acids analysis, 40 mg of bacterial cell pellet from each strain was subjected to a series of four different reagents followed by saponification and methylation of fatty acids, thus enabling their cleavage from lipids. The fatty acid methyl esters (FAME) thus obtained were analyzed by a gas chromatograph equipped with Sherlock MIS software (Microbial ID; MIDI 6.0 version; Agilent: 6850)[2]. The peaks obtained were then labeled, and the equivalent chain length (ECL) values were computed by the Sherlock software.

The polar lipids profile was analyzed by extracting cells with methanol-chloroform-saline (2:1:0.8, v/v/v) from 1 g of freeze-dried bacterial cells. Separation of lipids was performed by two-dimensional chromatography on a silica gel thin-layer chromatography plate (Kieselgel 60 F254; Merck) using chloroform-methanol-water (75:32:4, v/v/v) in the first dimension and chloroform–methanol–acetic acid–water (86:16:15:4, v/v/v/v) in the second dimension. The total polar lipids profile was detected by spraying with 6% ethanolic molybdophosphoric acid.

[1]https://rast.nmpdr.org/

[2]http://midi-inc.com

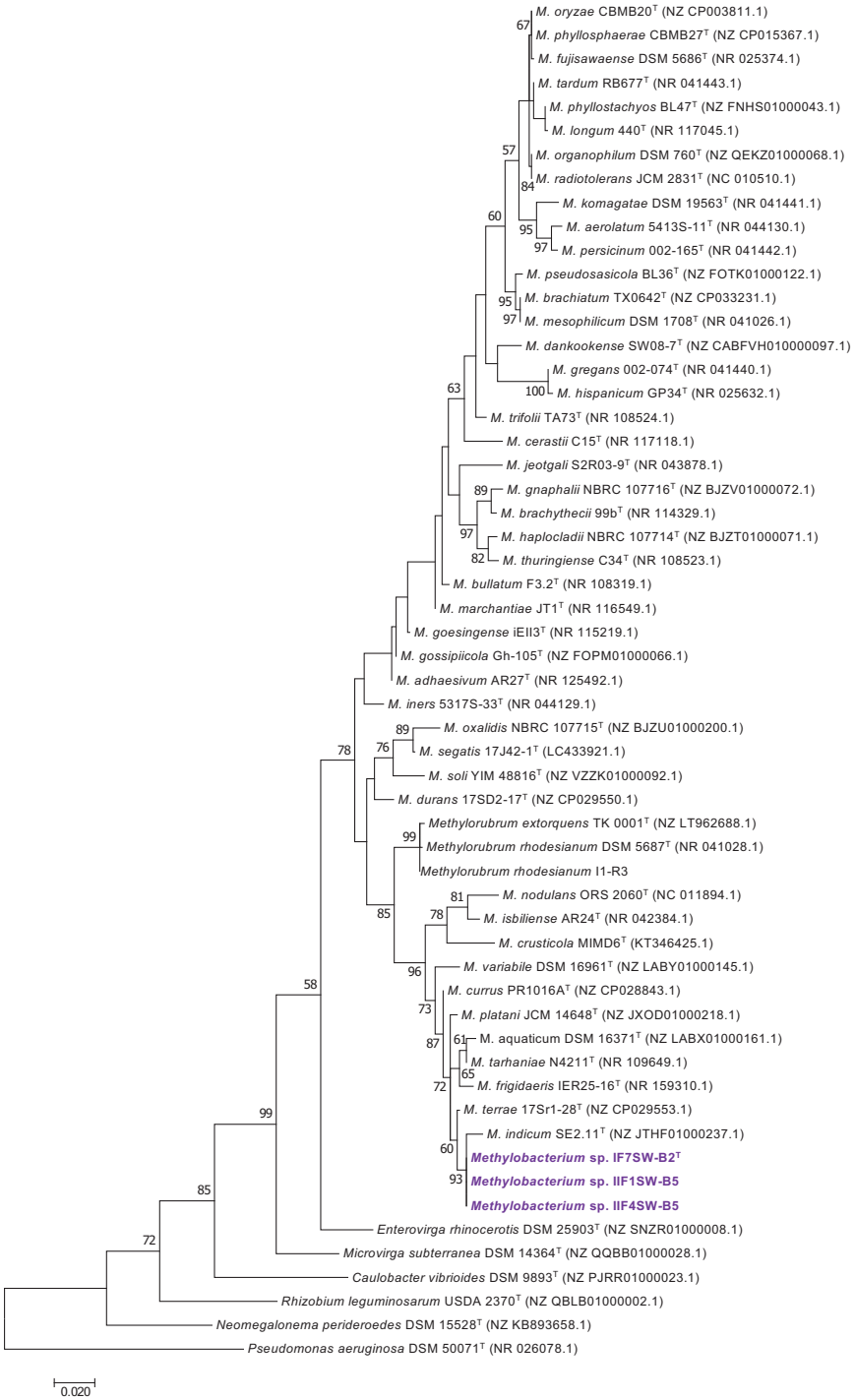


FIGURE 1 | Maximum likelihood phylogenetic tree based on 16S rRNA gene sequences shows the relationship of *Methylobacterium ajmalii* sp. nov. with members of the family *Methylobacteriaceae*. Bootstrap values from 1,000 replications are shown at branch points. Bar, 0.02 substitution per site.

The respiratory isoprenoid quinone was extracted with a chloroform-methanol mixture (2:1, v/v), evaporated under vacuum, re-extracted with acetone, and analyzed using high-performance lipid chromatography as per established methods (Ramaprasad et al., 2018).

RESULTS AND DISCUSSION

This study reports the isolation and identification of four strains belonging to the family *Methylobacteriaceae*, collected from different locations on the ISS. Three of the strains, referred

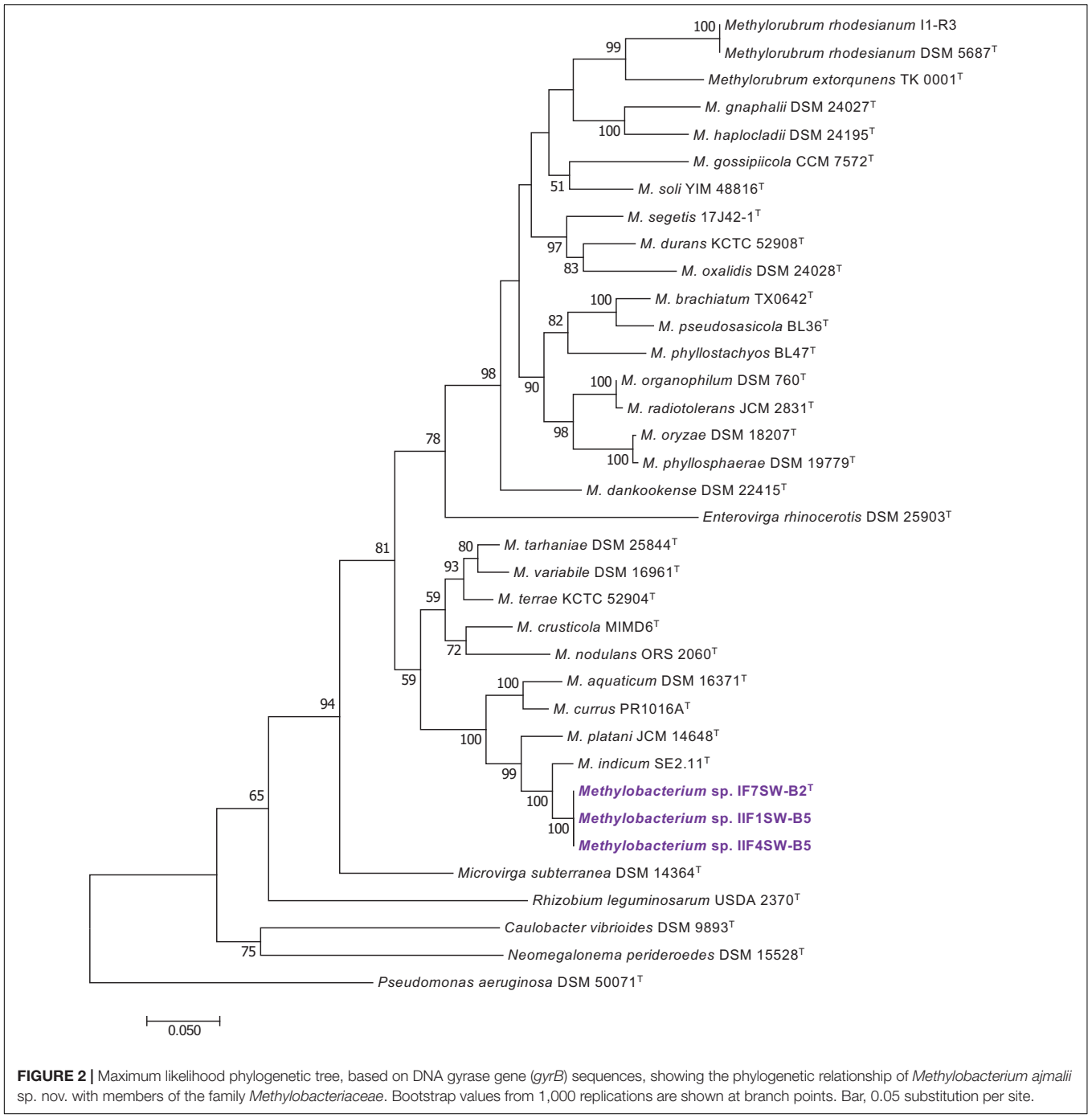


FIGURE 2 | Maximum likelihood phylogenetic tree, based on DNA gyrase gene (*gyrB*) sequences, showing the phylogenetic relationship of *Methylobacterium ajmalii* sp. nov. with members of the family *Methylobacteriaceae*. Bootstrap values from 1,000 replications are shown at branch points. Bar, 0.05 substitution per site.

to as IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5, were identified based on the traditional and genomic taxonomic approaches. The fourth strain, which was isolated from a HEPA filter and referred to as I1-R3, was identified based on genomic analyses only.

Phylogenetic Analysis of Novel ISS Strains

To confirm that three of the ISS strains (IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5) belong to a novel species, their phylogenetic affiliations were analyzed with other species belonging to the genus *Methylobacterium*. The sequence similarity of these three ISS strains with validly described *Methylobacterium* species was <99.4% for 16S rRNA gene (**Supplementary Table 1**) and <97.3% for *gyrB* gene with the closest being *M. indicum* SE2.11^T. Phylogenetic analysis of these three ISS strains was carried out by constructing a maximum likelihood tree based on 16S rRNA (**Figure 1**), *gyrB* (**Figure 2**), *atpD* (**Supplementary Figure 1**), *recA* (**Supplementary Figure 2**), *dnaK* (**Supplementary Figure 3**), *rpoB* (**Supplementary Figure 4**), and *glnI*

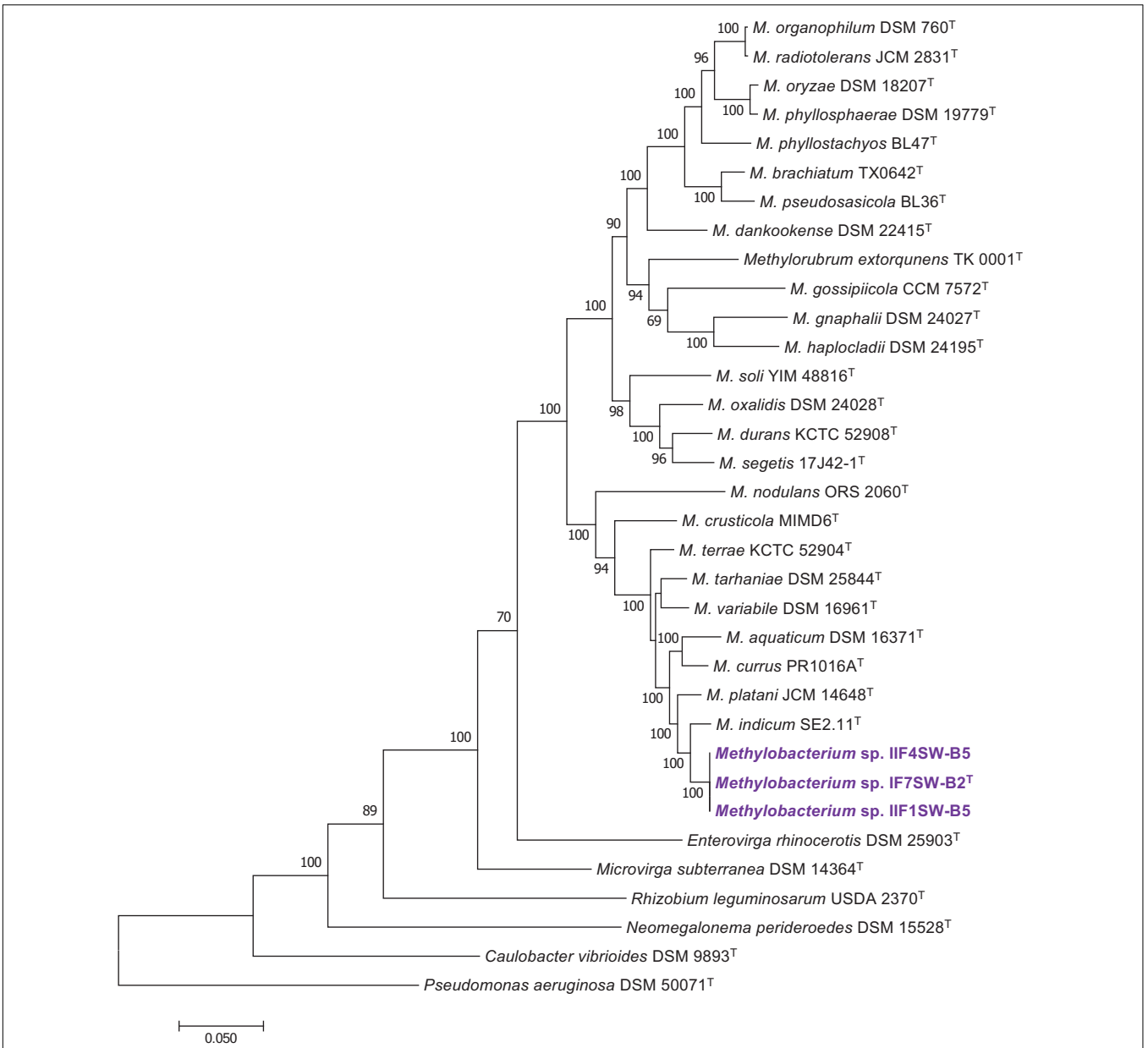


FIGURE 3 | Maximum likelihood phylogenetic tree, based on six gene sequences (*atpD*, *recA*, *dnaK*, *rpoB*, *glnI*, and *gyrB*) concatenated manually, showing the phylogenetic relationship of *Methylobacterium ajmalii* sp. nov. with members of the family *Methylobacteriaceae*. Bootstrap values from 1,000 replications are shown at branch points. Bar, 0.05 substitution per site.

(**Supplementary Figure 5**) gene sequences. In addition, MLSA was carried out by concatenating the six housekeeping genes manually (**Figure 3**). In addition, a phylogenetic tree based on WGS was generated (**Figure 4**). The phylogenetic trees constructed based on all these genes, MLSA, and WGS showed that these three ISS strains (IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5) are clustered together and in the same clade with *M. indicum* SE2.11^T. The 16S rRNA gene-sequencing, housekeeping gene-based analyses, MLSA, and genome-based tree further supported the concept that these three ISS strains belong to the same species but are closely related to *M. indicum*. In addition, the identity of the ISS strain I1-R3 was further confirmed to be *M. rhodesianum* based on its 16S rRNA gene (**Figure 1**) and *gyrB* (**Figure 2**) phylogenetic affiliation to the type strain *M. rhodesianum* DSM 5687^T.

Whole Genome Sequence–Based Phylogenetic Analysis

The genomes of the four isolated ISS strains were sequenced, with their draft genome assembled and annotated. The results are summarized in **Table 1**. The genome varied in size from 6.1 to

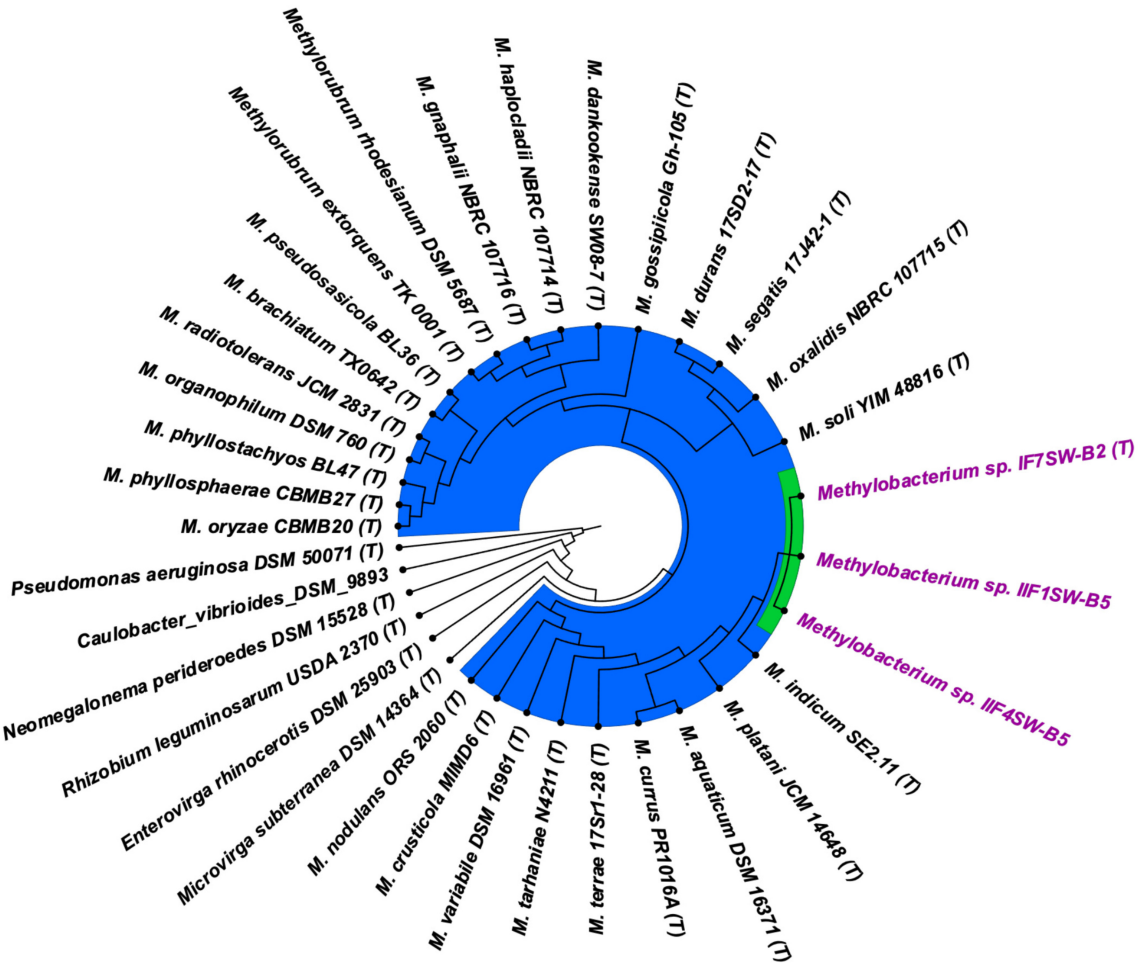


FIGURE 4 | Genome-based phylogenetic tree showing the phylogenetic relationship of *Methylobacterium ajmalii* sp. nov. with members of the family *Methylobacteriaceae*.

6.8 Mbp with GC content between 68 and 71%, similar to other members of the family *Methylobacteriaceae*.

Due to higher sequence similarities of three ISS strains with *M. indicum* SE2.11^T (99.4% for 16S rRNA gene and 97.3% for *gyr*B gene), the draft genomes of three ISS strains were subjected to ANI and dDDH analysis with other species belonging to family *Methylobacteriaceae* (**Table 2**). The ANI indices of three ISS strains (IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5) with *M. indicum* SE2.11^T were 92.7 to 93%, and dDDH values were 45.8 to 46.4%. The ANI and dDDH values obtained for three ISS strains with other *Methylobacterium* species were below the threshold of 95% ANI (Yoon et al., 2017) and 70% dDDH values (Auch et al., 2010), which were established for prokaryotic species delineation. This suggested that these three ISS strains are novel species of the genus *Methylobacterium*. These three ISS strains exhibited ANI and dDDH values around 99–100% with each other, indicating that they belong to the same species. The entire genomes of these three ISS strains, *M. indicum* SE2.11^T, and *M. platani* PMB02^T were aligned to detect their divergence and similarity using the MUMmer 3.0 system (Kurtz et al., 2004). As shown in **Supplementary Figure 6**, genomes of these three ISS strains aligned perfectly, while the closest genomes of *M. indicum* and *M. platani* exhibited divergence with the ISS type strain IF7SW-B2^T. Since these three ISS strains were isolated at different time periods and from various locations, their persistence in the ISS environment and ecological significance in the closed systems warrant further study.

The fourth strain I1-R3 was identified as *M. rhodesianum* based on highly similar 16S rRNA (99.9%), *gyrB* (100%), ANI (98.9%), and dDDH (91.6%) genomic parameters with *M. rhodesianum* DSM 5687^T. The pigmentation of the strain I1-R3 (light pink) was also different from the novel ISS *Methylobacterium* strains (reddish pink). The ANI and dDDH values between I1-R3 and the three novel ISS *Methylobacterium* strains were ∼82% and 24%, respectively. Hence, genomic and morphological analyses confirmed the phylogenetic affiliation of strain I1-R3 as *M. rhodesianum*. In this communication, phylogenetic affiliations of only IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5 strains were presented.

Phenotypic Characterization of Novel ISS Strains

The minimal information about the ISS strain genome characteristics are given in **Supplementary Table 2**. The differential phenotypic characteristics of IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5 are listed in **Table 3**, in comparison with other related *Methylobacterium* species. Three strains belonging to *Methylobacterium* sp. nov. are reddish pink–pigmented, Gram-stain-negative, catalase-positive, oxidase-positive, motile, and rod-shaped. These strains grew well on nutrient agar and R2A. These three strains grew optimally at temperatures between 25 and 30°C, were viable only at pH 6.0 to 8.0, and exhibited poor tolerance to salt (0 to 1%). Absence of growth was observed when grown at 7, 37, and 45°C. These strains were positive for assimilation of L-arabinose, D-glucose, maltose, D-mannitol, D-mannose, malic acid, potassium gluconate, and trisodium citrate. These strains also exhibited esterase lipase and trypsin enzymatic activities. The complete results of phenotypic characteristics determined using API 20 NE, API ZYM, and API 50 CH are detailed in **Supplementary Tables 3-5**, respectively. The majority of the phenotypic characteristics of the ISS strains were similar to other *Methylobacterium* species. Phenotypically, these three ISS strains were different from the closest genomic relative *M. indicum* in assimilating glucose, malic acid, maltose, mannitol, potassium gluconate, and trisodium citrate. Furthermore, unlike *M. indicum*, these ISS strains did not exhibit growth at pH 5.0. In comparison to other *Methylobacterium* species, *M. aquaticum* and *M. terrae* exhibit similar carbon substrate utilization and enzyme production profiles. However, malic acid was assimilated by these ISS strains but not by *M. aquaticum*. Maltose was also utilized by these ISS strains but not by *M. terrae* cells.

The main phenotypic characteristics of the ISS strains IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5 were in accordance with the description of the genus *Methylobacterium*, with the most important being reddish pink pigmentation (Green and Bousfield, 1982). The optimum growth conditions (temperature, pH, salt tolerance) of the ISS *Methylobacterium* strains were similar to other members belonging to the genus *Methylobacterium*. Also, these three ISS strains shared the properties of exhibiting catalase activity and motility with other *Methylobacterium* species. However, the three novel ISS *Methylobacterium* strains differed from other members of the genus *Methylobacterium* in some of the phenotypic characteristics, as shown in **Table 3**. For instance, they exhibited properties like assimilation of certain sugars, which was absent in some of the *Methylobacterium* species. They also did not show cystine arylamidase activity as opposed to several related *Methylobacterium* species.

Chemotaxonomic Characterization of Novel ISS Strains

The FAME profiling of three ISS strains and other related *Methylobacterium* species are given in **Table 4**. The major fatty acids in these ISS strains were $C_{18:1}$ ω7*c* and/or $C_{18:1}$ ω6*c* (Sum

TABLE 1 | Summary of the draft whole-genome sequences of four strains belonging to the family *Methylobacteriaceae*, isolated from the ISS.

Species/Strain	NCBI Accession no.	Isolation location	No. of scaffolds	Genome size (bp)	N_{50} (bp)	Average Coverage	G + C content (%)	Filtered reads used for assembly (million)	Coding sequences
Methylobacterium ajmalii IF7SW-B2^T	JACWCTOOOOOOOO	Lab 3 overhead	192	6,802,552	59,313	698	71.07	33.55	6,255
Methylobacterium ajmalii IIF1SW-B5	JACWCUOOOOOOOO	Cupola	193	6,593,618	50,984	754	71.03	36.29	6,076
Methylobacterium ajmalii IIF4SW-B5	JACWCVOOOOOOOO	Dining table	966	6,534,937	10,467	538	70.77	25.01	6,538
Methylorubrum rhodesianum I1-R3	JACWCWOOOOOOOO	HEPA filter	160	6,159,250	98,200	302	68.96	13.07	5,555

TABLE 2 | Genomic analyses of *Methylobacterium ajmalii* in comparison to other species of the family *Methylobacteriaceae*.

Species (NCBI accession no.)	ANI value (%)			dDDH (%) (Formula 2)		
	IF7SW-B2^T	IIF1SW-B5	IIF4SW-B5	IF7SW-B2^T	IIF1SW-B5	IIF4SW-B5
Methylobacterium ajmalii IF7SW-B2^T (JACWCT000000000)	100	100	100	100	100	99.9
Methylobacterium ajmalii IIF1SW-B5 (JACWCUOOOOOOOOO)	100	100	100	100	100	99.9
Methylobacterium ajmalii IIF4SW-B5 (JACWCVOOOOOOOOO)	99.7	99.6	100	99.9	99.9	100
Methylobacterium indicum SE2.11^T (GCA_001043895.1)	93.0	92.9	92.7	45.8	45.8	46.4
Methylobacterium currus PR1016A^1 (GCA_003058325.1)	90.5	90.5	90.8	36.4	36.4	37.5
Methylobacterium terrae KCTC 52904^T (GCA_003173755.1)	90.4	90.4	90.4	34.8	34.8	35.8
Methylobacterium platani JCM 14648^T (GCA_001043885.1)	90.2	90.3	90.4	35.4	35.5	36.4
Methylobacterium tarhaniae DSM 25844^T (GCA_001043955.1)	89.7	89.5	89.7	34.5	34.6	35.6
Methylobacterium aquaticum DSM 16371^T (GCA_001043915.1)	89.0	89.0	89.1	33.6	33.6	34.5
Methylobacterium variabile DSM 16961^T (GCA_001043975.1)	88.9	88.8	88.9	33.5	33.6	34.4
Methylobacterium crusticola MIMD6^T (GCA_003574465.1)	84.7	84.9	85.1	26.4	26.4	27.1
Methylobacterium nodulans ORS 2060^T (GCA_000022085.1)	82.7	82.7	82.8	24.2	24.3	24.9
Methylobacterium dankookense DSM 22415^T (GCA_902141855.1)	81.6	81.5	81.7	22.9	23.0	23.5
Methylobacterium segetis 17J42-1^1 (GCA_004348265.1)	81.2	81.2	81.4	22.7	22.8	23.4
Methylobacterium oxalidis DSM 24028^T (GCA_007992195.1)	81.2	81.1	81.2	22.3	22.4	23.0
Methylobacterium durans KCTC 52908^T (GCA_003173715.1)	81.0	80.9	81.0	22.4	22.5	23.1
Methylobacterium organophilum DSM 760^T (GCA_003096615.1)	80.9	80.8	81.2	22.4	22.4	22.9
Methylobacterium radiotolerans JCM 2831^T (GCA_000019725.1)	80.9	80.8	81.2	22.2	22.3	22.8
Methylobacterium brachiatum TX0642^T (GCA_003697185.1)	80.9	80.9	81.1	22.5	22.5	22.9
Methylobacterium soli YIM 48816^T (GCA_008806385.1)	80.8	80.7	80.9	22.2	22.2	22.8
Methylobacterium pseudosasicola BL36^T (GCA_900114535.1)	80.7	80.6	80.8	21.7	21.8	22.3
Methylorubrum extorquens TK 0001^T (GCA_900234795.1)	80.6	80.6	80.9	22.0	22.0	22.5
Methylobacterium oryzae DSM 18207^T (GCA_000757795.1)	80.6	80.4	80.6	21.9	22.0	22.5
Methylobacterium phyllosphaerae DSM 19779^T (GCA_001936175.1)	80.6	80.5	80.6	21.8	21.9	22.4
Methylobacterium phyllostachyos BL47^T (GCA_900103445.1)	80.4	80.4	80.4	21.8	21.8	22.3
Methylobacterium gossipiicola CCM 7572^T (GCA_900113485.1)	80.3	80.4	80.4	21.8	21.9	22.5
Methylobacterium haplocladii DSM 24195^T (GCA_007992175.1)	80.3	80.2	80.4	21.7	21.8	22.3
Methylobacterium gnaphalii DSM 24027^T (GCA_007992215.1)	79.9	79.8	80.0	21.4	21.5	21.9
Microvirga subterranea DSM 14364^T (GCA_003350535.1)	79.0	78.9	78.8	20.8	20.8	21.2
Enterovirga rhinocerotis DSM 25903^T (GCA_004363955.1)	78.1	78.1	78.3	20.8	20.8	20.9

in Feature 8; 82 to 85%) with small amounts of $C_{18:0}$ 3-OH, $C_{16:0}$, $C_{17:0}$, Sum in Feature 3, Sum in Feature 2, $C_{18:0}$ and $C_{12:0}$, and traces of $C_{11:0}$, $C_{13:0}$, and $C_{14:0}$. The fatty acids, $C_{18:1}$ ω7*c* and/or $C_{18:1}$ ω6*c*, were observed to be dominant in these ISS strains, similar to other species. However, complete FAME profiles were not consistent among *Methylobacterium* species and some significant differences in the proportions of certain fatty acids were observed (**Table 4**). The notable difference in the FAME profile was the lower abundance of $C_{18:1}$ ω7*c* in *M. indicum* (46%) when compared with these ISS strains (82 to 85%).

These ISS strains contained Q-10 as the major respiratory isoprenoid quinone, which is common in members of the genus *Methylobacterium*. The polar lipids present in these three strains were diphosphatidylglycerol (DPG), phosphatidylglycerol (PG), phosphatidyl choline (PC), phosphatidyl-ethanolamine (PE), and an unidentified lipid (**Supplementary Figure 7**). The total polar lipid profile of these ISS strains was consistent with their close relatives, predominated with phospholipids, DPG, PG, and PE. Furthermore, the chemotaxonomic data together with the results of the genomic and phylogenetic analysis support the affiliation of strains IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5 to the genus *Methylobacterium*.

Functional Characteristics of the Novel ISS Strain

The genome of the ISS strain IF7SW-B2^T, type strain, was annotated and analyzed to determine biotechnologically important genetic determinants. The whole genome and annotation analysis predicted a total of 6,531 genes in the assembled draft genome. Among these, 1,430 fell into various RAST categories, contributing to 2,067 predicted features described in **Table 5**. All the 1,430 feature and subsystems have been documented in **Supplementary Data 1**. A major fraction of the annotated genes was composed of amino acids and derivatives (408), carbohydrate metabolism (246), protein metabolism (198), genes associated with cofactors, vitamins, prosthetic groups, pigments metabolism (190), and respiration (151) (**Table 5**). Genes responsible for motility and chemotaxis

TABLE 3 | Differential phenotypic characteristics of *Methylobacterium ajmalii* and related species of genus *Methylobacterium*.

Characteristic	1	2	3	4	5	6	7	8	9
Growth temperature (°C)	25-30*	18-42	20-30	20-30	20-30	10-37	18-37	20-37	10-40
Growth pH	6.0-8.0	5.0-9.0	5.0-7.0	5.0-8.0	6.0-8.0	4.0-9.0	6.0-8.0	5.0-8.0	4.0-7.0
Catalase	+	W	+	W	+	+	+	+	+
Oxidase	+	+	−	−	+	+	+	−	+
Reduction of nitrate to nitrite	−	w	w	w	w	−	−	+	W
Starch hydrolysis	−	−	+	−	−	+	+	−	−
Assimilation (API-20NE) of:									
D-glucose	+	w	W	+	−	+	+	−	−
L-arabinose	+	+	+	−	+	+	+	+	+
D-mannose	+	−	W	−	−	+	+	−	−
D-mannitol	+	−	+	−	−	-	+	−	−
Maltose	+	−	+	−	−	−	−	−	−
Potassium gluconate	+	+	+	−	+	−	−	−	−+
Malic acid	+	+	−	+	−	+	+	−	w
Trisodium citrate	+	+	−	−	−	+	+	−	w
Phenyl acetic acid	−	.-	W	−	−	−	+	−	−
Enzymatic activity (API-ZYM) of:									
Esterase lipase	+	+	W	w	+	−	+	N.D.	+
Cystine arylamidase	−	+	W	w	w	+	+	N.D.	−
Trypsin	+	+	W	w	w	−	+	N.D.	−

Strains: 1: Methylobacterium sp. nov. (n = 3, where n: IF7SW-B2^T, IIF1SW-B5 and IIF4SW-B5) (this study); 2: M. currus PR1016A^T (Park et al., 2018); 3: M. aquaticum DSM 16371^T (Gallego et al., 2005a; Park et al., 2018); 4: M. variable DSM 16961^T (Gallego et al., 2005b; Park et al., 2018); 5: M. platani JCM 14648^T (Kang et al., 2007; Park et al., 2018); 6: M. tarhaniae DSM 25844^T (Veyisoglu et al., 2013; Kim et al., 2019); 7: M. terrae KCTC 52904^T (Kim et al., 2019); 8: M. indicum SE2.11^T (Chaudhry et al., 2016); 9: M. frigidaeris IER25-16^T (Lee and Jeon, 2018). w, weakly positive; +, Positive; -, Negative; N.D., not determined. All strains exhibited 0-1% NaCl tolerance and were motile.

**Growth was tested at 7, 25, 30, 37, and 45°C and cells were grown optimally at 25 and 30°C but no growth was observed at 7, 37, and 45°C. Growth lower than 25°C might be possible but not tested.*

(95), metabolism of aromatic compounds (47), and stress response (72) were also observed.

Based on the genome annotation, genes for nitrogen metabolism were predicted in the genome of the ISS strain IF7SW-B2^T. Most of the subsystem features aligned with the ammonia assimilation pathway (11 genes), which is a preferred nitrogen source for the bacteria (Leigh and Dodsworth, 2007). In addition, metabolic factors similar to high-affinity phosphate transporter and control of Pho regulon were also identified in the ISS strain IF7SW-B2^T (Wanner, 1993, 1996). Interestingly, a higher number of stress tolerance genes, especially the oxidative stress response factors, were observed in the ISS strain IF7SW-B2^T when compared with other novel species isolated from the ISS; *Methylobacterium* sp. IF7SW-B2^T exhibited 58 features, whereas 36 features were identified in *Solibacillus kalamii* (Seuylemezian et al., 2017) and 18 features were identified in *Kalamiella piersonii* (Singh et al., 2019). The results obtained agree with the previous reports that showed altered regulation of the stress response factors in microorganisms, in the presence of microgravity conditions (Orsini et al., 2017; Aunins et al., 2018). Further studies on the role of oxidative stress in species selection are warranted. The WGS assembly of these three ISS strains reported here will enable the comparative genomic characterization of ISS isolates with Earth counterparts in future studies. This will further aid in the identification of genetic determinants that might potentially be responsible for promoting plant growth under microgravity conditions and contribute to the development of self-sustainable plant crops for long-term space missions in future.

Genes Essential for Interaction With Plants in the ISS Strain

A thorough genomic analysis of the ISS strain IF7SW-B2^T revealed the presence of genes that have been involved in promoting plant growth. The isopentenyl tRNA transferase (*miaA*) essential for cytokinin production reported in *M. aquaticum* strain 22A (Tani et al., 2015) was also found in genome of the ISS strain IF7SW-B2^T with high similarity. The product of the *miaA* gene was reported to be responsible for isopentenylation of a specific adenine in some tRNAs and confirmed the secretion of zeatin originated from tRNA in *M. extorquens* (Koenig et al., 2002). Furthermore, multiple components of the cobalamin synthesis pathway, such as cobalamin biosynthesis protein BluB, L-threonine 3-O-phosphate decarboxylase (EC 4.1.1.81), adenosylcobinamide-phosphate guanylyltransferase (EC 2.7.7.62), cobyric acid synthase (EC 6.3.5.10), nicotinate-nucleotide—dimethylbenzimidazole phosphoribosyltransferase (EC 2.4.2.21), adenosylcobinamide-phosphate synthase (EC

TABLE 4 | Percentage of total cellular fatty acids from *Methylobacterium ajmalii* and related species of genus *Methylobacterium*.

Fatty acids	1	2	3	4	5	6	7	8	9	10	11
C9:0	-	-	-	0.2	-	0.5	-	-	-	-	-
C11:0	0.31	0.66	0.32	0.7	-	0.8	-	-	-	-	-
C12:0	1.06	0.75	1.81	-	-	-	tr	-	-	1.29	7.0
C13:0	0.76	0.65	0.45	-	-	-	-	-	-	-	-
C14:0	0.71	0.42	0.49	0.28	tr	0.9	tr	-	tr	1.18	-
C16:0	3.02	2.66	2.22	8.7	4.1	8.0	7.6	5.7	5.8	6.16	4.2
C17:0	3.09	2.32	2.02	0.4	-	-	-	-	-	1.07	-
C18:0	1.09	0.71	0.57	5.9	1.0	6.6	1.71	2.8	1.6	2.43	2.8
C19:0	-	-	-	-	tr	-	-	-	tr	-	-
$C_{8:0}$3-OH	-	-	-	1.1	-	1.5	-	-	-	-	-
iso- $C_{10:0}$	-	-	-	-	2.1	-	-	-	2.4	-	-
C10:02-OH	-	-	-	0.2	-	0.4	-	-	-	-	-
C11:03-OH	-	-	-	-	tr	-	-	-	tr	-	-
$C_{12:1}$ at 11-12	-	-	-	-	tr	-	-	-	tr	-	-
iso- $C_{13:0}$	-	-	-	-	tr	-	-	-	tr	2.46	-
C13:0 2-OH	-	-	-	-	1.3	-	-	-	-	-	-
anteiso- C14:0	-	-	-	-	-	-	-	-	-	1.94	-
C14:1-ω5c	-	-	-	-	2.4	-	tr	-	2.4	2.21	-
anteiso- $Ci_{5:0}$	-	-	-	-	tr	-	-	-	1.2	3.10	-
iso- $C_{15:0}$	-	-	-	-	-	-	-	-	-	2.22	-
iso- $C_{15:0}$ 3-OH	-	-	-	-	-	-	-	-	-	-	1.6
iso- $C_{15:1}$ F	-	-	-	-	tr	-	-	-	tr	-	-
iso- $C_{15:1}$ G	-	-	-	-	-	-	-	-	-	1.86	-
C16:0 N-alcohol	-	-	-	-	-	-	tr	-	-	-	-
C16:1- ω5c	-	-	-	-	4.9	-	-	-	4.3	-	-
anteiso- C17:0	-	-	-	-	-	-	-	-	-	1.80	-
anteiso- C17:1-ω9c	-	-	-	-	tr	-	-	-	Tr	2.07	-
iso- C17:1-ω5c	-	-	-	-	-	-	-	-	-	2.86	-
iso- $C_{17:0}$ 3-OH	-	-	-	-	tr	-	-	-	tr	1.76	-
C17:1- ω7c	-	-	-	-	tr	-	-	-	tr	-	-
iso-C18:0	0.58	0.46	0.40	-	-	-	-	-	-	-	-
iso-C18:1 H	-	-	-	-	tr	-	-	-	tr	-	-
C18:0- ω5c	0.26	0.28	0.26	-	-	0.7	-	-	-	-	-
C18:1- ω5c	-	-	-	0.6	-	-	-	-	-	-	-
C18:1- ω9c	-	-	-	-	-	-	tr	-	-	-	-
C18:0-3OH	3.14	2.41	3.84	2.4	1.4	1.7	3.46	-	2.0	-	4.9
C18:3- ω6c	-	-	-	-	-	-	tr	-	-	4.99	-
iso-C19:0	-	-	-	-	tr	-	-	-	-	-	-
C19:0 10-methyl	-	-	-	-	-	-	tr	-	-	-	-
C19:0 cyclo- ω8c	-	-	-	-	-	-	-	-	-	-	1.5
C20:1- ω7c	-	-	-	0.4	-	-	-	-	-	-	-
C20:2- ω6,9c	-	-	-	-	-	-	-	-	-	-	1.5
Sum In Feature 8*	82.09	85.08	83.97	66.7	57.6	60.3	81.0	86.4	57.5	46.03	73.8
Sum In Feature 5*	-	-	-	0.6	-	0.7	-	-	-	-	-
Sum In Feature 3*	2.43	2.05	2.37	1.3	10.3	1.0	2.24	1.9	11.2	-	0.9
Sum In Feature 2*	2.17	1.99	1.78	1.8	2.2	2.1	1.98	2.8	1.5	3.22	1.9

Strains: 1: IIF7SW-B2^T (this study), 2: IIF1SW-B5 (this study); 3: IIF4SW-B5 (this study); 4: M. currus PR1016A^T (Park et al., 2018); 5: M. aquaticum DSM 16371^T (Kim et al., 2019); 6: M. variable DSM 16961^T (Park et al., 2018); 7: M. platani JCM 14648^T (Kang et al., 2007); 8: M. tarhaniae DSM 25844^T (Kim et al., 2019); 9: M. terrae KCTC 52904^T (Kim et al., 2019); 10: M. indicum SE2.11^T (Chaudhry et al., 2016); 11: M. frigidaeris IER25-16^T (Lee and Jeon, 2018). -: Not detected, tr: trace amount (<1%).

**Summed features represent groups of two or three fatty acids that cannot be separated using the MIDI system. Summed feature 2 (iso-C16:1 I and/or C14:0 3-OH), summed feature 3 (C16:1 ω7c and/or C16:1 ω6c), Sum In Feature 5 comprises of C18:0-ω6,9c and ante-C18:0 and summed feature 8 (C18:1 ω7c and/or C18:1ω6c).*

TABLE 5 | Genes belonging to different functional categories based on annotation generated using RAST for *Methylobacterium ajmalii* IF7SW-B2^T.

Functional description	Predicted genes*
Cofactors, Vitamins, Prosthetic Groups, Pigments	190
Cell Wall and Capsule	26
Virulence, Disease, and Defense	61
Potassium metabolism	10
Photosynthesis	11
Miscellaneous	17
Phages, Prophages, Transposable elements, Plasmids	25
Membrane Transport	69
RNA Metabolism	40
Nucleosides and Nucleotides	92
Protein Metabolism	198
Cell Division and Cell Cycle	2
Motility and Chemotaxis	95
Regulation and Cell signaling	49
Secondary Metabolism	5
DNA Metabolism	99
Fatty Acids, Lipids, and Isoprenoids	94
Nitrogen Metabolism	14
Dormancy and Sporulation	1
Respiration	151
Stress Response	72
Metabolism of Aromatic Compounds	47
Amino Acids and Derivatives	408
Sulfur Metabolism	17
Phosphorus Metabolism	28
Carbohydrates	246

**Total protein coding genes as per annotated genome.*

6.3.1.10), cob(I)alamin adenosyltransferase. (EC 2.5.1.17), cobalamin synthase (EC 2.7.8.26), and adenosylcobinamide kinase (EC 2.7.1.156), were identified in genome of the ISS strain IF7SW-B2^T. The metabolic pathway for cobalamin synthesis predicted in the ISS strain is presented (**Supplementary Figure 8**). Supporting this prediction, previous study also reported that *Methylobacterium* strains harbor genes involved in the production of a variety of vitamins, such as cobalamin, biotin, thiamin, and riboflavin, indicating the potential of methylobacteria promoting algal growth (Krug et al., 2020). In addition, genes associated with siderophore production, i.e., ferric siderophore transport system, biopolymer transport protein ExbB, and multiple flagellar proteins, were identified in genome of the ISS strain IF7SW-B2^T and are listed in the **Supplementary Figure 8**. Genes involved in iron acquisition and metabolism in which microalgae benefit from bacterial siderophores have been reported previously in *Methylobacterium* spp. (Krug et al., 2020). In the "carbon for iron mutualism" concept, algae assimilated iron complexed in bacterial siderophores and in return provided essential dissolved organic matter for the bacteria (Amin et al., 2015). Similar studies are warranted to confirm the plant-growth promoting activities in the IF7SW-B2^T ISS strain.

In summary, the phylogenetic and genetic distinctiveness and differential phenotypic properties were sufficient to categorize these three ISS strains as members of a species distinct from other recognized *Methylobacterium* species. Therefore, on the basis of the data presented, strains IF7SW-B2^T, IIF1SW-B5, and IIF4SW-B5 represent a novel species of the genus *Methylobacterium*, for which the name *Methylobacterium ajmalii* sp. nov. is proposed. The type strain is IF7SW-B2^T (NRRL B-65601^T and LMG 32165^T).

Description of *Methylobacterium ajmalii* sp. nov.

Methylobacterium ajmalii (aj.ma'li.i. N.L. gen. n. *ajmalii* named after Ajmal Khan, a renowned Indian scientist on biodiversity). Cells are Gram-stain-negative, aerobic, and motile rods showing oxidase- and catalase-positive reactions. Cells are 1.6–1.8 μm wide and 2.2–3.2 μm long. Colonies on R2A agar are reddish pink–pigmented, circular, convex, and smooth, with a diameter of approximately 0.6–1.0 mm after 3 days of incubation on R2A agar. Growth occurs at 25–30°C (optimum, 30°C), at pH 6.0–8.0 (optimum, pH 7.0) and in the presence of 0–1.0% (w/v) NaCl (optimum, 0%). In API ZYM tests, the strain is positive for Alkaline phosphatase, Esterase (C4), Esterase lipase (C8), Leucine arylamidase, Trypsin, Acid phosphatase, and Naphthol-AS-BI-phosphohydrolase, but negative for other enzyme activities. Cells utilize Adipic acid, D-glucose, D-maltose, D-mannitol, D-mannose, L-arabinose, Malic acid, N-acetyl-glucosamine, Potassium gluconate, and Trisodium citrate for growth, but not other substrates in API 20NE. Cells are capable of weakly fermenting inulin and D-melezitose as observed in API 50 CH. Ubiquinone Q-10 is the predominant respiratory isoprenoid quinone. The major fatty acid is summed feature 8 (comprising C18:1 ω7*c* and/or C18:1 ω6*c*). The major polar lipids are diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine, and phosphatidylglycerol. The genomic DNA G + C content of the type strain is 71.07 mol%.

The type strain IF7SW-B2^T is isolated from the International Space Station.

DATA AVAILABILITY STATEMENT

The 16S rRNA gene sequences of *Methylobacterium* sp. IIF1SW-B5, and *Methylobacterium* sp. IIF4SW-B5 are submitted under accession numbers KY218843 and KY218865, respectively. The WGS and the raw data deposited under BioProject accession number PRJNA634337. The WGS accession numbers are mentioned in **Table 1**. The WGS was also deposited in GeneLab under GeneLab dataset (GLDS-300; https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS339 300). The version described in this paper is the first version.

AUTHOR CONTRIBUTIONS

KV and NKS conceived and designed the experiments. SB, VE, and NKS performed the experiments. NKS analyzed the genomic

data inclusive of *de novo* assemblies and verification, scaffold quality assessment, and annotation and generation of the whole genome and protein level alignment for positional description of organism in the tree of life. SB independently verified the genome assembly, generated alignments for all gene trees in the manuscript, and manually curated the tree images. KV and NKS isolated the type strain, and NKS carried out the phenotypic assays and biochemical characterization. KV compiled the contribution of write-ups from all authors associated with phenotype, NKS generated genotype and tables, and SB generated phylogenetic trees and figures. VE conducted the SB generated chemotaxonomic analysis. All authors read and approved the final manuscript. CEM generated the genomic library and sequenced the genomes of all strains. CCCW and ARP reviewed the manuscript.

FUNDING

The research described in this manuscript was funded by a 2012 Space Biology NNH12ZTT001N Grant No. 19-12829-26 under Task Order NNN13D111T awarded to KV, and NASA's 2018 Space Biology (ROSBio) NNH18ZTT001N-FG App B: Flight and Ground Space Biology Research Grant No. 80NSSC19K1501 awarded to CCCW.

ACKNOWLEDGMENTS

The research described in this manuscript was performed at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA and University of Southern California. We would like to thank Aleksandra Checinska-Sielaff for isolating the strain. We thank astronauts Captain Terry Virts for collecting samples aboard the ISS and the Implementation Team at NASA Ames Research Center (Fathi Karouia) for coordinating this effort. We also thank Ryan Kemp (Zymo Corp.) for extracting DNA and Dan Butler (Cornell Medicine) for performing shotgun sequencing using the NovaSeq platform. We also acknowledged the Jet Propulsion Laboratory supercomputing facility staff, notably Narendra J. Patel (Jimmy) and Edward Villanueva, for their continuous support in providing the best possible infrastructure for BIG-DATA analysis. ©2021 California Institute of Technology. Government sponsorship acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.639396/full#supplementary-material

REFERENCES

Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522, 98–101. doi: 10.1038/nature14488

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. *Standards Genom. Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Aunins, T. R., Erickson, K. E., Prasad, N., Levy, S. E., Jones, A., Shrestha, S., et al. (2018). Spaceflight modifies *Escherichia coli* gene expression in response to antibiotic exposure and reveals role of oxidative stress response. *Front. Microbiol.* 9:310. doi: 10.3389/fmicb.2018.00310

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 Genom.* 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

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5:8365.

Chaudhry, V., Baindara, P., Pal, V. K., Chawla, N., Patil, P. B., and Korpole, S. (2016). Methylobacterium indicum sp. nov., a facultative methylotrophic bacterium isolated from rice seed. *Syst. Appl. Microbiol.* 39, 25–32. doi: 10.1016/j.syapm.2015.12.006

Checinska, A., Probst, A. J., Vaishampayan, P., White, J. R., Kumar, D., Stepanov, V. G., et al. (2015). Microbiomes of the dust particles collected from the international space station and spacecraft assembly facilities. *Microbiome* 3:50.

Checinska Sielaff, A., Urbaniak, C., Mohan, G. B. M., Stepanov, V. G., Tran, Q., Wood, J. M., et al. (2019). Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. *Microbiome* 7:50.

Gallego, V., Garcia, M. T., and Ventosa, A. (2005a). Methylobacterium hispanicum sp. nov. and *Methylobacterium aquaticum* sp. nov., isolated from drinking water. *Int. J. Syst. Evol. Microbiol.* 55, 281–287. doi: 10.1099/ijs.0.63319-0

Gallego, V., Garcia, M. T., and Ventosa, A. (2005b). Methylobacterium variabile sp. nov., a methylotrophic bacterium isolated from an aquatic environment. *Int. J. Syst. Evol. Microbiol.* 55, 1429–1433. doi: 10.1099/ijs.0.63597-0

Green, P. N., and Ardley, J. K. (2018). Review of the genus Methylobacterium and closely related organisms: a proposal that some Methylobacterium species be reclassified into a new genus. methylorubrum gen. nov. *Int. J. Syst. Evol. Microbiol.* 68, 2727–2748. doi: 10.1099/ijsem.0.002856

Green, P. N., and Bousfield, I. J. (1982). A taxonomic study of some Gram-negative facultatively methylotrophic bacteria. *Microbiology* 128, 623–638. doi: 10.1099/00221287-128-3-623

Grossi, C. E. M., Fantino, E., Serral, F., Zawoznik, M. S., Fernandez, Do Porto, D. A., et al. (2020). Methylobacterium sp. 2A is a plant growth-promoting rhizobacteria that has the potential to improve potato crop yield under adverse conditions. *Front. Plant Sci.* 11:71. doi: 10.3389/fpls.2020.00071

Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., et al. (2018). RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860.

Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Jones, D. M. (1981). Manual of methods for general bacteriology. *J. Clin. Pathol.* 34:1069. doi: 10.1136/jcp.34.9.1069-c

Jurtshuk, P. Jr., and McQuitty, D. N. (1976). Use of a quantitative oxidase test for characterizing oxidative metabolism in bacteria. *Appl. Environ. Microbiol.* 31, 668–679. doi: 10.1128/aem.31.5.668-679.1976

Kang, Y. S., Kim, J., Shin, H. D., Nam, Y. D., Bae, J. W., Jeon, C. O., et al. (2007). Methylobacterium platani sp. nov., isolated from a leaf of the tree Platanus orientalis. *Int. J. Syst. Evol. Microbiol.* 57, 2849–2853. doi: 10.1099/ijs.0.65262-0

Kelly, D. P., McDonald, I. R., and Wood, A. P. (2014). "The family methylobacteriaceae," in *The Prokaryotes- Alphaproteobacteria and Betaproteobacteria*, eds E. Rosenberg, E. F. DeLong, S. Lory, E.

Stackebrandt, and F. Thompson, (Berlin: Springer), 313–340. doi: 10.1007/978-3-642-30197-1_256

Kim, J., Chhetri, G., Kim, I., Kim, H., Kim, M. K., and Seo, T. (2019). Methylobacterium terrae sp. nov., a radiation-resistant bacterium isolated from gamma ray-irradiated soil. *J. Microbiol.* 57, 959–966. doi: 10.1007/s12275-019-9007-9

Koenig, R. L., Morris, R. O., and Polacco, J. C. (2002). tRNA is the source of low-level trans-zeatin production in *Methylobacterium spp*. *J. Bacteriol.* 184, 1832–1842. doi: 10.1128/jb.184.7.1832-1842.2002

Krug, L., Morauf, C., Donat, C., Muller, H., Cernava, T., and Berg, G. (2020). Plant growth-promoting methylobacteria selectively increase the biomass of biotechnologically relevant microalgae. *Front. Microbiol.* 11:427. doi: 10.3389/fmicb.2020.00427

Kumar, M., Tomar, R. S., Lade, H., and Paul, D. (2016). Methylotrophic bacteria in sustainable agriculture. *World J. Microbiol. Biotechnol.* 32:120.

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

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.

Kwak, M. J., Jeong, H., Madhaiyan, M., Lee, Y., Sa, T. M., Oh, T. K., et al. (2014). Genome information of *Methylobacterium oryzae*, a plant-probiotic methylotroph in the phyllosphere. *PLoS One* 9:e106704. doi: 10.1371/journal.pone.0106704

Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35, 4162–4164. doi: 10.1093/bioinformatics/btz188

Lee, Y., and Jeon, C. O. (2018). *Methylobacterium frigidaeris* sp. nov., isolated from an air conditioning system. *Int. J. Syst. Evol. Microbiol.* 68, 299–304. doi: 10.1099/ijsem.0.002500

Leigh, J. A., and Dodsworth, J. A. (2007). Nitrogen regulation in bacteria and archaea. *Annu. Rev. Microbiol.* 61, 349–377. doi: 10.1146/annurev.micro.61.080706.093409

Madhaiyan, M., Suresh Reddy, B. V., Anandham, R., Senthilkumar, M., Poonguzhali, S., Sundaram, S. P., et al. (2006). Plant growth-promoting methylobacterium induces defense responses in groundnut (*Arachis hypogaea L.*) compared with rot pathogens. *Curr. Microbiol.* 53, 270–276. doi: 10.1007/s00284-005-0452-9

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Goker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* 14:60. doi: 10.1186/1471-2105-14-60

Orsini, S. S., Lewis, A. M., and Rice, K. C. (2017). Investigation of simulated microgravity effects on Streptococcus mutans physiology and global gene expression. *NPJ Microgravity* 3:4.

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.

Parasuraman, P., Pattnaik, S., and Busi, S. (2019). *Phyllosphere Microbiome: Functional Importance in Sustainable Agriculture, New and Future Developments in Microbial Biotechnology and Bioengineering*. Amsterdam: Elsevier, 135–148.

Park, C., Lee, Y. S., Park, S. Y., and Park, W. (2018). Methylobacterium currus sp. nov., isolated from a car air conditioning system. *Int. J. Syst. Evol. Microbiol.* 68, 3621–3626. doi: 10.1099/ijsem.0.003045

Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. doi: 10.1371/journal.pone.0030619

Patt, T. E., Cole, G. C., and Hanson, R. S. (1976). Methylobacterium, a new genus of facultatively methylotrophic bacteria. *Int. J. Syst. Bacteriol.* 26, 226–229. doi: 10.1099/00207713-26-2-226

Ramaprasad, E. V. V., Mahidhara, G., Sasikala, C., and Ramana, C. V. (2018). Rhodococcus electrodiphilus sp. nov., a marine electro active actinobacterium isolated from coral reef. *Int. J. Syst. Evol. Microbiol.* 68, 2644–2649. doi: 10.1099/ijsem.0.002895

Ramaprasad, E. V. V., Sasikala, C., and Ramana, C. V. (2015). Flectobacillus rhizosphaerae sp. nov., isolated from the rhizosphere soil of *Oryza sativa* (L.), and emended description of the genus *Flectobaicillus*. *Int. J. Syst. Evol. Microbiol.* 65, 3451–3456. doi: 10.1099/ijsem.0.000432

Seuylemezian, A., Singh, N. K., Vaishampayan, P., and Venkateswaran, K. (2017). Draft genome sequence of solibacillus kalamii, isolated from an air filter aboard the international space station. *Genome Announc.* 5:e00696-17.

Singh, N. K., Wood, J. M., Mhatre, S. S., and Venkateswaran, K. (2019). Metagenome to phenome approach enables isolation and genomics characterization of *Kalamiella piersonii* gen. nov., sp. nov. from the international space station. *Appl. Microbiol. Biotechnol.* 103, 4483–4497.

Tani, A., Ogura, Y., Hayashi, T., and Kimbara, K. (2015). Complete genome sequence of methylobacterium aquaticum strain 22a, isolated from *Racomitrium japonicum* moss. *Genome Announc.* 3:e00266-15.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624.

Tindall, B., Sikorski, J., Smibert, R., and Krieg, N. (2007). "Phenotypic characterization and the principles of comparative systematics," in *Methods for General and Molecular Microbiology*, eds C. Reddy, T. Beveridge, J. Breznak, G. Marzluf, T. Schmidt, and L. Snyder, (Washington, DC: ASM Press), 330–393.

Veyisoglu, A., Camas, M., Tatar, D., Guven, K., Sazak, A., and Sahin, N. (2013). Methylobacterium tarhaniae sp. nov., isolated from arid soil. *Int. J. Syst. Evol. Microbiol.* 63, 2823–2828.

Wanner, B. L. (1993). Gene regulation by phosphate in enteric bacteria. *J. Cell. Biochem.* 51, 47–54.

Wanner, B. L. (1996). "Phosphorus assimilation and control of the phosphate regulon," in *Escherichia coli and Salmonella: cellular and molecular biology*, ed. F. C. Neidhardt, (Washington, DC: ASM press), 1357–1381.

Yoon, S. H., Ha, S. M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286.

Conflict of Interest: The author(s) declare that there are no conflicts of interest. This manuscript was prepared as an account of work sponsored by NASA, an agency of the US Government. The US Government, NASA, California Institute of Technology, Jet Propulsion Laboratory, and their employees make no warranty, expressed or implied, or assume any liability or responsibility for the accuracy, completeness, or usefulness of information, apparatus, product, or process disclosed in this manuscript, or represents that its use would not infringe upon privately held rights. The use of, and references to any commercial product, process, or service does not necessarily constitute or imply endorsement, recommendation, or favoring by the U.S. Government, NASA, California Institute of Technology, or Jet Propulsion Laboratory. Views and opinions presented herein by the authors of this manuscript do not necessarily reflect those of the U.S. Government, NASA, California Institute of Technology, or Jet Propulsion Laboratory, and shall not be used for advertisements or product endorsements.

Copyright © 2021 Bijlani, Singh, Eedara, Podile, Mason, Wang and Venkateswaran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 May 2021
doi: 10.3389/fmicb.2021.632731

Genomics, Exometabolomics, and Metabolic Probing Reveal Conserved Proteolytic Metabolism of *Thermoflexus hugenholtzii* and Three Candidate Species From China and Japan

Scott C. Thomas[1], Devon Payne[1†], Kevin O. Tamadonfar[1†], Cale O. Seymour[1], Jian-Yu Jiao[2,3], Senthil K. Murugapiran[1,4†], Dengxun Lai[1], Rebecca Lau[5,6], Benjamin P. Bowen[5,6], Leslie P. Silva[5,6], Katherine B. Louie[5,6], Marcel Huntemann[5,6], Alicia Clum[5,6], Alex Spunde[5,6], Manoj Pillay[5,6], Krishnaveni Palaniappan[5,6], Neha Varghese[5,6], Natalia Mikhailova[5,6], I-Min Chen[5,6], Dimitrios Stamatis[5,6], T. B. K. Reddy[5,6], Ronan O'Malley[5,6], Chris Daum[5,6], Nicole Shapiro[5,6], Natalia Ivanova[5,6], Nikos C. Kyrpides[5,6], Tanja Woyke[5,6], Emiley Eloe-Fadrosh[5,6], Trinity L. Hamilton[4], Paul Dijkstra[7], Jeremy A. Dodsworth[8], Trent R. Northen[5,6], Wen-Jun Li[2,3] and Brian P. Hedlund[1,9]**

[1] School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, NV, United States, [2] School of Life Sciences, Sun Yat-sen University, Guangzhou, China, [3] State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China, [4] Department of Plant and Microbial Biology, The BioTechnology Institute, University of Minnesota, St. Paul, MN, United States, [5] The Department of Energy Joint Genome Institute, Berkeley, CA, United States, [6] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, [7] Department of Biological Sciences, Center of Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, United States, [8] Department of Biology, California State University, San Bernardino, CA, United States, [9] Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, Las Vegas, NV, United States

OPEN ACCESS

Edited by:
Jesse G. Dillon, California State University, Long Beach, United States

Reviewed by:
Andrew Decker Steen, The University of Tennessee, Knoxville, United States
Vera Thiel, German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), Germany

***Correspondence:**
Scott C. Thomas
scott.c.thomas19@gmail.com
Brian P. Hedlund
brian.hedlund@unlv.edu

†Present address:
Devon Payne, Department of Microbiology and Immunology, Montana State University, Bozeman, MT, United States
Kevin O. Tamadonfar, Washington University, St. Louis, MO, United States
Senthil K. Murugapiran, Diversigen, New Brighton, MN, United States

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

Received: *23 November 2020*
Accepted: *02 March 2021*
Published: *03 May 2021*

Thermoflexus hugenholtzii JAD2^T, the only cultured representative of the *Chloroflexota* order *Thermoflexales*, is abundant in Great Boiling Spring (GBS), NV, United States, and close relatives inhabit geothermal systems globally. However, no defined medium exists for *T. hugenholtzii* JAD2^T and no single carbon source is known to support its growth, leaving key knowledge gaps in its metabolism and nutritional needs. Here, we report comparative genomic analysis of the draft genome of *T. hugenholtzii* JAD2^T and eight closely related metagenome-assembled genomes (MAGs) from geothermal sites in China, Japan, and the United States, representing "*Candidatus* Thermoflexus japonica," "*Candidatus* Thermoflexus tengchongensis," and "*Candidatus* Thermoflexus sinensis." Genomics was integrated with targeted exometabolomics and ^{13}C metabolic probing of *T. hugenholtzii*. The *Thermoflexus* genomes each code for complete central carbon metabolic pathways and an unusually high abundance and diversity of peptidases, particularly Metallo- and Serine peptidase families, along with ABC transporters for peptides and some amino acids. The *T. hugenholtzii* JAD2^T exometabolome provided evidence of extracellular proteolytic activity based on the accumulation of free amino

acids. However, several neutral and polar amino acids appear not to be utilized, based on their accumulation in the medium and the lack of annotated transporters. Adenine and adenosine were scavenged, and thymine and nicotinic acid were released, suggesting interdependency with other organisms *in situ*. Metabolic probing of *T. hugenholtzii* JAD2^T using ^{13}C-labeled compounds provided evidence of oxidation of glucose, pyruvate, cysteine, and citrate, and functioning glycolytic, tricarboxylic acid (TCA), and oxidative pentose-phosphate pathways (PPPs). However, differential use of position-specific ^{13}C-labeled compounds showed that glycolysis and the TCA cycle were uncoupled. Thus, despite the high abundance of *Thermoflexus* in sediments of some geothermal systems, they appear to be highly focused on chemoorganotrophy, particularly protein degradation, and may interact extensively with other microorganisms *in situ*.

Keywords: exometabolomics, thermophile, genomics, *Chloroflexi*, *Thermoflexus*, *Thermoflexus hugenholtzii*, metagenome-assembled genomes

INTRODUCTION

The bacterial phylum *Chloroflexota* (synonym *Chloroflexi*) continues to be expanded, revealing a global distribution containing broad phylogenetic and physiological diversity. Currently, isolates capable of anoxygenic photosynthesis, obligate organohalide respiration, autotrophy, chemolithotrophy, carboxydotrophy, and fermentation have been described (Moe et al., 2009; Ward et al., 2018; Islam et al., 2019), and recent reports implicate their importance in thermophilic nitrification (Sorokin et al., 2012; Spieck et al., 2019). According to List of Prokaryotic Names with Standing in Nomenclature (Parte et al., 2020), nine classes of *Chloroflexota* have been validly named; however, the Genome Taxonomy Database (Parks et al., 2018) lists 11 classes, only four of which are represented by axenic cultures. *Chloroflexota* are found in freshwater, marine, and hypersaline environments, contaminated groundwater, and terrestrial geothermal springs, among other habitats (Moe et al., 2009; Krzmarzick et al., 2012; Cole et al., 2013; Dodsworth et al., 2014; Hanada, 2014; Denef et al., 2016; Gomez-Saez et al., 2017; Bayer et al., 2018; Kato et al., 2018; Mehrshad et al., 2018; Ward et al., 2018; Thiel et al., 2019; Kochetkova et al., 2020). However, our knowledge of the physiology and ecology of *Chloroflexota* is far from complete, as exemplified by the high abundance and diversity of marine *Chloroflexota* in the poorly understood class *Dehalococcoidia* and the uncultivated SAR202 cluster (Lloyd et al., 2018; Mehrshad et al., 2018).

Many members of the *Chloroflexota* are difficult to isolate and grow in the laboratory, making detailed physiological investigations challenging, even when an isolate is obtained (Yamada et al., 2006, 2007; Bowman et al., 2013; Dodsworth et al., 2014). Several require or are stimulated by complex organic mixtures (e.g., yeast extract, peptone, environmental extracts) (Yamada et al., 2006, 2007; Löffler et al., 2013; Dodsworth et al., 2014). The lack of a defined medium makes one of the most basic biological questions, "what does it eat?", difficult to answer. Slow growth, low growth yield, and the common filamentous morphology of *Chloroflexota* can make quantification of growth challenging, furthering difficulties associated with describing physiological characteristics (Bowman et al., 2013; Dodsworth et al., 2014). New approaches are needed to cultivate and characterize hard-to-grow and yet-to-be isolated microorganisms, including many *Chloroflexota*, to better address their physiology and ecology.

Thermoflexus hugenholtzii JAD2^T was isolated from high-temperature (∼80°C) sediments in Great Boiling Spring (GBS), Nevada, United States, where it can be one of the most abundant organisms (estimated 3.2–60% relative abundance) (Costa et al., 2009; Cole et al., 2013; Dodsworth et al., 2014; Thomas et al., 2019). Similar 16S rRNA gene sequences have been recovered from terrestrial geothermal environments around the world (Engel et al., 2013; Hou et al., 2013; Kato et al., 2018), ranging from 63 to 85°C at circumneutral pH, where they can be abundant [e.g., >8% of 16S rRNA gene sequences (Hou et al., 2013)]. The abundance of *T. hugenholtzii* and close relatives in these springs suggests they contribute significantly to biogeochemical cycling in these systems. Yet, little is known about their metabolic capabilities. Axenic cultures of *T. hugenholtzii* remain difficult to study due to low growth yields (<1 mg dry cell mass L^{-1}), filamentous morphology (up to ∼500 μm long), lack of a defined medium, and dependence on complex organic extracts from GBS water for optimal growth. Furthermore, in culture, *T. hugenholtzii* may have the narrowest growth temperature range of any bacterium or archaeon known (67.5–75°C) (Dodsworth et al., 2014).

The genomic revolution has provided a plethora of information regarding the potential activities of microorganisms, yet there is a need to connect this inferred potential to the actual physiology of the organisms. Better linking genomes to phenomes stands to advance our understanding of microorganisms and microbial communities by going beyond genetic surveys and providing evidence of precise functions and critical links between genetic potential and ecosystem function. To gain an understanding of the activity of microorganisms, one needs to look at the consequences of enzymatic action, in conjunction with genomic, transcriptomic, or proteomic

information, which provide predictions of metabolic capability and evidence of expression. The advancement of exometabolomics, the analysis of metabolites found outside the cell, enables large-scale interpretations of the activities of microorganism through their interactions with molecules in the environment (Mapelli et al., 2008; Silva and Northen, 2015). Similarly, the use of stable isotope-labeled organic compounds can provide information on both catabolic and anabolic activity of specific compounds. The use of position-specific ^{13}C-labeled compounds (i.e., isotopomers) provides even more information, including activities of specific enzymes and rates of different metabolic pathways (Dijkstra et al., 2011a,b; Leighty and Antoniewicz, 2013).

Here, we combined analysis of the draft genome of *T. hugenholtzii* and closely related metagenome-assembled genomes (MAGs) with a study of the exometabolome and stable isotope probing of *T. hugenholtzii* to better understand its metabolism and potential ecological role and help inform environmental studies in systems where *Thermoflexus* species are abundant.

TABLE 1 | *Thermoflexus hugenholtzii* JAD2^T genome statistics[a].

Attribute	Value	% of total
Genome size (bp)	3,216,964	100.00
DNA coding (bp)	2,875,571	89.39
DNA G+C (bp)	2,166,171	67.34
DNA scaffolds	78	100
Total genes	2,997	100
Protein coding genes	2,944	98.23
RNA genes	53	1.77
Pseudo genes	0	0
Genes in internal clusters	427	14.25
Genes with function prediction	2,319	77.38
Genes assigned to COGs	1,928	64.33
Genes with Pfam domains	2,396	79.95
Genes with signal peptides	111	3.70
Genes with transmembrane helices	798	26.63
CRISPR repeats	6	

[a]*Genome statistics obtained from JGI IMG (taxon ID 2140918011).*

MATERIALS AND METHODS

Genome Sequencing

Cultivation of *T. hugenholtzii* JAD2^T for genome sequencing was described in Dodsworth et al. (2014). The genome project for strain JAD2^T was created in the Genomes OnLine Database (Mukherjee et al., 2021) (Go0015989) and genome sequencing, assembly, and annotation performed by the Department of Energy Joint Genome Institute (Berkley, CA, United States) (Huntemann et al., 2015). A summary of the project information associated with MIGS version 2.0 compliance (Field et al., 2008) is provided in **Supplementary Table 1**. Contigs and reads were deposited in GenBank (FYEK00000000 and SRP054824).

A total of eight MAGs were analyzed for comparison to the *T. hugenholtzii* JAD2^T genome (**Tables 1**, **2** and **Supplementary Table 2**). Sample information and sequencing, assembly, and binning information for GBS85_2, GBS70_5, GBS60_20, and GXS_4 can be obtained from the Integrated Microbial Genomes and Microbiomes system (IMG/M, Chen et al., 2020) (300020145, 3300020139, 3300020153, and 3300000865, respectively), and for HR22 from Kato et al. (2018) and under Bioproject ID PRJDB6348.

For MAGs JZ2_71, QQ_20, and QQ_28, raw reads were filtered as in Hua et al. (2015) and high-quality reads were assembled using SPAdes (v3.9.0) (Bankevich et al., 2012). Scaffolds were generated with BBMap (v38.85[1]). Scaffolds with a length > 2.5 kbp were assigned to genomic bins using MetaBAT, which is based on read abundance and tetranucleotide word frequency (Kang et al., 2015). Gene calling was performed using Prodigal (Hyatt et al., 2010). Additional site information for JZ2_71, QQ_20, and QQ_28 can be found under GOLD Biosample IDs Gb0159120 (JZ2) and Gb0187827 (QQ) and in Hedlund et al. (2012).

[1]https://sourceforge.net/projects/bbmap/

All MAGs were checked for contamination and completeness using the CheckM (v1.0.11) lineage workflow (Parks et al., 2015). Ribosomal RNA presence and copy number was predicted using metaxa2 (v2.2) (16S, 23S) (Bengtsson-Palme et al., 2015, 2016), and RNAmmer (v1.2) (16S, 23S, 5S) (Lagesen et al., 2007). Transfer RNA count was predicted using tRNAscan-SE (v2.0.2) (Lowe and Eddy, 1997). MIMAG quality (Bowers et al., 2017) determination was made for each MAG based on these results. All genomes were run through GTDB-Tk (v0.1.1) for taxonomic assignment and identification of protein-coding genes (Hyatt et al., 2010; Matsen et al., 2010; Eddy, 2011; Jain et al., 2017; Parks et al., 2018).

For genome-based phylogenetic analysis, 120 ubiquitous single-copy protein-coding genes (i.e., bac120) from *Thermoflexus* genomes were identified and aligned using the Genome Taxonomy Database Toolkit (GTDB-tk) (Parks et al., 2018). These sequences were combined with a selection of other *Chloroflexota* with species-level assignments in GTDB release 86 along with a single *Escherichia coli* K-12 MG1655 marker alignment as an outgroup. GCF_900187885.1 was omitted from the alignment because this genome is duplicated as IMG 2140918011. IQ-Tree (v1.6.7.a) (Nguyen et al., 2014) was used to construct a phylogenomic tree from the produced alignment. Ultrafast bootstrap (Hoang et al., 2017) and SH-like alrt (Nguyen et al., 2014) values as implemented in IQ-Tree were used at 1,000 replicates for each to assess support for nodes of the tree.

Evaluation of Metabolic Potential

The IMG/M system (Chen et al., 2020), in combination with MAPLE, BlastKOALA, and selected searches and manual curation, was utilized to evaluate the *T. hugenholtzii* genome. Protein sequences were obtained from IMG (IMG Taxon ID: 2140918011) or from NCBI for MAGs and were submitted to MAPLE (Metabolic and Physiological potentiaL Evaluator, v2.3.1) (Takami et al., 2012, 2016; Arai et al., 2018) to determine

TABLE 2 | Genome information and source for *Thermoflexus hugenholtzii* JAD2^T and MAGs[a].

Name	Length (mbp)	# contigs	N50 contigs	% complete	% contam[b]	Source[c]
T. hugenholtzii JAD2^T	3.22	87	139933	97.27	0.91	Isolated from GBS sediment, United States
T. hugenholtzii GBS85_2	3.90	355	15878	94.26	1.82	85°C GBS sediment, United States
T. hugenholtzii GBS70_5	2.83	256	15285	87.27	1.82	70°C GBS sediment, United States
T. hugenholtzii GBS60_20	2.89	324	11092	86.36	3.82	60°C GBS sediment, United States
"*Candidatus* T. sinensis QQ28"	3.50	558	22079	90.91	0.91	68°C QQ sediment, China
"*Candidatus* T. sinensis GXS_4"	3.01	362	10284	77.27	2.42	74°C GXS sediment, China
"*Candidatus* T. sinensis JZ2_71"	3.03	503	7572	85.91	2.36	63°C JZ sediment, China
"*Candidatus* T. tengchongensis QQ20"	3.96	112	80288	95.45	1.82	68°C QQ sediment, China
"*Candidatus* T. japonica HR22"	2.93	175	34479	90.45	1.09	70°C Bioreactor, Japan

[a] *See* ***Supplementary Table 2*** *for additional MAG statistics.*
[b] *% contam., % contamination.*
[c] *GBS, Great Boiling Spring; QQ, Qiao Quan (also called Qiaobianrequan); GXS, Gongxiaoshe: JZ, Jinze [see Hedlund et al. (2012, 2015), Hou et al. (2013), and Thomas et al. (2019) for site locations and details for MAGs from GBS and China].*

Kyoto Encyclopedia of Genes and Genomes (KEGG) functional module completion ratios (MCRs) based on the presence or absence of KEGG orthology groups (KOs) (Takami et al., 2012), using the NCBI BLAST search engine with the bi-directional best hit annotation method for KO assignment, using all organisms in the KEGG database. MAPLE automatically assigns KOs to query genes using the KEGG automatic-annotation server (KAAS), maps the assigned KOs to KEGG functional modules, then calculates MCRs based on the presence of KOs within each functional module. MAPLE also assigns a *Q*-value to each MCR, to aid in the predication of functionally operable metabolic pathways based on the presence or absence of genes in a genome. *Q*-values provide a statistical measure of the likelihood that the module was identified by chance, as many modules share KOs and thus MCR should not be interpreted alone (see Takami et al., 2012, 2016). Each MCR, using the whole community category, was evaluated with a *Q*-value < 0.5 considered biologically feasible, meaning the presence of associated genes in an individual genome suggest that the metabolic pathway/biochemistry (i.e., the KEGG module) is capable of functioning[2].

Genes coding for peptidases and peptidase inhibitors and peptidase genomic abundance comparison to other organisms was done using the MEROPS database batch Blast (v10) (Rawlings et al., 2014, 2018). MAGs and *T. hugenholtzii* protein sequences also were analyzed using the MEROPS database (v12) using BLASTP (NCBI BLAST v2.5.0+) and the merops_scan.lib library for comparison between MAGs and *T. hugenholtzii*. An E-value cutoff of 1E-10 was used and the maximum target sequences matched was set to one.

Thermoflexus hugenholtzii proteins were also submitted to BlastKOALA (v2.1) (Kanehisa et al., 2016) to populate KEGG maps for exploring metabolic potential, and the SignalP server (v5.0) (Nielsen et al., 1997; Armenteros et al., 2019) and SecretomeP server (v2.0) (Bendtsen et al., 2005) for identification of predicted signal peptide sequences.

For evaluating the presence of gene clusters encoding predicted nitrous oxide reductase systems (*nos*DYLZ) and aerobic carbon monoxide dehydrogenase systems (*coxMSLF*) in MAGs, BLAST+ (Camacho et al., 2009) in the web-based Galaxy platform (Cock et al., 2015) was performed using the translated *T. hugenholtzii* JAD2^T genes (*nos*DYLZ, 2413742816-18 and 2143742820; *coxMSLF*, 2143740265-68, 2143742206-09) as the query. Individual amino acid databases for protein-coding genes for MAGs were created and the *T. hugenholtzii* JAD2^T sequences were queried using Megablast (Zhang et al., 2000; Morgulis et al., 2008) against each metagenomic bin database. The top hit in each MAG was carefully examined to assess the quality of the annotation, as reported in **Supplementary Table 5** and described in results.

Cultivation of *Thermoflexus hugenholtzii* JAD2^T for Exometabolomics

Thermoflexus hugenholtzii was cultivated and metabolites were identified in the medium before and after cultivation to determine substrates and products of growth. The cultivation medium was prepared according to the enrichment medium used in Dodsworth et al. (2014) containing the complex carbon sources peptone and yeast extract, except that peptone was increased to 1.0 g/L. Briefly, 20 mL of GBS salts medium, prepared anaerobically, was distributed to 165 mL serum bottles and pressurized with 1 atm of overpressure of N_2. After autoclaving and relieving the excess pressure, peptone, phosphate buffer, and vitamin solutions were added anaerobically just before inoculation. Filter-sterilized air was added to each bottle for a final concentration of 1% O_2. An exponential-phase inoculum was added at 1/100 vol/vol. Control replicates were treated identically but received a sham inoculation of 1/100 vol/vol sterile medium. The bottles were returned to 1 atmosphere over pressure with N_2. Sterile controls, representing the starting medium, were stored at 4°C in the dark for 7 days ($n = 4$). The inoculated replicates ($n = 5$) and another set of sterile controls ($n = 5$) were incubated in the dark with no shaking at ∼74 (±2)°C for 7 days.

After 7 days of incubation, all samples were placed on ice and then centrifuged at 18,514 rcf for 30 min at 4°C in sterile

[2] https://maple.jamstec.go.jp/maple/maple-2.3.1/help.html

50 mL Falcon tubes. The supernatant was decanted and filter-sterilized using 0.2 μm PES filters (VWR) into sterile 50 mL Falcon tubes. Filters were rinsed by the passage of 20 mL of sterile nanopure water prior to filtering the supernatant to remove potential chemical contaminants. Supernatants were stored at −80°C and shipped to Joint Genome Institute (JGI) on dry ice for high-performance liquid chromatography tandem mass spectrometry (HPLC–MS/MS) analysis. All samples were processed together.

Cell density was determined by concentrating 1.0 mL of culture from each replicate for 10 min at 22,442 rcf at 4°C. The supernatant was decanted, and the cell pellet was re-suspended in 200 μL of nanopure water. 5.0 μL of the concentrated subsample was loaded onto a Petroff-Hausser counter (#3900; Hausser Scientific Partnership) and photographed using an Olympus BX51 phase-contrast microscope fitted with a V-TV1x-2 camera (Olympus). Measurements of filament length and density were used to determine cell numbers using an average individual cell length of 4.0 μM (Dodsworth et al., 2014).

Exometabolomics Measurements by HPLC–MS/MS

HPLC–MS/MS was used to identify metabolites. 1 mL media samples with or without *T. hugenholtzii* growth were desalted and extracted using solid-phase extraction cartridges (Bond Elut PPL, 6 mL, 500 mg, #12255001, Agilent). Each cartridge was pre-equilibrated with 1 mL methanol (MeOH) (3×), then 1 mL H_2O (5x), then all water expelled with air. Each sample was then acidified with HCl by adding 20 μL of 6 M HCl to 1 mL media, briefly vortexing, then flowing through the PPL cartridge. Each cartridge was then rinsed with 1 mL of 0.01 M HCl (2×) and air-dried. Eluent was collected following rinses of each cartridge with 1 mL MeOH (2×) and 1 mL acetonitrile (2×) into a 5 mL Eppendorf tube. Eluent extracts of the desalted media were then dried in a SpeedVac (SPD111V, Thermo Scientific) and stored at −80°C.

In preparation for HPLC-MS/MS analysis, dried extracts (eluents) were resuspended in 110 μL MeOH with internal standards (2-Amino-3-bromo-5-methylbenzoic acid, 1 μg mL^{-1}, #R435902; d4-lysine, 10 μg mL^{-1}, #61619210; d5-benzoic acid, 10 μg mL^{-1}, #217158 – Sigma), centrifuge-filtered through a 0.22 μm hydrophilic PVDF membrane (UFC40GV0S, Millipore), and placed into HPLC vials. HPLC-MS/MS was performed on extracts using an Agilent 1290 LC stack, with MS and MS/MS data collected using a Q Exactive Orbitrap MS (Thermo Scientific, San Jose, CA, United States) in centroid format in both positive and negative ion mode. Full MS spectra were collected from *m/z* 70–1,050 at 70,000 resolution, with MS/MS fragmentation data acquired using stepped 10, 20, and 30 eV collision energies at 17,500 resolution. Source settings of the mass spectrometer included a sheath gas flow rate of 55 (au), auxiliary gas flow of 20 (au), sweep gas flow of 2 (au), spray voltage of 3 kV and capillary temperature of 400°C. Between each sample injection, a blank was run consisting of 100% methanol. Normal-phase chromatography was performed using a ZIC-pHILIC column (Millipore SeQuant ZIC-pHILIC, 150 × 2.1 mm, 5 μm, polymeric), at 40°C, at a flow rate of 0.25 mL^{-1} with a 2 μL injection volume for each sample. The HILIC column was equilibrated with 100% buffer B (90:10 ACN:H_2O w/5 mM ammonium acetate) for 1.5 min, diluting buffer B down to 50% with buffer A (H_2O w/5 mM ammonium acetate) for 23.5 min, down to 40% B over 3.2 min, to 0% B over 6.8 min, and followed by isocratic elution in 100% buffer A for 3 min. Metabolites were identified based on exact mass and retention time and comparing MS/MS fragmentation spectra to purchased standards. Raw data files can be obtained through the JGI genome portal under project name "Hedlund 2017 exometabolomics of *Thermoflexus hugenholtzii* JAD2," Project ID: 1196374[3].

HPLC-MS/MS data were analyzed using a custom Python code (Yao et al., 2015). Metabolite identification was performed by comparing detected *m/z*, retention time and MS/MS spectra from experimental data to that of compound standards run using the same LC–MS methods. A positive identification was given when retention time and *m/z* matched that of the standard. For peaks that had associated MS/MS, the highest level of positive identification was achieved when the spectra matched that of the standard. This information is summarized in **Supplementary Table 6**.

HPLC-MS/MS peak-height values for compounds identified in each treatment were compared to determine biological activity, thermal degradation, or thermal production. The treatment with *T. hugenholtzii* growth for 7 days was compared to incubated abiotic controls to determine biological activity, while the sterile starting medium was compared with incubated abiotic controls to determine abiotic effects of high-temperature incubation. Each metabolite was classified according to the Human Metabolome Database (HMDB) hierarchical classification system (Wishart et al., 2018), to simplify links between the LC–MS/MS-identified compounds and genomic data. Metabolites were deemed to have been significantly consumed/degraded or produced if they passed all of the following criteria: (i) at least two of the three treatments' peak height values were normally distributed according to a Shapiro-Wilk test ($p > 0.05$); (ii) either of the comparisons between the non-incubated treatment and the incubated control treatment or the incubated control and the culture treatment showed significant differences when subjected to a Tukey's HSD test ($p < 0.05$); and (iii) at least one treatment had a mean peak height intensity (au) > 10^5. If samples did not match *m/z* and retention times for standards, then they were removed from analysis. If significant compounds were found to have a peak height below 10^5 in all treatments, then they were removed from analysis. If significant compounds were found to have a peak height at or below 10^5 in some but not all treatments, then individual chromatograms were manually inspected. Compounds were excluded from analysis if satisfactory peak shape was not found upon manual inspection.

[3]https://genome.jgi.doe.gov/portal/Hed201tabolomics_FD/Hed201tabolomics_FD.info.html

Shapiro-Wilk and Tukey's HSD tests were performed using R version 3.4.3.

Cultivation of *Thermoflexus hugenholtzii* JAD2^T for ^{13}C-Labeled Substrate Metabolic Probing

^{13}C-labeled metabolic probing was conducted with both position-specific and uniformly labeled substrates, and oxidation of the labeled carbon was assessed by using an isotope spectrometer. The cultivation medium used was based on Dodsworth et al. (2014) and was similar to the exometabolomics medium described above but was scaled up to accommodate multiple head space gas samples (see **Supplementary Table 7**). An exponential-phase inoculum was added to triplicate bottles at 1/100 vol/vol, 15 mL of pure CO_2 was added to provide enough CO_2 (300–2,000 μmol mol^{-1}) for subsequent ^{13}C-CO_2 analysis (see below), and then cultures were incubated at 75°C for the duration of the experiment.

At 98.75 h of growth (early exponential phase), position-specific ^{13}C-labeled substrates or uniformly ^{13}C tricarboxylic acid (TCA) metabolites or amino acids were administered to cultures in the peptone-based medium described above. Each ^{13}C treatment was performed in triplicate. ^{13}C position-specific substrate additions consisted of filter-sterilized solutions (21.4 μmol substrate-C mL^{-1}) of sodium pyruvate (1-^{13}C and 2,3-^{13}C), sodium acetate (1-^{13}C and 2-^{13}C), and glucose (1-^{13}C and uniformly (U) ^{13}C-labeled) (99 atom fraction %; Cambridge Isotope Laboratories, Andover, MA, United States). Uniformly ^{13}C-labeled substrate additions consisted of citrate, L-serine, L-cysteine, L-alanine, and succinate (99 atom fraction %; Cambridge Isotope Laboratories, Andover, MA, United States) at a final concentration of 4.0 μg mL^{-1}. ^{13}C-CO_2 production rate controls were given natural abundance (i.e., non-^{13}C-enriched) pyruvate, acetate, and glucose (as described above for the ^{13}C-labeled compounds). A time 0-h headspace sample (10 mL) was taken immediately prior to ^{13}C-labeled substrate additions, and 1–2 headspace samples (10 mL each) were taken per 24-h period for the next ∼180 h. Cooling was minimized during sampling by placing bottles in a pre-heated (75°C) water bath.

The 10 mL headspace samples were injected into a Tedlar air-sample bag (Zefon International, Ocala, FL, United States) and increased in volume by diluting with CO_2-free air after injecting samples. This was done to facilitate a sample run time of ∼10 min. on a Picarro 2101-*i* CO_2 and CH_4 isotope spectrometer (Picarro Inc., Sunnyvale, CA, United States). Picarro data were recorded as 30-s averages of $\delta^{13}CO_2$ over a period of near-constant delta readings.

Cultures for monitoring the rate of CO_2 production were set up as described above but without ^{13}C-labeled compounds. Headspace samples (10 mL) were taken over the duration of the experiment and run on a LICOR 6262 (Licor Inc., Lincoln, NE, United States) to determine headspace CO_2 concentrations.

Triplicate compound stability controls were also performed. For the sterile compound stability tests, 20 mL of GBS salts medium, prepared anaerobically, was distributed into 165 mL serum bottles and prepared essentially as described above except that no inoculum or additional CO_2 were added. These controls were incubated for ∼180 h at 75°C to mimic the conditions of the ^{13}C-labeled compound additions in the larger Wheaton bottle cultures (see **Supplementary Text 1**). One final sample (∼30 mL) was taken for analysis on the Picarro as described above.

To evaluate the stability of added compounds at growth temperatures, the $^{13}CO_2$ production rate from *T. hugenholtzii* cultures was compared with that of sterile controls by converting the volume of CO_2 present to moles of CO_2 using the ideal gas law and calculating the ^{13}C atom fraction. Using the calculated atom fraction values, we applied a mass balance equation for isotope mixing to determine the contribution of ^{13}C-CO_2 from biotic and abiotic degradation processes (see **Supplementary Text 1** and **Supplementary Table 3** for additional information).

Data Availability

All genomic data are available on one or more data servers, as summarized in **Supplementary Table 2**. Raw metabolomics data are available on IMG through the genome portal under project name "Hedlund 2017 exometabolomics of *Thermoflexus hugenholtzii* JAD2," Project ID: 1196374[3].

RESULTS

Thermoflexus hugenholtzii JAD2^T Genome Overview

The *T. hugenholtzii* JAD2^T draft genome is 3,216,964 bp in size and consists of 78 scaffolds (size range, 121–4,05,611 bp), with a G + C content of 67.34%. The genome encodes 2,997 predicted genes, of which 2,944 are protein-coding (89.39%). Also annotated are 48 tRNA-encoding genes, a single copy of 5S and 16S rRNA genes, and a fragmented 23S rRNA gene. A total of 1,928 genes (64.33%) and 1,141 genes (38.07%) were assigned to COGs and KO groups, respectively (**Supplementary Table 4**). Additional details concerning the genome and interpretations can be found in **Table 1** and **Supplemental Information**.

Key metabolic features of *T. hugenholtzii* were predicted from the genome (**Figure 1**) based on IMG/M annotations, BlastKOALA, and selected manual annotations, and pathways were evaluated for feasibility based on MAPLE MCRs, where *Q*-values below 0.5 were considered feasible (Takami et al., 2012, 2016; Arai et al., 2018). Most central carbon metabolic pathways [e.g., glycolysis, TCA cycle, pentose-phosphate pathway (PPP), gluconeogenesis] are feasible, except for the Entner-Doudoroff pathway (**Supplementary Table 5**). Notably, a gene encoding an archaeal-type fructose 1,6-bisphosphatase aldolase/phosphatase (K01622) (Say and Fuchs, 2010) (**Supplementary Table 6**), involved in gluconeogenesis, was found. In addition, transporters for carbohydrates, arabinogalactan oligomer/maltooligosaccharide, monosaccharides, multiple sugars, rhamnose, and ribose were identified (**Supplementary Table 5**). Transporters for thiamine (IMG gene ID# 2143740997 and 2143740999) and ascorbate [phosphotransferase system (PTS), 2143742986–2143742988] were also identified (**Supplementary Table 5**). Genes coding for nucleoside ABC transporters and a

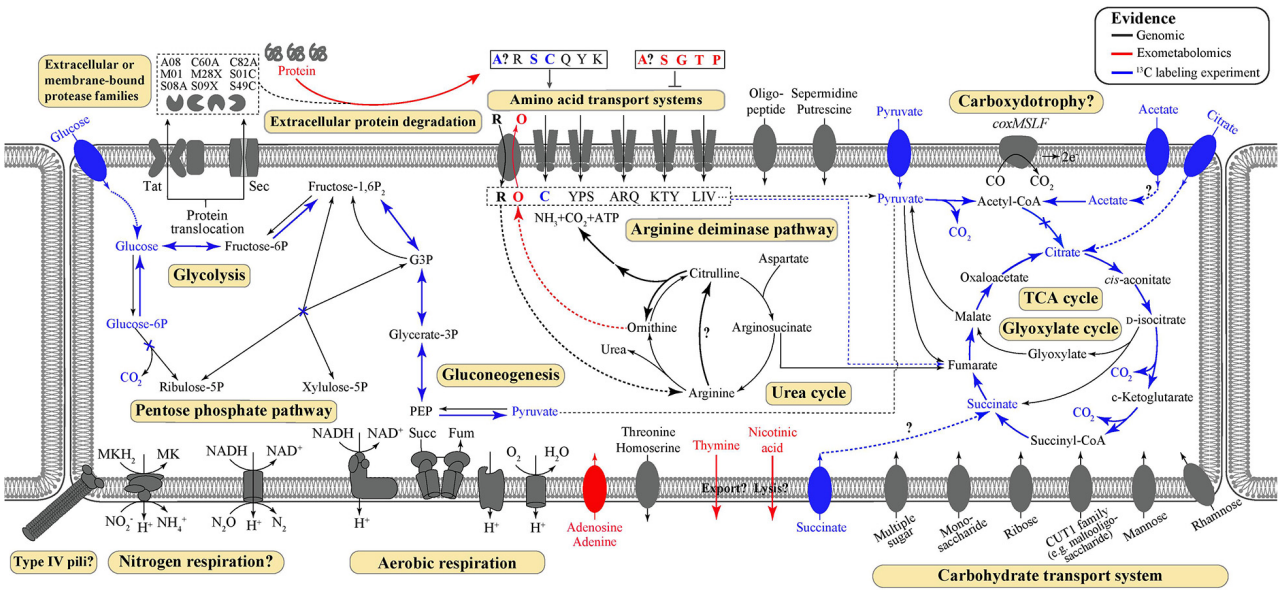


FIGURE 1 | Predicted and demonstrated metabolic features of *Thermoflexus hugenholtzii* JAD2^T. Genomic predictions are shown in black/gray. Evidence derived from exometabolomics experiments and ^{13}C-labeling experiments are shown in red and blue, respectively. Experimental evidence was also supported by genomic data, except for thymine and nicotinic acid secretion. Extracellular or membrane-bound protease families refer to MEROPS protease families. Amino acid transport systems show single letter amino acid codes or strings thereof. Succ, succinate; Fum, fumarate; PEP, phosphoenolpyruvate; MK, menaquinone; G3P, glycerol-3-phosphate.

putative hydroxymethylpyrimidine transporter were found (**Supplementary Table 5**), along with those encoding numerous ABC type-II transporters (**Supplementary Table 5**).

While genes coding for a nitrous oxide reductase system (2413742816-18 and 2143742820, **Supplementary Data Sheet 1**) and a dissimilatory nitrite reduction to ammonium system (*nrfHA* 2143740544 and 2143740545) were identified manually, the MAPLE analysis provided no support for any complete nitrogen or sulfur metabolism modules, including nitrogen fixation (M00175), ammonia oxidation (M00528), complete denitrification (M00529), dissimilatory nitrate reduction to ammonium (M00530), assimilatory nitrate/nitrite reduction (M00531), complete nitrification (i.e., comammox; M00804), assimilatory sulfate reduction (M00176), dissimilatory sulfate reduction (M00596), or respiratory thiosulfate oxidation (SOX pathway, M00595) (**Supplementary Table 5**). The urea cycle (M00029) (**Supplementary Table 5**) appears feasible via a bifunctional carbamate kinase (EC2.7.2.2), fulfilling the role of a carbamoyl-phosphate synthase (EC6.3.4.16) (**Supplementary Table 5**). No components for nitrate/nitrite transport (M00438) or sulfate transport systems (M00185) were found, although all components for a NitT/TauT family transport system (M00188), involved in sulfonate/nitrate/taurine transport, were present (**Supplementary Table 5**). A full aerobic type-I *coxMSLF* was identified (2143740265-68 and 2143742206-09) (**Supplementary Data Sheet 1**). Genes coding for NADH:quinone oxidoreductase, succinate dehydrogenase, cytochrome c oxidase, and an F-type ATPase lacking a prototypical delta subunit were also identified (**Supplementary Table 4**), which is typical of some other *Chloroflexota* (Takami et al., 2012; Chadwick et al., 2018). Components for neither photosystem (M00597 and M00598), nor the 3-hydroxypropionate autotrophic pathway (**Supplementary Table 5**) were detected, and no other autotrophic pathways were encoded in the genome. KEGG modules for the synthesis of bacteriochlorophylls, carotenoids, and rhodopsins were largely unpopulated, and manual searches failed to reveal homologs of key genes for these biosynthetic pathways.

Predicted Protein and Amino Acid Metabolism

Thermoflexus hugenholtzii JAD2^T contains an unusually high abundance and diversity of annotated peptidases, with 133 genes coding for peptidases and five peptidase inhibitors (**Figure 2A**, **Table 3** and **Supplementary Table 5**). 4.4% of total genes coded for members of peptidase families, placing *T. hugenholtzii* JAD2^T in the top 3.6% of *Bacteria* and *Archaea* for the percentage of genes belonging to a MEROPS protein family[4]. 17 of the annotated endo- and exopeptidases are predicted to be membrane-bound or extracellular (**Supplementary Table 5**). Secretion of these proteases would be feasible through the Sec pathway (SecD/F, SecGYA, YidC, Ftsy, and Ffh) and the twin-arginine translocation system (TatAC), along with family I and II signal peptidases (LepB and LspA) (**Supplementary Table 5**). The most abundant protease families are the Metallo (M) and Serine (S) peptidases, with 51 and 53 genes, respectively.

[4]http://merops.sanger.ac.uk/cgi-bin/compgen_index?type=p; accessed on 02/11/2019.

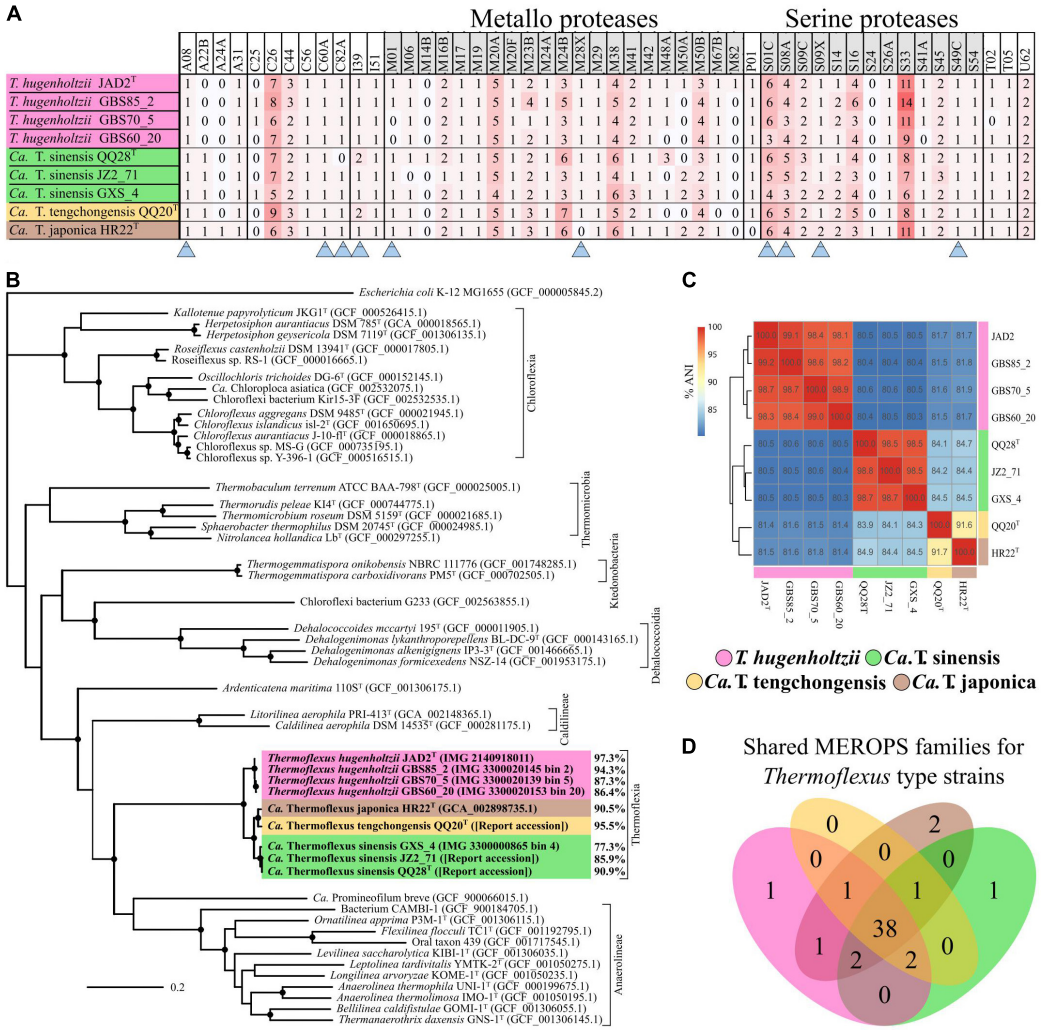


FIGURE 2 | *Thermoflexus* evolutionary relationships and abundance and conservation of MEROPS families. **(A)** Heatmap for the presence and abundance of MEROPS families in all MAGs and *T. hugenholtzii* JAD2^{T}, with families containing predicted extracellular or lipid-anchored proteases marked (arrows). **(B)** Phylogenomic bac120 tree for the phylum *Chloroflexota*. The alignments for all *Chloroflexota* GTDB genomes with classification at the species level and *E. coli* K-12 MG1655 (outgroup) were used to generate the phylogenomic tree, black circles indicate bootstrap value ≥ 95% (1,000 replicates). Percentages associated with *T. hugenholtzii* JAD2^{T} and *Thermoflexus* MAGs represent estimated genome completeness. **(C)** Average nucleotide identity (ANI) matrix for all MAGs and *T. hugenholtzii* JAD2^{T}. **(D)** Venn diagram for MEROPS families shared between *T. hugenholtzii* JAD2^{T} and MAGs representing type material for "*Candidatus* Thermoflexus" species.

Predicted family M01 (2143742583) and M28F (2143740571), both aminopeptidases, have lipoprotein signal peptides (**Supplementary Table 5**). M28F has been shown to result in free arginine, lysine, and leucine (Fundoiano-Hershcovitz et al., 2005). Other genes coding for M-family exopeptidases or peptidases that result in small peptide fragments or free amino acids were predicted to be cytoplasmic and might be important for processing transported oligopeptides or protein turnover (e.g., M01, M03B, M16, M17, M19, M24, M28, M29, M32, M42, and M79). For example, one M19 dipeptidase (2143740867) is predicted to generate free glycine and eight family M20 exopeptidases are predicted to generate free amino acids.

Nine genes belong to the S1 family of Serine proteases, including seven members of the S1C family. This family of endopeptidases resides in the periplasm of Gram-negative bacteria and can also serve as a general molecular chaperone (Krojer et al., 2002) (**Supplementary Table 5**). Five of these seven genes contained lipoprotein signal peptides (214374072, 2143741777, and 2143742813), Sec signal peptides (2143740725), or non-classical secretion sequences (2143742057) (**Supplementary Table 5**). Four peptidases belonged to family S8A, subtilisin endopeptidases with broad substrate specificity, with three containing lipoprotein signal peptides (2143741277) or non-classical secretion sequences (2143740323 and 2143742883) (**Supplementary Table 5**). A gene

TABLE 3 | MEROPS statistics for *Thermoflexus hugenholtzii* JAD2^{T} [a].

MEROPS members of peptidase families	133 (not including 5 inhibitors)
% of total genes coding for members of peptidase families	4.44%
Most abundant families	Metallo (51) and Serine (53) peptidases
Bacteria and Archaea with ≥ 133 members*	13.7%
Bacteria with ≥ 133 members*	14.1%
Bacteria and Archaea with ≥ 4.44% of total genes coding for members of peptidase families*	3.6%
Bacteria with ≥ 4.44% of total genes coding for members of peptidase families*	3.7%
Endopeptidase count	52 (39.10% of total)
Exo-, amino-, carboxy-, di- peptidase count	56 (42.11% of total)

[a]*Data generated using MEROPS version 10 and comparison to Peptidases in Whole Genome Sequences (*), https://www.ebi.ac.uk/merops/cgi-bin/compgen_index?type=P, accessed 02/11/2019. Six erroneous organism entries were removed.*

(2143742360) belonging to family S9B prolyl endopeptidases contained a Sec signal peptide (**Supplementary Table 5**). Family S33, aminopeptidases that preferentially cleave proline from peptides, contained 18 genes, but no secretion sequences were identified. Seven genes encoded family C26 peptidases, consisting of gamma-glutamyl hydrolases closely linked to pyrimidine biosynthesis, arginine biosynthesis, and the urea cycle. Genes for inhibitors belonging to families I39 (2), a broad inhibitor of endopeptidases, I51 (1), an inhibitor of serine carboxypeptidases, and I87 (2), an inhibitor of FtsH, were found (**Supplementary Table 5**).

Several ABC transporters might enable the transport of oligopeptides, free amino acids, or other protein degradation products. Genes for ABC transporters for branched-chain amino acids (LivKHGFM), oligopeptides (OppABDCF), and spermidine/putrescine (PolDGBA) were present (**Supplementary Table 5**). A particularly large gene cluster (IMG gene IDs 2143741899-2143741889) coding for peptide/nickel, polar amino acid, branched-chain amino acid, and hydrophobic amino acid transporters was identified (**Supplementary Table 5**), along with a putative glutamine transport system (2143742870-2143742872; **Supplementary Table 5**). A gene for an amino acid/polyamine/organocation transporter (2143740310), along with two genes for ornithine carbamoyltransferases (2143743137 and 2143741572), and one for a carbamate kinase (2143741573), were found (**Supplementary Table 5**). These genes are part of the arginine deiminase pathway that is responsible for the import and catabolic use of arginine and the export of ornithine.

In contrast, no ABC transporters for general L-amino acids (AapJQMP), cysteine (TcyABC and TcyJKLMN), lysine (LysX1X2Y), histidine (HisJMQP), glutamine (GlnHPQ), arginine (ArtJIMQP), hydroxyproline (LhpPMNO), D-methionine (MetQIN) arginine/ornithine (AotJMQP), glutamate/aspartate (GltIKJL), arginine/lysine/histidine/glutamate (BgtBA), arginine/lysine/histidine (ArtPQR), lysine/arginine/ornithine/histidine/octopine (PA5152-55), neutral amino acids/histidine (NatBCDAE), dipeptide/heme/δ-aminolevulinic acid (DppABCDF), or dipeptide (DppEBCD) transport were found (**Supplementary Table 5**).

Several pathways were identified for the catabolic use of amino acids or interconversion of amino acids, which are likely important for the proteolytic lifestyle of *Thermoflexus*. For example, homoserine, threonine, and glycine could potentially be converted into pyruvate from serine (EC4.3.1.19) and aspartate could be degraded to fumarate (EC6.3.4.4, 4.3.2.2; 6.3.4.5, 4.3.2.2), feeding central carbon metabolism (**Supplementary Table 5**). Arginine, glutamate, or glutamine could also be broken down to 2-oxoglutarate, feeding the TCA cycle, suggesting they may be important substrates for *T. hugenholtzii* JAD2^{T}. Alternatively, using TCA cycle-derived 2-oxoglutarate or arginine, glutamate, or glutamine as substrates, ornithine, citrulline, and proline biosynthesis appears possible (**Supplementary Table 5**). Serine and isoleucine biosynthesis from aspartate by way of homoserine, threonine, and glycine appeared possible (M00018 and M00570) (**Supplementary Table 5**) and suggests these may also be important substrates. Cysteine biosynthesis from serine (M00021) or homocysteine and serine (M00338) was not feasible according to MAPLE (**Supplementary Table 5**) and no other routes for biosynthesis were observed (**Supplementary Table 5**), suggesting *Thermoflexus* might be auxotrophic for cysteine. *T. hugenholtzii* JAD2^{T} was also predicted to be incapable of *de novo* synthesis of asparagine, aspartate, cysteine, glycine, histidine, homocysteine, homoserine, isoleucine, lysine, methionine, phenylalanine, phosphoserine, serine, threonine, and tyrosine, suggesting amino acid scavenging and/or interconversion might be critical to *Thermoflexus* (see **Supplementary Text 1** for additional information).

Environmental Distribution and Metabolic Potential of *Thermoflexus* MAGs

Eight *Thermoflexus* MAGs were identified in public databases, including four high-quality MAGs (GBS85_2, QQ20, QQ28, and HR22) and four medium-quality MAGs (GBS70_5, GBS60_20, and GXS_4, JZ2_71) (**Figure 2B** and **Supplementary Table 2**). All bins contained one copy of each rRNA gene, except JZ2_71, which lacked a 5S rRNA, presumably due to genome incompleteness or a binning problem (**Supplementary Table 2**). Some rRNA genes were fragmented across different contigs. These MAGs derived from four different sediment samples within GBS, several springs in the Tengchong region of southwest China (Qiao Quan spring, Gongxiaoshe spring, and Jinze pool), and an enrichment culture derived from a subsurface gold mine in Japan. These thermal environments range from 60 to 85°C and pH from 6.7 to 7.3, which is generally consistent with the very narrow range for laboratory growth of *T. hugenholtzii* [67.5–75°C; pH 6.5–7.75 (Dodsworth et al., 2014)]. The reason for the high relative abundance of *T. hugenholtzii* in GBS sediments above maximum growth temperature in the laboratory

is unknown (Cole et al., 2013; Thomas et al., 2019). Additional information about these springs is provided elsewhere (Hedlund et al., 2012; Hou et al., 2013; Peacock et al., 2013; Kato et al., 2018; Thomas et al., 2019).

A phylogenetic analysis using the bac120 gene set showed that all MAGs formed a deep-branching monophyletic group within the phylum *Chloroflexota*, with *T. hugenholtzii* JAD2^T being the only cultured representative (**Figure 2B**). The phylogenetic analysis and average nucleotide identity (ANI) values showed that the genomes included four species-level groups. MAGs GBS85_2, GBS70-5, and GBS60_20 belonged to *T. hugenholtzii* (**Figure 2C**). GXS_4, JZ2_71, and QQ28^T belonged to a species cluster sharing 98.5–98.8% ANI, herein called "*Candidatus* Thermoflexus sinensis." HR22^T and QQ20^T formed a cluster but shared only 91.7% ANI and were therefore designated "*Candidatus* Thermoflexus japonica" and "*Candidatus* Thermoflexus tengchongensis," respectively.

All MAGs showed similar MCRs and were generally consistent with the metabolic potential of *T. hugenholtzii* JAD2^T (**Supplementary Table 5**). *T. hugenholtzii* JAD2^T was most similar to the con-specific high-quality MAG GBS85_2, with only ∼4.6% of modules having differing MCRs. With a few exceptions, MCRs for carbohydrate metabolism modules were similar across MAGs and mirrored *T. hughenholtzii* JAD2^T. However, "*Candidatus* Thermoflexus tengchongensis QQ20^T" did not encode the full gene complement for glycolysis (M00001 and M00002) or gluconeogenesis (M00003) modules. The non-oxidative PPP (M00007) was not feasible in "*Candidatus* Thermoflexus japonica HR22^T." The glyoxylate cycle (M00012) was only feasible in the *T. hugenholtzii* group and "*Candidatus* Thermoflexus tengchongensis QQ20^T." All MAGs lacked the delta subunit of the F-type ATPase. The aerobic type-I *coxMSLF* was conserved across all MAGs. However, *Ca.* T. sinensis GXS_4 and "*Candidatus* T. japonica HR22^T" had notably lower sequence similarity for *cox*MLF, and GBS60_20 had lower sequence similarity for *coxM*, with respect to *T. hugenholtzii* JAD2^T (**Supplementary Table 3**). The NreB-NreC (dissimilatory nitrate/nitrite reduction) two-component regulatory system (M00483) was only feasible in *T. hugenholtzii* JAD2^T, GBS85_2, and GBS60_20. A full nitrous oxide reductase system (*nosZDYL*) was found in all MAGs except for *Ca.* T. sinensis JZ2_71, which was missing *nosL*, and GBS60_20, which was missing *nosYL* and had a notably lower sequence similarity for *nosZD*, with respect to *T. hugenholtzii* JAD2^T (**Supplementary Table 5**). In the case of noted absences, this may be due to incomplete genomes from metagenome binning rather than true genomic absences.

Comparison of MEROPS families between MAGs and *T. hugenholtzii* JAD2^T revealed a total of 50 MEROPS protease families and two inhibitor families (**Figure 2D** and **Supplementary Table 5**). Of these, 46 protease families and the two inhibitor families (I39 and I51) were represented by at least one gene in every MAG and *T. hugenholtzii* JAD2^T (**Figure 2D** and **Supplementary Table 5**). MAGs "*Candidatus* T. sinensis QQ28^T," "*Candidatus* T. tengchongensis QQ20^T," "*Candidatus* Thermoflexus japonica HR22^T," and *T. hugenholtzii* JAD2^T shared 38 families, with only one family (M82) unique to *T. hugenholtzii* JAD2^T, one family (M14B) unique to "*Candidatus* T. sinensis QQ28^T," and two families (A24A, S24) unique to "*Candidatus* T. japonica HR22^T" (**Figure 2D**). Within the *T. hugenholtzii* group, 39 families were shared by all members (**Supplementary Data Sheet 2**). Within the "*Candidatus* T. sinensis" group, 41 families were shared by all members (**Supplementary Data Sheet 2**). M and S families were dominant across all MAGs and *T. hugenholtzii* JAD2^T, with 14 and 10 unique families found in all genomes, respectively (**Figure 2A**). Families C26, M20A, M38, S01C, S08A, and S33 were the most abundant families (average count: 7, 5, 5, 5, 4, and 9, respectively) (**Figure 2A** and **Supplementary Table 5**). These numbers are likely an under-estimation for some less complete MAGs.

With a few exceptions, nucleotide and amino acid metabolism modules were similar across all MAGs and mirrored *T. hugenholtzii* JAD2^T (**Supplementary Table 5**). However, inosine monophosphate biosynthesis (M00048) was not feasible in any of the "*Candidatus* T. sinensis" MAGs or the medium-quality GBS70_5 MAG, while it was in all others (**Supplementary Table 5**). Similarly, MAGs generally possessed the same transporters as *T. hugenholtzii* JAD2^T (**Supplementary Table 5**). A molybdate transport system (M00189) was not feasible in the *T. hugenholtzii* group or *Candidatus* T. tengchongensis QQ20^T but was in all others. A ribose transport system (M00212) was not feasible in any of the "*Candidatus* T. sinensis" MAGs; an inositol-phosphate transport system (M00599) was feasible in "*Candidatus* T. japonica HR22^T" and "*Candidatus* T. tengchongensis GXS_4" and "QQ28^T," but not in any others or *T. hugenholtzii* JAD2^T. A PTS transporter for ascorbate was found in all MAGs and *T. hugenholtzii* JAD2^T.

Exometabolomics

Thermoflexus hugenholtzii JAD2^T grew well (0.26–4.6 × 10^7 cells/mL) in cultures for exometabolomics, resulting in an average cell yield of 1.1 × 10^7 cells/mL (**Supplementary Text 1**). NMDS plots showed the exometabolomic profile representing *T. hugenholtzii* growth to have much higher variability than the sterile medium and sterile incubated controls (**Figure 3A** and **Supplementary Table 6**), demonstrating the difficulty to reproducibly grow this organism and limiting our ability to identify statistically significant differences in the abundance of substrates and products. Nevertheless, thirteen compounds that significantly increased or decreased in abundance due to biological or thermal activity were identified with high confidence (**Figures 1**, **3B** and **Supplementary Table 6**). Only two compounds contained in the medium were substrates for *T. hugenholtzii* JAD2^T and were represented by the HMDB classes *Imidazopyrimidines* (adenine) and *Purine nucleosides* (adenosine). Compounds produced due to biological activity were largely represented by the HMDB class *Carboxylic acids and derivatives* (6), with one representative of *Pyridines and derivatives* (nicotinic acid), and one representative of *Organoheterocyclic compounds* (thymine). All compounds of the *Carboxylic acids and derivatives* class belonged to the sub-class *Amino acids*, *peptides*, and *analogs*, with direct parent compounds of alpha-amino acids (glycine) or L-alpha amino

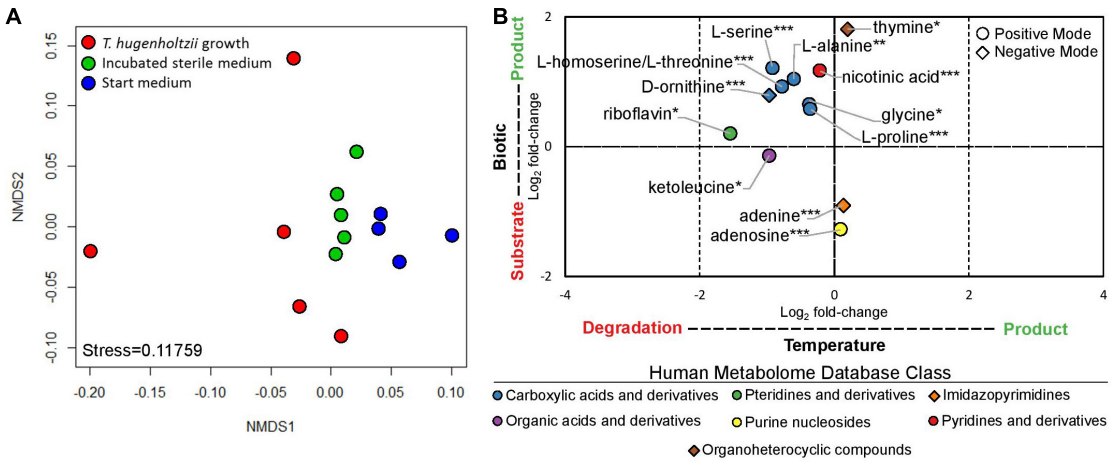


FIGURE 3 | Exometabolomic profile for *Thermoflexus hugenholtzii* JAD2^T. **(A)** NMDS plot and **(B)** Log_2 fold-changes in peak height for statistically significant results. NMDS plots were generated from positive and negative mode peak heights of compounds. Each dot represents a single sample. Replicate #1, with the most growth, is the left-most red data point. Log_2 fold-changes in peak height for statistically significant results were determined by ANOVA and *post hoc* Tukey Honest Significant Differences between treatments for each compound (Shapiro-Wilk test used to determine normality). Data were considered significant if either treatment comparison yielded a p-value < 0.05 for the Tukey HSD. HPLC–MS/MS positive mode (circle) or negative mode (diamond) peak height data were used for all compounds. If a compound was found significant in both positive and negative mode, positive mode data only are presented. Color indicates the Human Metabolome Database (HMDB) metabolite Class for specific compounds. Compound identification, 1, glycine; 2, alanine; 3 homoserine/threonine; 4, proline; 5, serine; 6, monomethyl glutaric acid; 7, ketoleucine; 8, riboflavin; 9, adenosine; 10, nicotinic acid; 11, ornithine; 12, adenine; 13, thymine. *s indicate confidence in compound identification (***, HPLC–MS/MS data matches a fragmented in-house standard; **, *m/z* and retention time match in-house standard but MS/MS fragmentation is difficult to interpret; *, *m/z* and retention time match in-house standard but no fragmentation data are available; if samples did not match *m/z* and retention times for standards, then they were removed from analysis).

acids (L-alanine, L-homoserine/L-threonine, L-proline, L-serine, and L-ornithine). L-homoserine and L-threonine were not distinguishable with HPLC–MS/MS. In addition, many di- and tri-peptides were identified as possible biological products but were not confirmed with purified standards ("Untargeted metabolomics"; see metabolomics data availability above).

In addition, thermal degradation of ketoleucine (4-methyl-2-oxovaleric acid) and riboflavin, belonging to the HMDB classes *Organic acids and derivatives* and *Pteridines and derivatives*, respectively, were observed.

^{13}C Metabolic Probing

Metabolic probing of *T. hugenholtzii* JAD2^T with ^{13}C-labeled compounds demonstrated heterotrophic activity on a variety of organic substrates, including glucose, organic acids, amino acids, and TCA-cycle intermediates (**Figures 1**, **4**). $^{13}CO_2$ was recovered from both universally ^{13}C-labeled and $^{13}C_1$-labeled glucose, the latter providing evidence for the oxidative PPP, albeit at a low rate relative to glucose oxidation through glycolysis. $^{13}C_1$-labeled pyruvate was oxidized to $^{13}CO_2$, providing evidence of pyruvate decarboxylase at the transition between glycolysis and the TCA cycle; however, no $^{13}CO_2$ was recovered from $^{13}C_{2,3}$-labeled pyruvate, demonstrating an uncoupling of glycolysis and the TCA cycle (**Figures 1**, **4**), suggesting acetate might be sequestered through the glyoxylate cycle or possibly for fatty acid biosynthesis (**Figures 1**, **4**). $^{13}CO_2$ was produced from the universally ^{13}C-labeled amino acids cysteine and serine, and the TCA-cycle intermediate citrate. Alanine and succinate also may have been oxidized to $^{13}CO_2$, although those data were not statistically significant.

DISCUSSION

Protein and Amino Acid Metabolism

Thermoflexus hugenholtzii JAD2^T only grows well in the laboratory on a complex medium containing peptone as a carbon, nitrogen, and energy source, suggesting peptides and amino acids sustain *T. hugenholtzii* growth. However, no growth on casamino acids or multiple single amino acids has been observed (Dodsworth et al., 2014). Here, we combined genomic and phenomic analyses to demonstrate that *T. hugenholtzii* JAD2^T does indeed digest extracellular peptides and that some free amino acids are transported and oxidized, whereas others accumulate in the extracellular *milieu*. In all, 17 of the 133 annotated peptidases in *T. hugenholtzii* JAD2^T were predicted to be extracellular or lipid-anchored (**Figures 1**, **2A**, **Table 3**, and **Supplementary Table 5**). *Thermoflexus* MAGs from several geothermal springs in China and Japan showed a similar repertoire of proteases (**Figure 2D**, **Supplementary Table 5**, and **Supplementary Data Sheet 2**), suggesting a conserved proteolytic lifestyle for the genus.

The extracellular accumulation of alanine, glycine, homoserine/threonine, ketoleucine, ornithine, proline, and serine in culture supernatants was consistent with the lack of substrate-specific, general amino acid, and neutral amino acid transporters. However, the ^{13}C metabolic probing

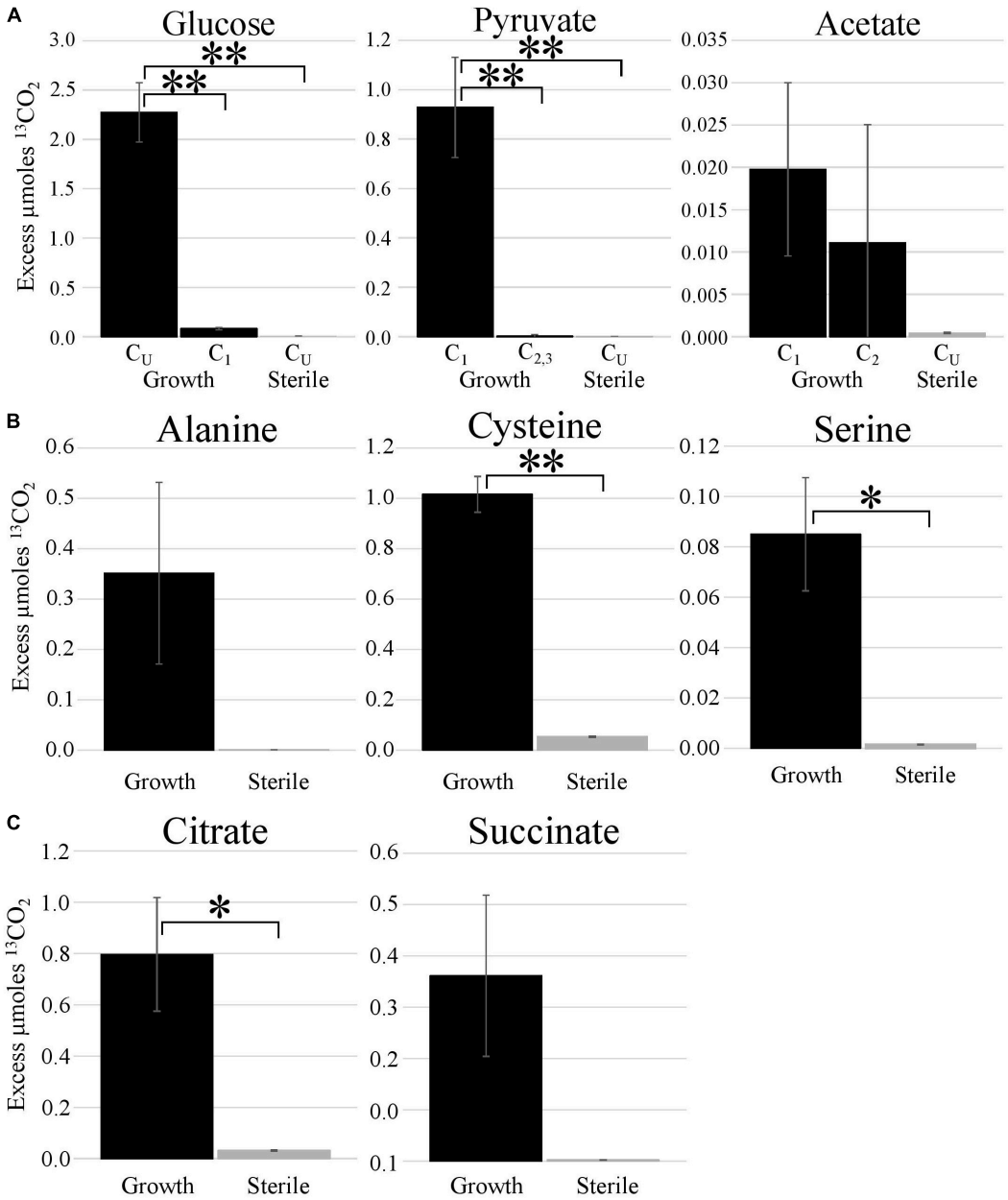


FIGURE 4 | Metabolic activities demonstrated by stable isotope experiments. Excess µmoles of $^{13}CO_2$ produced from ^{13}C-labeled substrates by *T. hugenholtzii* JAD2^T and sterile controls. Isotopomers of glucose, pyruvate, and acetate [**(A)**; CU, uniformly ^{13}C-labeled]; uniformly ^{13}C-labeled amino acids **(B)**; uniformly ^{13}C-labeled TCA metabolites **(C)**. **(A)** *s, indicate statistically different, ANOVA, *post hoc* Tukey HSD. *** <0.005. **(B,C)** *s, indicate statistically different, Student's *t*-test (two-tailed, unequal variance), ** <0.05, * <0.10.

experiments did provide evidence that serine and possibly alanine can be metabolized. It is possible that these amino acids are taken up by other transport systems, albeit at a low rate and/or affinity. The lack of annotated general and neutral amino acid transporters, along with more specific amino acid transporters, was surprising given the protease repertoire of *T. hugenholtzii* JAD2^T. Conversely, no branched-chain, hydrophobic, or charged amino acids accumulated, which is consistent with an abundance of branched-chain and hydrophobic amino acid transporters in the genome (**Supplementary Table 5**).

No amino acids decreased in abundance in the presence of *T. hugenholtzii* growth. This result suggests that individual amino acids were liberated from extracellular peptides at a rate similar to or less than uptake by *T. hugenholtzii*. Thus, the balance of extracellular peptidase activity may be finely tuned with amino acid uptake in *T. hugenholtzii*. Potentially, this helps *T. hugenholtzii* in the natural environment by ensuring energy

and biomass conserved in extracellular proteases is not wasted on amino acid production beyond cellular demand. This would also reduce the free amino acid pool in the extracellular environment and reduce competition.

Extracellular homoserine/threonine and proline accumulation was consistent with the presence of three genes coding for threonine/homoserine efflux transporters (RhtA). This may be indicative of a mechanism for balancing intracellular metabolite pools to facilitate the reactions of central carbon metabolic pathways when feeding on proteins (Livshits et al., 2003). In support of this hypothesis, all amino acids that accumulated in the medium are genomically predicted to be utilized in biosynthetic and catabolic pathways (**Supplementary Table 5**). Serine has also been shown to inhibit threonine and isoleucine biosynthesis (Hama et al., 1991), further suggesting metabolic inhibition may be taking place. Metabolic inhibition may contribute to the low cell density observed in *T. hugenholtzii* cultures when grown on peptides as a carbon and energy source but may be relieved *in situ* by cometabolism with neighboring species.

Consumption of amino acids as a primary carbon and energy source would also lead to excess intracellular nitrogen, which would have to be excreted. Genes coding for necessary gamma-glutamyl hydrolases and an alternative enzyme (EC2.7.2.2) enabling the urea cycle to function (**Supplementary Table 5**) provide such a mechanism. Additionally, gluconeogenesis would be expected under growth on amino acids, which is feasible by an archaeal-type fructose 1,6-bisphosphatase aldolase/phosphatase (Say and Fuchs, 2010). This enzyme may help *T. hugenholtzii* JAD2^T metabolize amino acid-derived heat-labile triosephosphates into heat-stable fructose 6-phosphate, rendering metabolite pools stable, and allow metabolic flexibility free from transcriptional regulation (Say and Fuchs, 2010). Malate dehydrogenase (EC1.1.1.40) appears to be responsible for the start of gluconeogenesis through pyruvate formation from amino acids fed into the TCA cycle, rather than a phosphoenolpyruvate carboxykinase, as was confirmed by a manual search for phosphoenolpyruvate carboxykinase.

Ornithine and ketoleucine accumulated in the medium. Ornithine is a by-product of the urea cycle and ketoleucine can be formed from the incomplete decomposition of branched-chain amino acids, both of which were predicted from the genome (i.e., a complete urea cycle and high abundance of branched-chain amino acid transporters). In addition, genes for parts of the arginine deiminase pathway, a pathway for the catabolism of arginine, were identified; however, an arginine deiminase (EC3.5.3.6) was not identified (Zuniga et al., 2002). This pathway results in the import of arginine, export ornithine, and production of ATP. Ornithine accumulation during *T. hugenholtzii* JAD2^T growth suggests this pathway may be active despite a gene coding for an arginine deiminase not being identified.

Thermoflexus appears to rely on a significant complement of amino acids, which is consistent with an obligately proteolytic lifestyle. For example, many amino acids appear to be metabolic dead ends, meaning that they do not feed into central carbon metabolic pathways, and others have no recognizable *de novo* biosynthetic clusters. For example, isoleucine, leucine, valine, methionine, phenylalanine, and tryptophan could be transported but are possible metabolic dead ends. The *de novo* biosynthesis of several amino acids did not appear possible due to the absence of single genes (i.e., histidine or glutamine, proline, ornithine, arginine, and citrulline or serine). Most of these absences are supported by comparative genomics with *Thermoflexus* MAGs, which suggests these are true absences and not artefacts from incomplete genome assembly. This could mean that these amino acids must be scavenged or interconverted, or alternatively that *Thermoflexus* harbors undefined genes capable of carrying out the missing reactions. In cases where a vast majority of genes are present for a pathway, the presence of undefined genes seems likely (e.g., histidine, tyrosine, phenylalanine, and phosophserine biosynthesis, LysW pathway). Attempts to design a defined medium using a diversity of amino acid mixtures and CCMP metabolites did not support growth. A defined medium would allow further exploration of the capacity for *de novo* amino acid biosynthesis by *T. hugenholtzii* JAD2^T and the identification of novel enzymes and metabolic pathways.

Broad Heterotrophic Activity and Central Carbon Metabolism

The ^{13}C metabolic probing experiments demonstrated broad heterotrophic activity of *T. hugenholtzii* JAD2^T, despite the challenge of growing it in pure culture. This result is generally consistent with genomic predictions, and broad heterotrophic activity demonstrated in GBS sediments where *Thermoflexus* is abundant (Murphy et al., 2013; Thomas et al., 2019). One surprising result is the apparent uncoupling of glycolysis and the TCA cycle, as evidenced by the decarboxylation of $^{13}C_1$ of pyruvate but not $^{13}C_{2,3}$ (**Figure 4**). The *T. hugenholtzii* JAD2^T genome contains two annotated pathways for oxidation of C_1 from pyruvate during formation of acetyl-CoA through pyruvate-ferredoxin oxidoreductase (IMG gene ID# 21437407219-2143740722, 2143741077, 2143741234, and 2143741244–2143741246) or pyruvate dehydrogenase (2143740152 and 2143740153). If the resulting acetyl group were transferred to oxaloacetate by citrate synthase (2143741275), then C_2 and C_3 of pyruvate would be oxidized over multiple cycles of the TCA cycle, which was not observed (**Figure 4**). The absence of this activity suggests acetate, produced from C_2 and C_3 of pyruvate, is either excreted or fully sequestered in biomass. Paradoxically, metabolic probing with ^{13}C-acetate suggested some acetate may be oxidized to $^{13}CO_2$, although the return of $^{13}CO_2$ from either of the isotopomers was not statistically significant (**Figure 4**). By comparison, *Chloroflexus aurantiacus* excretes acetate through an archaeal-type ADP-forming acetyl CoA synthetase (Schmidt and Schönheit, 2013), which is also present in *Thermoflexus* (2143741578). *C. aurantiacus* also assimilates acetate through the glyoxylate cycle when growing mixotrophically with H_2 and CO_2 (Zarzycki and Fuchs, 2011). It is possible that *Thermoflexus* has similar reactions with acetate, although more definitive experiments would be needed to probe these ideas.

The very high ratio of $^{13}CO_2$ production from universally labeled glucose, compared with $^{13}C_1$-glucose (∼25:1) indicates that glycolysis is highly active relative to the oxidative PPP, which would decarboxylate the C_1 position via 6-phosphogluconate dehydrogenase (2143742524). Interpretation of the ^{13}C-glucose and ^{13}C-pyruvate data together suggest a ten-fold higher rate of glycolysis relative to the oxidative PPP, since the only $^{13}CO_2$ production from universally labeled glucose would occur for C_3 and C_4 due to pyruvate-ferredoxin oxidoreductase or pyruvate dehydrogenase. This result might not be surprising given the presence of nucleotides and nucleosides in yeast extract coupled with predicted nucleoside transporters, the demonstrated uptake of adenine and adenosine (**Figure 3**), and the presence of a ribose transporter in all *Thermoflexus* genomes (**Figure 1**). A similarly high ratio of $^{13}CO_2$ production from universally labeled glucose compared with $^{13}C_1$-glucose was seen in 60°C GBS sediments (Thomas et al., 2019). The production of thymine and nicotinic acid are not understood based on incomplete biosynthetic pathways in all *Thermoflexus* genomes and warrants future work.

Potential Alternative Metabolic Strategies

Both nitrous oxide and nitrite were predicted to serve as terminal electron acceptors for anaerobic growth; however, neither metabolism could be confirmed with *T. hugenholtzii* cultures. A nitrous oxide reductase system was conserved across *Thermoflexus* species except MAG GBS60_20, which was obtained from 60°C sediments. Denitrification is active in GBS (Dodsworth et al., 2011). High rates of N_2O flux have been measured in the GBS source pool (∼82°C) and to a slightly lesser degree at a high-temperature shelf (∼82°C), with minimal flux observed at low-temperature sites (∼65°C) (Hedlund et al., 2011). N_2O released by leaky denitrification or other sources at high temperature may provide a terminal electron acceptor for *T. hugenholtzii* strains inhabiting this temperature range. At lower-temperature sites, the source for *T. hugenholtzii* GBS60_20 (Thomas et al., 2019), other organisms may have complete denitrification pathways or outcompete *T. hugenholtzii*, resulting in the loss of the nitrous oxide reductase system in *T. hugenholtzii* adapted to these temperatures. However, to date, no consumption of nitrous oxide has been observed for *T. hugenholtzii* JAD2^T cultures under anaerobic conditions (1% or 5% total volume headspace gas; data not shown). Similarly, no stimulation of growth under anaerobic conditions has been observed in the presence of nitrite (2 mM) (Dodsworth et al., 2014), so the function of the encoded dissimilatory nitrite reduction to ammonium system has also not been verified.

Genes coding for a type-I *coxMSLF* are conserved across all *T. hugenholtzii* MAGs, but carboxydotrophy has also not been observed for *T. hugenholtzii* JAD2^T. It has been suggested that this system may provide a means for *Chloroflexota* to persist in times of low nutrient availability and situations requiring dormancy by providing an alternative energy source (Islam et al., 2019). This system may provide a means for survival for *Thermoflexus* in times of low organic carbon availability, such as a lack of allochthonous C sources. However, no consumption of carbon monoxide (5% of headspace) was observed when *T. hugenholtzii* was grown in the presence of O_2 (1% of headspace) or anaerobically with nitrite (2 mM) or nitrous oxide (5% of headspace) (data not shown).

Ecological Implications and Potential Metabolic Interdependencies

From an ecological perspective, it is intriguing that *T. hugenholtzii* JAD2^T seems to be an obligate chemoheterotroph that depends on proteins and amino acids in light of the high abundance of this organism and close relatives in some hot spring sediments and non-photosynthetic mats (Cole et al., 2013; Hou et al., 2013). In GBS, *T. hugenholtzii* is an abundant member of the sediment community around 80°C (3.2–60% estimated relative abundance), several meters away from photosynthetic mats, which are well-formed in GBS sediments below ∼70°C (Cole et al., 2013). It seems unlikely that microbially derived, autochthonous proteinaceous substrates would be sufficient to support such an abundant organism. However, it is possible that high rates of phage-mediated microbial community turnover may enable *Thermoflexus* to grow to high abundance based on the use of microbial cell lysates serving as a primary source of proteins and extracellular biomass precursors (Breitbart et al., 2004). Similarly, predatory lifestyles have been reported for other *Chloroflexota* (e.g., *Herpetosiphon* spp.) (Livingstone et al., 2018) and perhaps *Thermoflexus* abundance follows a Lotka-Volterra predator-prey relationship, as the estimated abundance of *Thermoflexus* has been observed to fluctuate within GBS sediments over time (e.g., Cole et al., 2013; Thomas et al., 2019). Future environmental studies concerning *Thermoflexus* may benefit from co-occurrence analyses (Chaffron et al., 2010; Freilich et al., 2010; Barberán et al., 2012). Alternatively, or in conjunction with above, *T. hugenholtzii* may rely on allochthonous proteins, which could be addressed through analysis of the natural abundance stable isotopes. In addition, the presence of multiple carbohydrate and sugar importers and complete CCMPs suggest that *T. hugenholtzii* should be able to utilize these substrates as well, although these substrates do not support growth as sole carbon and energy sources (Dodsworth et al., 2014). In the natural environment, *T. hugenholtzii* and close relatives may serve as important players in the initial breakdown of allochthonous proteins, providing a pool of free amino acids for consumption by other community members. It is common to find plant, insect, and animal remains at the sediment-water interface in geothermal systems, and these biomass sources may serve as important proteinaceous substrates for *Thermoflexus*.

Heterotrophy is widespread within the *Chloroflexota*, including both photosynthetic and non-photosynthetic taxa that are abundant and common in circumneutral to alkaline pH geothermal features in Yellowstone National Park. For example, the genera *Roseiflexus* and *Chloroflexus*, both within the *Chloroflexaceae*, are highly abundant in phototrophic mats in the outflow channels of the Octopus Spring and Mushroom Spring in the Lower Geyser Basin, where *in situ* metabolism has been studied in some detail (e.g., van der Meer et al., 2005, 2007). Although these two genera are capable of autotrophy

via the 3-hydroxypropionate pathway, stable-isotope probing experiments have shown them to assimilate both bicarbonate and acetate *in situ*, which is consistent with their preferred mode of photoheterotrophic growth in culture (van der Meer et al., 2010). In these communities, heterotrophic growth is dominant under low light conditions and at night, when most carbon assimilated by *Chloroflexaceae* is derived from fermentation products and other photosynthates released by *Cyanobacteria*.

Thermoflexus extends this general heterotrophic lifestyle to higher temperatures within geothermal systems and similarly it is also likely to be interdependent on other microorganisms. Although mixotrophic *Aquificaceae* are present in both GBS and throughout geothermal springs in Tengchong (Dodsworth et al., 2015; Hedlund et al., 2015), they are not abundant in sediments hosting abundant *Thermoflexus*. However, each of these springs have a long water residence time [e.g., 1–2 days for GBS (Costa et al., 2009)] and they do host abundant *Aquificaceae* populations in the overlying water. Thus, it is possible that spatially uncoupled autotroph-heterotroph interactions exist between *Aquificaceae* and *Thermoflexus* that mirror those between photoautotrophic *Cyanobacteria* and *Chloroflexaceae* at lower temperatures. However, in the case of *Thermoflexus*, the metabolic focus might be detrital material or predation rather than direct metabolic coupling. A metabolism focused on detrital proteins or predation is consistent with the requirements of *Thermoflexus* for exogenous proteins, vitamins, cofactors, and unknown compounds present in organic mat extracts for optimal growth. Similarly, amino acids, thymine, and nicotinic acid released by *Thermoflexus* would likely be useful commodities for other community members. These ideas await more incisive experiments to probe these metabolisms in artificial consortia or *in situ*.

CONCLUSION

By combining genomic and exometabolomic data, insight into the physiology of *T. hugenholtzii* JAD2^T was gained. This synergistic approach allowed us to go beyond genomic predictions, observe the metabolic activity of this minimally culturable organism, and provide confirmation of some, but not all, predictions stemming from genomic analysis. By comparing the *T. hugenholtzii* JAD2^T genome to other *Thermoflexus* MAGs, it was further possible to hypothesize that similar yet-to-be cultivated organisms in geothermal environments around the world have comparable metabolic activity and contributions to biogeochemical cycling. These insights into *Thermoflexus* metabolic capabilities provide a new baseline for the continued cultivation effort of this genus and its relative *Chloroflexota*.

Descriptions of *Candidatus* Species

"*Candidatus* Thermoflexus sinensis" (si'nen.sis) Latin neut. adj. Sinae, Chinese; the Chinese Thermoflexus. The nomenclatural type is the metagenomic bin QQ_bins28 (JAEVEY000000000).

Currently known only from metagenomic sequence data from circumneutral pH geothermal springs in Tengchong, China. Habitat and genomic features suggest a phenotype conforming to the description of the genus *Thermoflexus*. Predicted to be proteolytic, based an abundance of proteases, and facultatively anaerobic, based on cytochrome c oxidase, nitrous oxide reductase, and a dissimilatory nitrite reduction to ammonium system. Possibly carboxydotrophic, based on a type I carbon monoxide dehydrogenase system. In addition to the nomenclatural type, also includes "*Candidatus* T. sinensis" GXS_4 (JAFLMU000000000) and "*Candidatus* T. sinensis" GZ2_71 (JAEVEZ000000000).

"*Candidatus* Thermoflexus tengchongensis" (teng.chong.en'sis) originating from Tenghchong, a region of Yunnan Province, China; the Thermoflexus from Tengchong. The nomenclatural type is the metagenomic bin QQ_bins20 (NCBI ID JAEVEX000000000).

Currently known only from metagenomic sequence data from circumneutral pH geothermal springs in Tengchong, China. Habitat and genomic features suggest a phenotype conforming to the description of the genus *Thermoflexus*. Predicted to be proteolytic, based an abundance of proteases, and facultatively anaerobic, based on cytochrome c oxidase, nitrous oxide reductase, and a dissimilatory nitrite reduction to ammonium system. Possibly carboxydotrophic, based on a type I carbon monoxide dehydrogenase system.

"*Candidatus* Thermoflexus japonica" (ja.pon'i.ca) Latin neut. adj. Japonicus, Japanese; the Japanese Thermoflexus. The nomenclatural type is the metagenomic bin HR22 (BEHY00000000.1).

Currently known only from metagenomic sequence data from an enrichment culture derived from a subsurface gold mine in Japan. Habitat and genomic features suggest a phenotype conforming to the description of the genus *Thermoflexus*. Predicted to be proteolytic, based an abundance of proteases, and facultatively anaerobic, based on cytochrome c oxidase, nitrous oxide reductase, and a dissimilatory nitrite reduction to ammonium system.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MH, AC, AS, MP, KP, NV, NM, I-MC, DS, TR, RO'M, CD, NS, NI, NK, TW, and EE-F sequenced and assembled the genome of *T. hugenholtzii*. J-YJ and W-JL assembled and provided the QQ20, QQ28, and JZ2_71 MAGs. ST, CS, SM, TH, DL, and JD completed the bioinformatic analysis. ST, KT, DP, BH, and PD contributed to the experimental design for ^{13}C work. ST, KT, and PD completed the analysis of ^{13}C work. ST, KT, and DP carried out all the culture work. ST, KT, BH, LS, BB, and TN contributed to the experimental design for exometabolomic work. RL, LS, BB, KL, and TN performed the exometabolomic analysis. ST, CS, and

SM completed the statistical analysis of exometabolomic data. ST interpreted, compiled, and wrote the manuscript with input from all authors.

FUNDING

This work was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was supported by the National Science Foundation (DEB 1557042 and OISE 0968421) and the National Natural Science Foundation of China (No. 91951205). Publication fees for this article were supported by the UNLV MSI Open Access Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.632731/full#supplementary-material

REFERENCES

Arai, W., Taniguchi, T., Goto, S., Moriya, Y., Uehara, H., Takemoto, K., et al. (2018). MAPLE 2. 3. 0: an improved system for evaluating the functionomes of genomes and metagenomes. *Biosci. Biotechnol. Biochem.* 82, 1515–1517. doi: 10.1080/09168451.2018.1476122

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5. 0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

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

Barberán, A., Bates, S., Casamayor, E., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351.

Bayer, K., Jahn, M. T., Slaby, B. M., Moitinho-Silva, L., and Hentschel, U. (2018). Marine sponges as *Chloroflexi* hot spots: genomic insights and high-resolution visualization of an abundant and diverse symbiotic clade. *mSystems* 3:e150-18.

Bendtsen, J. D., Kiemer, L., Fausbøll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol.* 5:58. doi: 10.1186/1471-2180-5-58

Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., et al. (2015). METAXA 2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.* 15, 1403–1414. doi: 10.1111/1755-0998.12399

Bengtsson-Palme, J., Thorell, K., Wurzbacher, C., Sjöling, Å, and Nilsson, R. H. (2016). Metaxa2 diversity tools: easing microbial community analysis with Metaxa2. *Ecol. Inform.* 33, 45–50. doi: 10.1016/j.ecoinf.2016.04.004

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893

Bowman, K. S., Nobre, M. F., Costa, M. S., Rainey, F. A., Moe, W. M., and Moe, W. M. (2013). *Dehalogenimonas alkenigignens* sp. nov., a chlorinated-alkane-dehalogenating bacterium isolated from groundwater. *Int. J. Syst. Evol. Microbiol.* 63, 1492–1498. doi: 10.1099/ijs.0.045054-0

Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T., and Rohwer, F. (2004). Phage community dynamics in hot springs. *Appl. Environ. Microbiol.* 70, 1633–1640. doi: 10.1128/AEM.70.3.1633

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST +: architecture and applications. *BMC Bioinformatics* 10:427. doi: 10.1186/1471-2105-10-421

Chadwick, G. L., Hemp, J., Fischer, W. W., and Orphan, V. J. (2018). Convergent evolution of unusual complex I homologs with increased proton pumping capacity: energetic and ecological implications. *ISME J.* 12, 2668–2680. doi: 10.1038/s41396-018-0210-1

Chaffron, S., Rehrauer, H., Pernthaler, J., and von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* 20, 947–959. doi: 10.1101/gr.104521.109

Chen, I. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2020). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi: 10.1093/nar/gkaa939

Cock, P. J. A., Chilton, J. M., Grüning, B., Johnson, J. E., and Soranzo, N. (2015). NCBI BLAST + integrated into galaxy. *Gigascience* 4:39. doi: 10.1186/s13742-015-0080-7

Cole, J. K., Peacock, J. P., Dodsworth, J. A., Williams, A. J., Thompson, D. B., Dong, H., et al. (2013). Sediment microbial communities in great boiling spring are controlled by temperature and distinct from water communities. *ISME J.* 7, 718–729. doi: 10.1038/ismej.2012.157

Costa, K. C., Navarro, J. B., Shock, E. L., Zhang, C. L., Soukup, D., and Hedlund, B. P. (2009). Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. *Extremophiles* 13, 447–459. doi: 10.1007/s00792-009-0230-x

Denef, V. J., Mueller, R. S., Chiang, E., Liebig, J. R., and Vanderploeg, A. (2016). *Chloroflexi* CL500-11 populations that predominate deep-lake hypolimnion bacterioplankton rely on nitrogen-rich dissolved organic matter metabolism and C 1 compound oxidation. *Appl. Environ. Microbiol.* 82, 1423–1432. doi: 10.1128/AEM.03014-15

Dijkstra, P., Blankinship, J. C., Selmants, P. C., Hart, S. C., Koch, G. W., Schwartz, E., et al. (2011a). Probing carbon flux patterns through soil microbial metabolic networks using parallel position-specific tracer labeling. *Soil Biol. Biochem.* 43, 126–132. doi: 10.1016/j.soilbio.2010.09.022

Dijkstra, P., Dalder, J. J., Selmants, P. C., Hart, S. C., Koch, G. W., Schwartz, E., et al. (2011b). Modeling soil metabolic processes using isotopologue pairs of position-specific 13C-labeled glucose and pyruvate. *Soil Biol. Biochem.* 43, 1848–1857. doi: 10.1016/j.soilbio.2011.05.001

Dodsworth, J. A., Gevorkian, J., Despujos, F., Cole, J. K., Murugapiran, S. K., Ming, H., et al. (2014). *Thermoflexus hugenholtzii* gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the *Chloroflexi*, thermoflexia classis nov., and description of thermoflexaceae fam. nov. and thermoflexales ord. nov. *Int. J. Syst. Evol. Microbiol.* 64, 2119–2127. doi: 10.1099/ijs.0.055855-0

Dodsworth, J. A., Hungate, B. A., and Hedlund, B. P. (2011). Ammonia oxidation, denitrification and dissimilatory nitrate reduction to ammonium in two US great basin hot springs with abundant ammonia-oxidizing archaea. *Environ. Microbiol.* 13, 2371–2386. doi: 10.1111/j.1462-2920.2011.02508.x

Dodsworth, J. A., Ong, J. C., Williams, A. J., Dohnalkova, A. C., and Hedlund, B. P. (2015). *Thermocrinis jamiesonii* sp. nov., a thiosulfate-oxidizing, autotropic thermophile isolated from a geothermal spring. *Int. J. Syst. Evolu. Microbiol.* 65, 4769–4775. doi: 10.1099/ijsem.0.000647

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Engel, A. S., Johnson, L. R., and Porter, M. L. (2013). Arsenite oxidase gene diversity among *Chloroflexi* and *Proteobacteria* from El tatio geyser field, Chile. *FEMS Microbiol. Ecol.* 83, 745–756. doi: 10.1111/1574-6941.12030

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Tatusova, T., et al. (2008). The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* 26, 541–547.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., and Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res.* 38, 3857–3868.

Fundoiano-Hershcovitz, Y., Rabinovitch, L., Shulami, S., Reiland, V., Shoham, G., and Shoham, Y. (2005). The ywaD gene from Bacillus subtilis encodes a double-zinc aminopeptidase. *FEMS Microbiol. Lett.* 243, 157–163. doi: 10.1016/j.femsle.2004.12.001

Gomez-Saez, G. V., Ristova, P. P., Sievert, S. M., Elvert, M., Hinrichs, K., Bühring, S. I., et al. (2017). Relative importance of chemoautotrophy for primary production in a light exposed marine shallow hydrothermal system. *Front. Microbiol.* 8:705. doi: 10.3389/fmicb.2017.00702

Hama, H., Kayahara, T., Tsuda, M., and Tsuchiya, T. (1991). Inhibition of homoserine dehydrogenase I by L-serine in *Escherichia coli*. *J. Biochem.* 109, 604–608. doi: 10.1093/oxfordjournals.jbchem.a123427

Hanada, S. (2014). "The phylum *Chloroflexi*, the family chloroflexaceae, and the related phototrophic families oscillochloridaceae and roseiflexaceae," in *The Prokaryotes*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer), doi: 10.1007/978-3-642-38954-2_165

Hedlund, B. P., Cole, J. K., Williams, A. J., Hou, W., Zhou, E., Li, W., et al. (2012). A review of the microbiology of the Rehai geothermal field in tengchong, yunnan province, China. *Geosci. Front.* 3, 273–288. doi: 10.1016/j.gsf.2011.12.006

Hedlund, B. P., Mcdonald, A. I., Lam, J., Dodsworth, J. A., Brown, J. R., and Hungate, B. A. (2011). Potential role of *Thermus thermophilus* and *T. oshimai* in high rates of nitrous oxide (N_2O) production in ∼80°C hot springs in the US Great Basin. *Geobiology* 9, 471–480. doi: 10.1111/j.1472-4669.2011.00295.x

Hedlund, B. P., Reysenbach, A.-L., Huang, L., John, O. C., Zizhang, L., Dodsworth, J. A., et al. (2015). Isolation of diverse members of the aquificales from geothermal springs in tengchong, China. *Front. Microbio.* 6:157. doi: 10.3389/fmicb.2015.00157

Hoang, D. T., Chernomor, O., Haeseler, A., Von Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.5281/zenodo.854445

Hou, W., Wang, S., Dong, H., Jiang, H., Briggs, B. R., Peacock, J. P., et al. (2013). A comprehensive census of microbial diversity in hot springs of tengchong, yunnan province china using 16S rRNA gene pyrosequencing. *PLoS One* 8:e53350. doi: 10.1371/journal.pone.0053350

Hua, Z., Han, Y., Chen, L., Liu, J., Hu, M., Li, S., et al. (2015). Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* 9, 1280–1294. doi: 10.1038/ismej.2014.212

Huntemann, M., Ivanova, N. N., Mavromatis, K., James Tripp, H., 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, 4–9. doi: 10.1186/s40793-015-0077-y

Hyatt, D., Chen, G., 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

Islam, Z. F., Cordero, P. R. F., Feng, J., Sean, Y. C., Thanavit, K. B., Gleadow, R. M., et al. (2019). Two *Chloroflexi* classes independently evolved the ability to persist on atmospheric hydrogen and carbon monoxide. *ISME J.* 13, 1801–1813. doi: 10.1038/s41396-019-0393-0

Jain, C., Rodriguez-r, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2017). High- throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

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

Kang, D. D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. doi: 10.7717/peerj.1165

Kato, S., Sakai, S., Hirai, M., Tasumi, E., Nishizawa, M., Suzuki, K., et al. (2018). Long-term cultivation and metagenomics reveal ecophysiology of previously uncultivated thermophiles involved in biogeochemical nitrogen cycle. *Microbes Environ.* 33, 107–110. doi: 10.1264/jsme2.ME17165

Kochetkova, T. V., Zayulina, K. S., Zhigarkov, V. S., Minaev, N. V., Chichkov, B. N., Novikov, A. A., et al. (2020). Tepidiforma bonchosmolovskayae gen. nov., sp. nov., a moderately thermophilic *Chloroflexi* bacterium from a Chukotka hot spring (Arctic, Russia), representing a novel class, Tepidiformia, which includes the previously uncultivated lineage OLB14. *Int. J. Syst. Evol. Microbiol.* 70, 1192–1202. doi: 10.1099/ijsem.0.003902

Krojer, T., Garrido-franco, M., Huber, R., Ehrmann, M., and Clausen, T. (2002). Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* 416, 455–459. doi: 10.1038/416455a

Krzmarzick, M. J., Crary, B. B., Harding, J. J., Oyerinde, O. O., Leri, A. C., and Myneni, S. C. B. (2012). Natural niche for organohalide-respiring *Chloroflexi*. *Appl. Environ. Microbiol.* 78, 393–401. doi: 10.1128/AEM.06510-11

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, 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

Leighty, R. W., and Antoniewicz, M. R. (2013). COMPLETE-MFA: complementary parallel labeling experiments technique for metabolic flux analysis. *Metab. Eng.* 20, 49–55. doi: 10.1016/j.ymben.2013.08.006

Livingstone, P. G., Morphew, R. M., Cookson, A. R., and Whitworth, D. E. (2018). Genome analysis, metabolic potential, and predatory capabilities of *Herpetosiphon llansteffanense* sp. nov. *Appl. Environ. Microbiol.* 84:e1040-18. doi: 10.1128/AEM.01040-18

Livshits, V. A., Zakataeva, N. P., Aleshin, V. V., and Vitushkina, M. V. (2003). Identification and characterization of the new gene rhtA involved in threonine and homoserine efflux in *Escherichia coli*. *Res. Microbiol.* 154, 123–135. doi: 10.1016/S0923-2508(03)00036-6

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. *mSystems* 3:e55-18. doi: 10.1128/mSystems.00055-18

Löffler, F. E., Yan, J., Ritalahti, K. M., Adrian, L., Edwards, E. A., Konstantinidis, K. T., et al. (2013). Dehalococcoides mccartyi gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, dehalococcoidia classis nov., order dehalococcoidales ord. nov. and family dehalococcoidaceae fam. nov., within the phylum *Chloroflexi*. *Int. J. Syst. Evol. Microbiol.* 63, 625–635. doi: 10.1099/ijs.0.034926-0

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.

Mapelli, V., Olsson, L., and Nielsen, J. (2008). Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. *Trends Biotechnol.* 26, 490–497. doi: 10.1016/j.tibtech.2008.05.008

Matsen, F. A., Kodner, R. B., and Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538. doi: 10.1186/1471-2105-11-538

Mehrshad, M., Rodriguez-Valera, F., Amoozegar, M. A., Lopez-Garcia, P., and Ghai, R. (2018). The enigmatic SAR202 cluster up close: shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. *ISME J.* 12, 655–668. doi: 10.1038/s41396-017-0009-5

Moe, W. M., Yan, J., Nobre, M. F., da Costa, M. S., and Rainey, F. A. (2009). *Dehalogenimonas* lykanthroporepellens gen. nov., sp. nov., a reductively dehalogenating bacterium isolated from chlorinated solvent-contaminated groundwater. *Int. J. Syst. Evol. Microbiol.* 59, 2692–2697. doi: 10.1099/ijs.0.011502-0

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., and Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics* 24, 1757–1764. doi: 10.1093/bioinformatics/btn322

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., et al. (2021). Genomes OnLine database (GOLD) v.8: overview and updates. *Nucleic Acids* 49, D723–D733. doi: 10.1093/nar/gkaa983

Murphy, C. N., Dodsworth, J. A., Babbitt, A. B., and Hedlund, B. P. (2013). Community microrespirometry and molecular analyses reveal a diverse energy economy in great boiling spring and sandy's spring west in the U.S. great basin. *Appl. Environ. Microbiol.* 79, 3306–3310. doi: 10.1128/AEM.00139-13

Nguyen, L., Schmidt, H. A., Haeseler, A., Von, and Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nielsen, H., Engelbrcht, J., Brunak, S., and Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

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.

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. System. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Peacock, J. P., Cole, J. K., Murugapiran, S. K., Dodsworth, J. A., Fisher, J. C., Moser, D. P., et al. (2013). Pyrosequencing reveals high-temperature cellulolytic microbial consortia in Great Boiling Spring after in situ lignocellulose enrichment. *PLoS One* 8:e0059927. doi: 10.1371/journal.pone.0059927

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, 624–632. doi: 10.1093/nar/gkx1134

Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, D503–D509. doi: 10.1093/nar/gkt953

Say, R. F., and Fuchs, G. (2010). Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* 464, 1077–1081. doi: 10.1038/nature08884

Schmidt, M., and Schönheit, P. (2013). Acetate formation in the photoheterotrophic bacterium *Chloroflexus aurantiacus* involves an archaeal type ADP-forming acetyl-CoA synthetase isoenzyme I. *FEMS Microbiol Lett.* 349, 171–179. doi: 10.1111/1574-6968.12312

Silva, L. P., and Northen, T. R. (2015). Exometabolomics and MSI: deconstructing how cells interact to transform their small molecule environment. *Curr. Opin. Biotechnol.* 34, 209–216. doi: 10.1016/j.copbio.2015.03.015

Sorokin, D. Y., Lücker, S., Vejmelkova, D., Kostrikina, N. A., Kleerebezem, R., Rijpstra, W. I., et al. (2012). Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum *Chloroflexi*. *ISME J.* 6, 2245–2256. doi: 10.1038/ismej.2012.70

Spieck, E., Spohn, M., Wendt, K., Bock, E., Shively, J., Frank, J., et al. (2019). Extremophilic nitrite-oxidizing *Chloroflexi* from yellowstone hot springs. *ISME J.* 14, 364–379. doi: 10.1038/s41396-019-0530-9

Takami, H., Taniguchi, T., Arai, W., Takemoto, K., Moriya, Y., and Goto, S. (2016). An automated system for evaluation of the potential functionome: MAPLE version 2. 1. 0. *DNA Res.* 23, 467–475. doi: 10.1093/dnares/dsw030

Takami, H., Taniguchi, T., Moriya, Y., Kuwahara, T., Kanehisa, M., and Goto, S. (2012). Evaluation method for the potential functionome harbored in the genome and metagenome. *BMC Genomics* 13:699. doi: 10.1186/1471-2164-13-699

Thiel, V., Fukushima, S.-I., Kanno, N., and Hanada, S. (2019). *Encyclopedia of Microbiology*, 4th Edn. Cambridge: Academic Press, 651–662. doi: 1016/B978-0-12-809633-8.20771-1

Thomas, S. C., Tamadonfar, K. O., Seymour, C. O., Lai, D., Dodsworth, J. A., Murugapiran, S. K., et al. (2019). Position-specific metabolic probing and metagenomics of microbial communities reveal conserved central carbon metabolic network activities at high temperatures. *Front. Microbiol.* 10:1427. doi: 10.3389/fmicb.2019.01427

van der Meer, M. T., Klatt, C. G., Wood, J., Bryant, D. A., Bateson, M. M., Lammerts, L., et al. (2010). Cultivation and genomic, nutritional, and lipid biomarker characterization of *Roseiflexus* strains closely related to predominant in situ populations inhabiting yellowstone hot spring microbial mats. *J. Bacteriol.* 192, 3033–3042. doi: 10.1128/JB.01610-9

van der Meer, M. T., Schouten, S., Bateson, M. M., Nübel, U., Wieland, A., Kühl, M., et al. (2005). Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone National Park. *Appl. Environ. Microbiol.* 71, 3978–3986. doi: 10.1128/AEM.71.7.3978-3986.2005

van der Meer, M. T., Schouten, S., Damsté, J. S., and Ward, D. M. (2007). Impact of carbon metabolism on 13C signatures of cyanobacteria and green non-sulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). *Environ. Microbiol.* 9, 482–491. doi: 10.1111/j.1462-2920.2006.01165.x

Ward, L. M., Hemp, J., Shih, P. M., McGlynn, S. E., and Fischer, W. W. (2018). Evolution of phototrophy in the *Chloroflexi* phylum driven by horizontal gene transfer. *Front. Microbiol.* 9:260. doi: 10.3389/fmicb.2018.00260

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., et al. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 46, D608–D617. doi: 10.1093/nar/gkx1089

Yamada, T., Imachi, H., Ohashi, A., Harada, H., Hanada, S., Kamagata, Y., et al. (2007). *Longilinea arvoryzae* gen. nov., sp. nov., strictly anaerobic, filamentous bacteria of the phylum *Chloroflexi* isolated from methanogenic propionate-degrading consortia. *Int. J. Syst. Evol. Microbiol.* 57, 2299–2306. doi: 10.1099/ijs.0.65098-0

Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H., et al. (2006). *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae* classis nov. and *Caldilineae* classis. *Int. J. Syst. Evol. Microbiol.* 56, 1331–1340. doi: 10.1099/ijs.0.64169-0

Yao, Y., Sun, T., Wang, T., Ruebel, O., Northen, T., and Bowen, B. P. (2015). Analysis of metabolomics datasets with high-performance computing and metabolite atlases. *Metabolites* 5, 431–442. doi: 10.3390/metabo5030431

Zarzycki, J., and Fuchs, G. (2011). Coassimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in *Chloroflexus aurantiacus*. *Appl. Environ. Microbiol.* 77, 6181–6188. doi: 10.1128/AEM.00705-11

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* 7, 203–214. doi: 10.1089/10665270050081478

Zuniga, M., Perez, G., and Gonzalez-Candelas, F. (2002). Evolution of arginine deiminase (ADI) pathway genes. *Mol. Phylogenet. Evol.* 25, 429–444. doi: 10.1016/S1055-7903(02)00277-4

Conflict of Interest: 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.

Citation: Thomas SC, Payne D, Tamadonfar KO, Seymour CO, Jiao J-Y, Murugapiran SK, Lai D, Lau R, Bowen BP, Silva LP, Louie KB, Huntemann M, Clum A, Spunde A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Chen I-M, Stamatis D, Reddy TBK, O'Malley R, Daum C, Shapiro N, Ivanova N, Kyrpides NC, Woyke T, Eloe-Fadrosh E, Hamilton TL, Dijkstra P, Dodsworth JA, Northen TR, Li W-J and Hedlund BP (2021) Genomics, Exometabolomics, and Metabolic Probing Reveal Conserved Proteolytic Metabolism of Thermoflexus hugenholtzii and Three Candidate Species From China and Japan. Front. Microbiol. 12:632731. doi: 10.3389/fmicb.2021.632731

Copyright © 2021 Thomas, Payne, Tamadonfar, Seymour, Jiao, Murugapiran, Lai, Lau, Bowen, Silva, Louie, Huntemann, Clum, Spunde, Pillay, Palaniappan, Varghese, Mikhailova, Chen, Stamatis, Reddy, O'Malley, Daum, Shapiro, Ivanova, Kyrpides, Woyke, Eloe-Fadrosh, Hamilton, Dijkstra, Dodsworth, Northen, Li and Hedlund. 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: 24 June 2021
doi: 10.3389/fmicb.2021.649594

Temporal Microbial Community Dynamics Within a Unique Acid Saline Lake

Noor-Ul-Huda Ghori[1,2], Michael J. Wise[2,3] and Andrew S. Whiteley[1,4]*

[1] School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, [2] The Marshall Centre of Infectious Diseases, School of Biological Sciences, The University of Western Australia, Perth, WA, Australia, [3] Department of Computer Science and Engineering, The University of Western Australia, Perth, WA, Australia, [4] Centre for Environment and Life Sciences, CSIRO Land Water, Perth, WA, Australia

Lake Magic is an extremely acidic, hypersaline lake found in Western Australia, with the highest concentrations of aluminum and silica in the world. Previous studies of Lake Magic diversity have revealed that the lake hosts acid- and halotolerant bacterial and fungal species. However, they have not canvassed microbial population dynamics across flooding, evapo-concentration and desiccation stages. In this study, we used amplicon sequencing and potential function prediction on sediment and salt mat samples. We observed that the bacterial and fungal diversity in Lake Magic is strongly driven by carbon, temperature, pH and salt concentrations at the different stages of the lake. We also saw that the fungal diversity decreased as the environmental conditions became more extreme. However, prokaryotic diversity was very dynamic and bacteria dominated archaeal species, both in abundance and diversity, perhaps because bacteria better tolerate the extreme variation in conditions. Bacterial species diversity was the highest during early flooding stage and decreased during more stressful conditions. We observed an increase in acid tolerant and halotolerant species in the sediment, involved in functions such as sulfur and iron metabolism, i.e., species involved in buffering the external environment. Thus, due to activity within the microbial community, the environmental conditions in the sediment do not change to the same degree as conditions in the salt mat, resulting in the sediment becoming a safe haven for microbes, which are able to thrive during the extreme conditions of the evapo-concentration and desiccation stages.

Keywords: poly-extremophilic, acid saline lake, taxogenomics, amplicon sequencing, Haloacidophiles

OPEN ACCESS

Edited by:
Brian P. Hedlund, University of Nevada, Las Vegas, United States

Reviewed by:
Melanie R. Mormile, Missouri University of Science and Technology, United States
Bonnie K. Baxter, Westminster College, United States

****Correspondence:***
Noor-Ul-Huda Ghori
noor-ul-huda.ghori@uwa.edu.au

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

Received: *05 January 2021*
Accepted: *04 June 2021*
Published: *24 June 2021*

Citation:
Ghori N-U-H, Wise MJ and Whiteley AS (2021) Temporal Microbial Community Dynamics Within a Unique Acid Saline Lake. Front. Microbiol. 12:649594. doi: 10.3389/fmicb.2021.649594

INTRODUCTION

Acid saline lakes represent one of the most extreme aquatic environments on Earth. They are poly-extreme ecosystems, exhibiting extremely acidic pH and salinities close to saturation. Such environments are of significant microbiological interest as they host organisms that are not only capable of withstanding pH and salinity stress, but also survive in the presence of additional stressors such as high metal concentrations and low nutrients (Mormile et al., 2007; Heidelberg et al., 2013; Johnson et al., 2015; Zaikova et al., 2018). Moreover, they serve as a reservoir of novel microbial functions such as acidophilic microorganisms that have been used for extracting metal ores from sulfide minerals (Dopson et al., 2017). Hence, such environments are

of significant interest to scientists interested in understanding the microbial diversity dynamics of communities residing there, as well as applied areas developing microbial consortia for bioprocessing applications (Dopson et al., 2017).

Lake Magic is one of the most extreme acidic hypersaline lakes (ca. 1 km in diameter) present within the Yilgarn Craton in Western Australia. This unique lake exhibits extremely low pH (<1.6) coupled to very high salinity (32% TDS) with the highest concentration of aluminum (1,774 mg/L) and silica (510 mg/L) in the world (Bowen and Benison, 2009; Conner and Benison, 2013). Lake Magic, similar to other lakes in Western Australia, has dynamic and characteristic stages of lake transformation, including flooding, evapo-concentration and desiccation that are driven by the local seasons (Bowen and Benison, 2009; Conner and Benison, 2013). The lake is fed via both regional acidic groundwater and infrequent precipitation (Benison et al., 2007).

Recent studies of the microbial diversity of Lake Magic have revealed that the lake hosts acidophilic, acid-tolerant, halophilic, and halotolerant bacterial species (Zaikova et al., 2018). In addition, fluid inclusions of halite crystals from Lake Magic exhibited the presence of micro-algae and prokaryotes trapped within them (Conner and Benison, 2013) while metagenomic analyses of lake water, groundwater and within the sediment of Lake Magic revealed that the lake is dominated by only a few species, such as *Salinisphaera*, and has a low representation of other bacterial species (Zaikova et al., 2018). Interestingly, eukaryotes, including fungi and green algae, were abundant in the pelagic zone, giving rise to the bright yellow color of the lake (Conner and Benison, 2013; Zaikova et al., 2018). These studies provide indicators of the population residing within the lake and the functional niches they occupy. However, they do not examine how the microbial populations in the lake change during different transformational stages of the lake. Molecular approaches to understand the dynamics of Yilgarn Craton lakes in the past have focused primarily on spatial composition at a single timepoint (Mormile et al., 2009; Johnson et al., 2015; Zaikova et al., 2018; Aerts et al., 2019). In contrast, temporal approaches to study microbial community dynamics have revealed considerable variability within microbial community compositions over time in various studies (Ju and Zhang, 2014; Nagarkar et al., 2018; Chénard et al., 2019; Cruaud et al., 2019), showing that some taxa remain consistent in their abundance whilst others exhibit sudden blooms (Nagarkar et al., 2018).

We hypothesize that the large fluctuations in environmental parameters during lake transformations are key drivers which will lead to marked changes in microbial populations and contribute to the mechanisms driving the dynamics of these communities (Cruaud et al., 2019). Studying the temporal dynamics of microbial communities in poly-extreme ecosystems such as Lake Magic could reveal crucial information about the drivers of community diversity and stability (Nagarkar et al., 2018). This study attempts to investigate the diversity and potential functional dynamics of prokaryotic and fungal communities over a period of 1 year. Using 16S rRNA gene and ITS gene amplicon sequencing data generated for salt mat and sediment samples, we elucidate the diversity patterns generated during lake transformations and assess likely functional profile of the resident microorganisms.

MATERIALS AND METHODS

Sample Collection

Core samples were collected from Lake Magic at timepoints across the flooding and evapo-concentration yearly cycle of the lake for 1 year (July 2017–July 2018). Multiple sampling sites were chosen around the lake for a single timepoint. Two types of samples were obtained from each core, namely a salt mat and a sediment sample (**Figure 1**). The salt mat layer changes physically during different stages of the lake, with the salt mat becoming more solid as the lake dries. With this in mind, the top 1 cm of the core was transferred to a clean falcon tube using a sterile spatula and regarded as the salt mat. The rest of the core was labeled as sediment (∼6 cm deep). All cores were taken with sterile core samplers and spatulas, and after every core the core samplers and spatulas were sterilized with 70% ethanol. A total of 5 sediment samples and 5 salt mat samples were collected. Temperature, pH, and salinity were measured for each sampling point using Orion 3 star Benchtop instrument. The pH and salinity were calibrated using standard solutions provided with the instrument kit. The pH and EC were measured in 1:5 sample-water extract.



FIGURE 1 | A single core of sample (∼6 cm deep) showing the salt mat and sediment sections. The bottom figure shows the salt mat (∼1 cm deep) with salt crystals on the top of sediment, sampled as a separate sample.

All samples were kept frozen at −20°C in 50 ml Falcon tubes in a portable freezer and moved to the laboratory freezer (−20°C) until further processing.

DNA Extraction, PCR and Amplicon Sequencing

DNA from sediment and salt mat samples was extracted using a method developed for acid saline sediments. Briefly, 0.4 *g* sediment and salt mat samples were taken in 2 ml tubes containing glass beads. To this 900 μl of extraction buffer (consisting of 0.2 M sodium phosphate buffer, 0.2% CTAB, 0.1 M NaCl, and 50 mM EDTA), 100 μl of 10% SDS and 10 μl of Proteinase K (20 mg/ml) were added. The mixture was kept at -80°C for 5 min and then heated at 70°C for 20 min. The mixture was subjected to mechanical agitation at 20 Hz for 20 min after which the tubes were kept on ice for 5 min. After bead beating the mixture was centrifuged at 10,000 × *g* for 5 min. To the supernatant 750 μl of chilled phenol-chloroform-isoamyl (pH 8) solution was added and centrifuged for 20 min at 10,000 × *g*. Aqueous layer was transferred to sterile tube and 600 μl of chilled chloroform -isoamyl (pH 8) solution was added. The mixture was centrifuged at 10,000 × *g* for 5 min. Next, to the aqueous layer 650 μl of 20% PEG, 2.5 M NaCl was added and incubated at 4°C for overnight. The solution was centrifuged for 20 min at 10,000 × *g*. The obtained pellet was washed with 70% ice cold ethanol and 2 μl glycogen. The final pellet was dissolved in 50 μl of TE buffer. All extractions were carried out in triplicate. Tubes containing no sample were incorporated as extraction blanks (controls) and were treated identical to sample extractions. DNA concentration was measured through fluorometry using a Qubit dsDNA HS Assay Kit with a Qubit 2.0 fluorometer (Life Technologies).

Extracted DNA from all timepoints were diluted 10-fold prior to PCR amplification of the 16S rRNA and ITS genes. All PCR reactions were carried out in triplicate. PCR amplification of the 16S rRNA gene V4-V5 region was performed using the universal PCR primer set 515F and 806R, targeting members within both bacterial and archaeal domains (Whiteley et al., 2012). The forward primer included the addition of an Ion Torrent PGM sequencing adapter, a GT spacer and a unique Golay barcode to facilitate multiplexed sequencing. Barcoded PCR reaction mixtures (20 μl) consisted of DNA template (1 μl), universal primer mix (untagged 515F and 806R at a final concentration of 0.2 μM), tagged 515F primer (0.2 μM), 600 ng BSA (Life Technologies) and 2.5 × 5′Hot Master Mix (5Primer, Australia). The PCR cycle was set at 94°C for 2 min followed by 25 cycles of 94 min for 45 s, 50°C for 60 s, and 65°C for 90 s. This was followed by 2 cycles of 94°C for 45 s, 65°C for 90 s and final extension at 65°C for 10 min.

Amplification of the fungal ITS regions was carried out using the universal primer set ITS1 F and ITS2 R (White et al., 1990; Gardes and Bruns, 1993), with the addition of an Ion Torrent PGM sequencing adapter, a GT spacer and a unique Golay barcode to the forward primer. The barcoded PCR primer mixtures (20 μl) included DNA template (1 μl), universal primer mix (untagged ITS1 F and ITS2 R at a final concentration of 0.2 μM), 600 ng BSA (Life Technologies), tagged ITS1 F primer (0.2 μM) and 2.5 × 5′ Hot Master Mix (5Primer, Australia). The PCR conditions included initial denaturation at 94°C for 2 min followed by 25 cycles of 94°C for 45 s, 50°C for 60 s, and 65°C for 90 s. This was followed by 9 cycles of 94°C for 45 s, 65°C for 90 s, and a final extension stage at 65°C for 10 min.

Several positive, no template (as negative controls) controls and DNA extraction controls (extraction blanks) were amplified along with the samples for both bacterial and fungal marker genes. PCR reaction performance was checked by loading PCR amplicons along with positive and negative controls on a 2% (w/v) agarose gel. The amplicons were quantified using a Qubit dsDNA HS Assay Kit on the Qubit 2.0 fluorometer (Life Technologies). All amplicons were subsequently pooled in one composite mixture at a concentration of 20 ng/μl, including negative controls. The pool was purified using AMPure XP (Beckman Coulter, Australia) and the quality of the pool was checked by visualizing on a 2% (w/v) agarose gel. The composite pool was sequenced on an Ion Torrent PGM.

Sequence Analysis and Statistical Analysis

Raw sequences were de-multiplexed and quality filtered through a custom QIIME version 1 pipeline (Quantitative Insights into Microbial Ecology; Caporaso et al., 2010) with a minimum average quality score of 20. The minimum sequence length was maintained at 130 b.p. and maximum sequence length of 350 b.p. Chimeric sequences were removed using USEARCH v6.1. No forward or reverse primer mismatches or barcode errors were allowed and maximum sequence homopolymers allowed were 15. The maximum number of ambiguous bases was set at six. *De novo* OTU picking was performed using ULCUST at 97% sequence identity cut off values and taxonomy was assigned through the Greengenes database (version 13.8). For fungal data, taxonomy was assigned using the SILVA v123 database (Quast et al., 2012).

The OTUs tables obtained for different levels of taxonomy were used as measures of taxa relative abundance in univariate statistical analysis. The OTUs detected in negative controls were manually removed from the data set. Alpha (α)- diversity at the phylum level for both 16S rRNA gene and ITS gene data was calculated using the richness, evenness and Shannon Weiner diversity index using the relative frequency table generated from a rarefied BIOM table. Data normality was checked with the Shapiro–Wilk test and log transformations of the data were performed where appropriate. Differences in the diversity for each stage and layer (sediment, salt mat), was calculated using a two-way analysis of variance (ANOVA). Tukey's HSD *post hoc* comparisons of groups were used to identify which groups were significantly different from each other.

Beta diversity (β) of microbial communities was calculated with non-metric multidimensional scaling (NMDS) using Bray-Curtis dissimilarity (Bray and Curtis, 1957) of OTUs table. Statistical significances of dissimilarity, based on temporal data and sample layer, was assessed using main effect and pairwise ANOSIM in R using the vegan package.

Phylogeny was inferred using Phylosift (Darling et al., 2014) for sequences that could not be classified past the domain level. The OTUs abundance and diversity patterns were calculated using the vegan package (Dixon, 2003) in R software (R Core Team, 2014). Plots and heat maps were produced using "ggplot2" (Wickham, 2011), ggpubr packages in R and the online suite Calypso (Zakrzewski et al., 2017).

Predicted microbial functions of the bacteria residing within Lake Magic were generated using FAPROTAX, using default settings. FAPROTAX is a manually constructed database that maps the microbial taxa to metabolic functions (Louca et al., 2016). The output from FAPROTAX was visualized in R using the ggplot2 package.

Chemical Analysis

Ten grams of sediment sample for each time point was oven dried at 60°C until completely desiccated. The sample was crushed, sieved to 2 mm fractions, packed in plastic bags and sent to the School of Agriculture and Environment, University of Western Australia chemical analysis laboratory. The analytical analysis included analysis of phosphorus, potassium, sulfur, organic carbon, nitrogen, iron, copper, sodium, boron, calcium, zinc, aluminum, magnesium, and manganese. These were extracted using Mehlich-3 extraction analyzed by inductively coupled plasma optical emission spectroscopy. Concentrations were expressed as percentage per weight and mg/kg.

RESULTS

Lake Stages and Physico-Chemical Parameters

A total of 5 time points representing 5 different stages of Lake Magic were sampled for this study. Although the lake did not go through some of the more extreme physical changes during the year, the transformation of one stage to the next was evident. Namely, we observed flooding, early evapo-concentration, mid evapo-concentration, late evapo-concentration and early flooding in the lake (**Figure 2**) details of which are presented in **Table 1** and **Table 2**. During the span of this study the lake did not reach complete desiccation due to heavy rainfall from 2017 to 2018 in Western Australia.

The first sampling point was during late winter to early summer in 2017, where the lake was filled with several centimeters of clear blue water (July 2017; **Figure 2A**). The lake sediment was rich in clay and a small amount of wet salt mat sample was acquired. In October 2017, the lake became shallower (**Figure 2B**) during early evapo-concentration (**Table 1**), characterized by clear water but where halite precipitation was evident on the lake shore. The lake transformed into a shallow yellow lake during January 2018 sampling (**Figure 2C**) where the surroundings of the lake were rich in halite precipitation, the lake itself exhibited a pungent acidic odor and the salt mat became desiccated and substantial. During mid-summer (March 2018) the lakebed became dry and a thick salt crust was observed on the surface of the salt mat (**Figure 2D**) with visible salt crystals (**Figure 1**), constituting the late evapo-concentration stage. At this stage the lake was also rich in iron oxide precipitation, which was evident due to its distinct red color of the sediment. The final sampling timepoint was the beginning of the flooding stage in July 2018 (**Figure 2E**) where the lake started to fill with water, with a concomitant dissolving of the halite and iron precipitation observed during the evapo-concentration stages. The chemical data has been summarized in **Table 2**.

Temporal Dynamics of the Lake Magic Microbiome

In order to assess the microbial community dynamics in the lake at different stages, we used 16S rRNA gene and ITS gene Ion

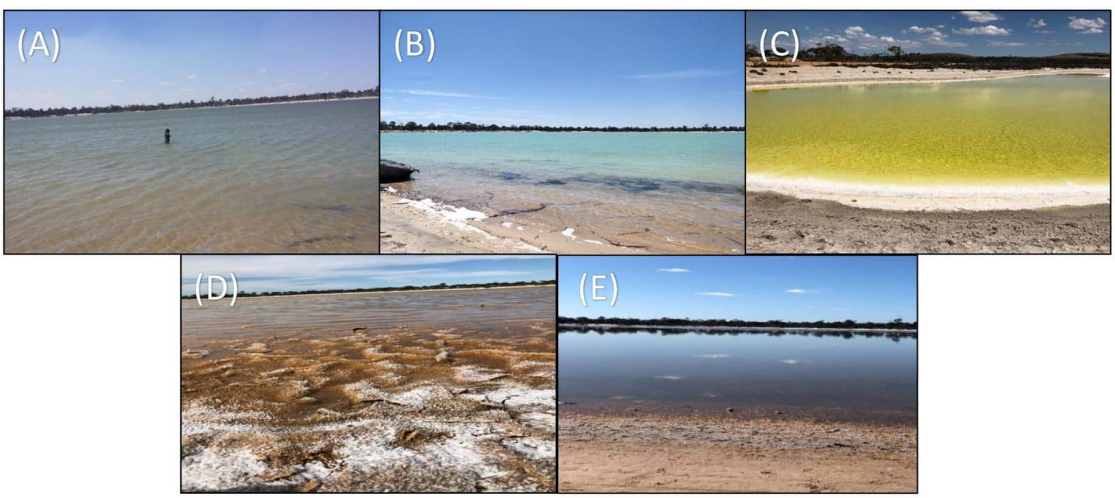


FIGURE 2 | Five stages of Lake Magic in a span of 1-year **(A)** July 2017, flooded **(B)** October 2017, early evapo-concentration **(C)** January 2018, mid evapo-concentration **(D)** March 2018, and late evapo-concentration **(E)** July 2018, early flooding.

TABLE 1 | Description of sampling site at different sampling time points for the water and sediment samples.

Lake stage	Sample code	Date collected	pH	EC(mS/cm)	Temperature (°C)	Lake description
Flooded	FL	July 2017	4.5 sediment 4.2 water	7 sediment 73.5 water	16	Clear blue lake water, very thin salt mat layer
Early evapo-concentration	EE	October 2017	4.5 sediment 4.1 water	11.3 sediment 78.3 water	25	Shallow blue water, salt mat becomes more evident
Mid evapo-concentration	ME	January 2018	4.5 sediment 3.4 water	15.95 sediment 146.9 water	28	Yellow slime water, salt foams forms on the lakeshore, thick salt mat layer develops, strong pungent smell in the air
Late evapo-concentration	LE	March 2018	4.2 sediment 2.7 water	41.8 sediment 223.8 water	33	A thick layer of salt mat with visible halite precipitation as crystals
Early flooding	EF	July 2018	4.8 sediment 3.41 water	55.9 sediment 226.7 water	24	Salt mat intact, water fills the dry lake bed, clear blue water

TABLE 2 | Chemical data for samples at different time points.

Stage	Sampling time	% Carbon	% Nitrogen	Al (mg/kg)	B (mg/kg)	Ca (mg/kg)	Cu (mg/kg)	Fe (mg/kg)	K (mg/kg)	Mg (mg/kg)	Mn (mg/kg)	Na (mg/kg)	P (mg/kg)	S (mg/kg)	Zn (mg/kg)
Flooded	July 2017	1.01	0.062	1,007	3.7	350	0.8	64	323	587	0.6	7,056	6	488	0.1
Early evapo-concentration	October 2017	0.597	0.061	1,023	3.6	11,308	0.4	47	357	733	0.9	9,526	10	8,716	0.1
Mid evapo-concentration	January 2018	1.30	0.088	1,256	5.7	2,750	0.6	52	598	1,462	1.5	19,260	12	2,591	0.2
Late evapo-concentration	March 2018	0.739	0.085	1,099	7.5	20,565	0.6	77	1,065	2,750	2.5	37,861	8	16,261	0.2
Early flooding	July 2018	1.56	0.064	1,186	6.8	10,790	0.7	33	781	2,661	1.5	33,759	7.4	9,190	0.2

All values shown are assessed using a single sample at each time point.

Torrent sequencing (15 Salt mats; 15 Sediments) in triplicate. After DNA extraction, all samples had detectable amounts of DNA, however, the DNA concentration was consistently higher for salt mat samples when compared to the sediment samples. Statistical analysis, however, showed that there was no significant difference (ANOVA $p > 0.05$) between diversity and richness indices of salt mat and sediment layers for both bacteria and fungi analyses (**Supplementary Figure 1**). Generally, we found DNA extraction easier from salt mat samples, relative to sediment samples, due to the lower amount of clay present in the salt mat samples.

A total of 330,913 reads were obtained for 16S rRNA gene microbiome analyses, representing 15,947 OTUs. The OTUs were assigned to 37 phyla, 135 classes, 260 orders, 426 families, and 739 genera of archaea and bacteria. Only OTUs which appeared in sequenced negative controls were discounted from the analyses and no other OTUs were filtered. This is because low abundance OTUs can have a significant effect on the diversity metrics for the microbial communities in low biomass environments, hence, filtering of OTUs with low representation in the microbiome can result in loss of crucial information. Analysis of the ITS gene sequences revealed a total of 724,490 reads, representing 4,005 OTUs and were assigned to 15 phyla, 44 classes, 86 order, 177 family and 259 genera.

Bacterial Communities Are Dynamic in Lake Magic

The Lake Magic microbiome OTUs richness was analyzed for salt mat and sediment samples on a relative frequency table constructed from a rarefied BIOM table (**Figure 3A**) at the phylum level which varied significantly between the five stages of the lake (ANOVA $p = 0.001$). The Chao 1 estimate varied from 9 to 23 for the 16S rRNA gene and the OTUs richness was significantly higher during the early flooding stage (mean richness = 21). The second highest OTUs richness was observed during the mid evapo-concentration stage (mean richness = 18). The late evapo-concentration stage showed the lowest mean richness of 13.6, whereas OTUs richness at the flooding stage was significantly different and showed a mean richness of 16. A mean richness of 15.5 was seen during the early evapo-concentration stage.

The Lake Magic microbiome diversity was also analyzed using a rarefied BIOM table at the phylum level. The diversity varied during the different lake stages (**Figure 3B**) and followed a similar trend to that of OTUs richness. The Shannon-Weiner diversity index ranged from 1.61 to 2.13 for bacterial communities, where the diversity was seen to increase during early evapo-concentration (mean diversity = 1.65) and the mid evapo-concentration stage (mean diversity = 1.87). However, the diversity decreased during the late evapo-concentration stage and was recorded as the lowest diversity index (mean diversity = 1.39) of all stages. The highest diversity was seen during the early flooding stage (mean diversity = 2.11). The ANOVA test was further analyzed with a *Tukey's post hoc test* for diversity and richness indices, which revealed that the diversity during flooding, late evapo-concentration and early flooding stages were significantly different from each other, whereas the richness was significantly different only during flooding and early flooding stages.

The Lake Magic bacterial composition under different lake stages was analyzed and is shown in **Supplementary Figure 2**.

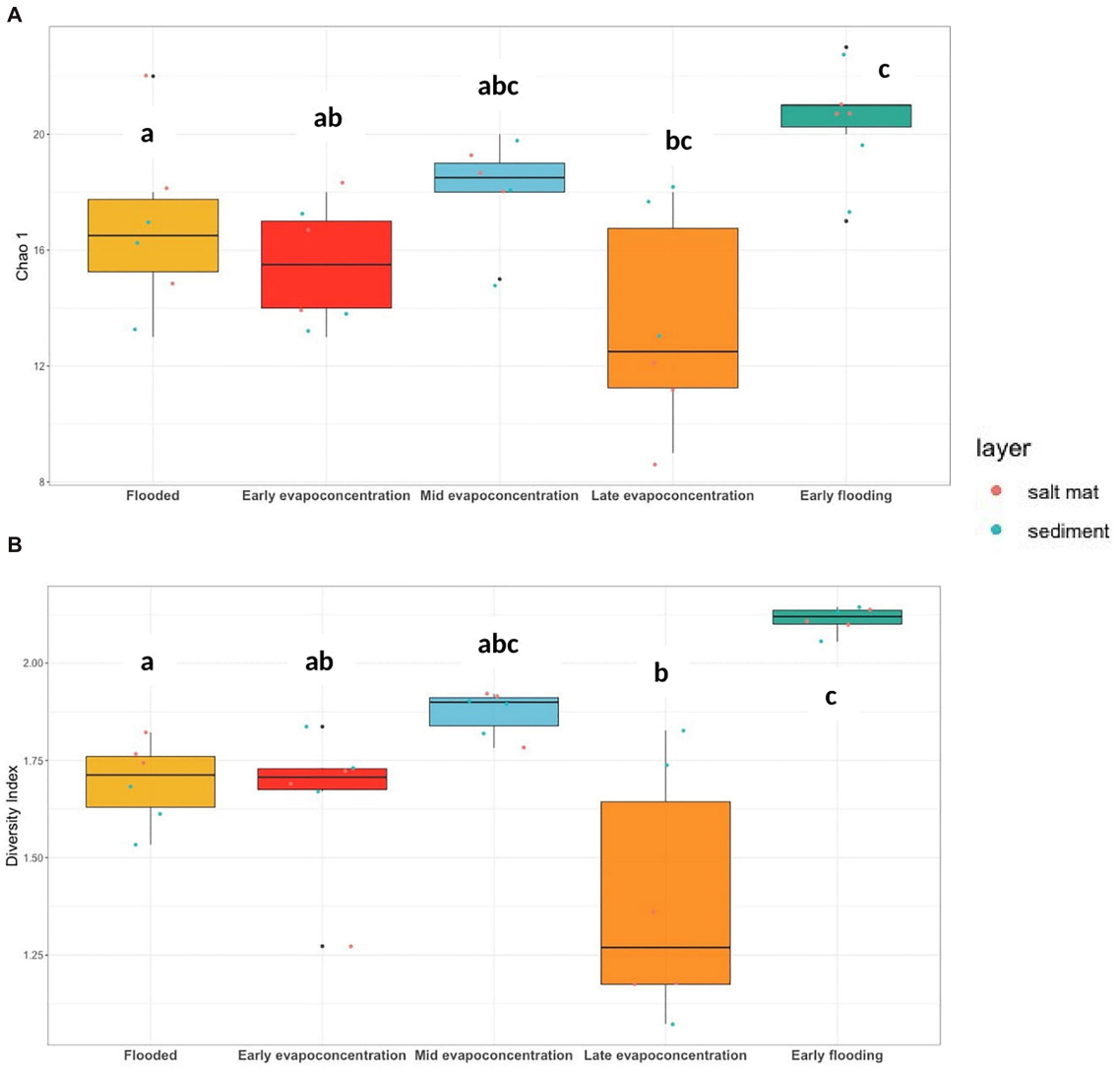


FIGURE 3 | (A) 16S rRNA gene OTUs richness, Chao 1 (ANOVA, $p = 0.001$) and **(B)** Shannon-Weiner diversity index (ANOVA, $p < 0.001$) at phylum level. The letters show significant differences as obtained with *Tukey's post hoc* test.

Archaeal community diversity in the microbiome was represented by only two phyla, the *Crenarchaeota* and the *Euryarchaeota,* whilst the bacterial domain contributed the dominant OTUs observed within the 16S rRNA gene analyses. The majority of the sequences for bacteria originated from two phyla: *Bacteriodetes* (20%) and *Proteobacteria* (39%), most of which could not be classified below family level, indicating that a large proportion of the bacterial taxa in Lake Magic appear to be relatively poorly characterized.

The Bacterial Community Becomes More Specialized as Stress Increases

For taxa which could be classified to genera, varying trends were observed during different lake stages and within the sample layers (**Supplementary Figure 3**). Specifically, key bacterial genera fluctuated within the salt mat and sediment during the various lake stages. The correlation analysis of chemical data with the bacterial diversity revealed that microbial dynamics is strongly driven by salinity, temperature, pH and carbon content in the lake (**Supplementary Figure 4**). For instance, members of the *Acidiphilium* genus were low in abundance during the flooding stage in both the salt mat and sediment samples, whilst their relative abundance was seen to increase within the salt mat during evapo-concentration stage as the lake conditions became more stressful. A significant increase (ANOVA, $p < 0.05$) in *Acidiphilium* relative abundance was observed within the salt mat during the late evapo-concentration stage and significantly decreased during the early flooding stage of the lake (**Supplementary Figure 3**, *Acidiphilium*). Similarly, the relative abundance

of *Acidobacterium* was high within sediments during the flooding stage and a significant increase in abundance within the salt mat was seen during all evapo-concentration stages (**Supplementary Figure 3**, *Acidobacterium*). Similar to *Acidiphilium,* the relative abundance of *Acidobacterium* decreased when the lake was in the early flooding stage. In contrast, sequences belonging to *Arthrobacter, Bacillus, Flavbacterium, and Sporosarcina* increased significantly in relative abundance during the early flooding stage in both the sediment and salt mat, whilst genera such as *Nitrososphaera* were only present during the early flooding stage in high relative abundance.

Interestingly, it was also observed (**Supplementary Figure 3**) that the relative abundance of *Sulfurimonas, Syntrophobacter, Halothiobacillus, Acidobacterium, Acidiphilium,* and *Alicyclobacillus* decreased during the early flooding stage in both the salt mat and the sediment, whilst, the *Syntrophobacter* population increased in relative abundance within the sediment during late evapo-concentration. During the flooding stage *Salinisphaera* was found to be more abundant in the sediment when compared to the salt mat. However, in the late evapo-concentration stage, its abundance increased in the salt mat (**Supplementary Figure 3**).

The Fungal Community Becomes Less Diverse Under Extreme Conditions Within Lake Magic

The diversity and richness of the fungal community within Lake Magic was analyzed similar to 16S rRNA gene analysis using the rarefied BIOM table at the phylum level. The diversity of fungi significantly decreased (ANOVA, $p < 0.001$) as the lake conditions became more extreme. The richness index ranged from 2 to 7 whereas the Shannon diversity index ranged from 0.004 to 1.28 (**Figure 4A,B**). The highest mean richness was observed for the flooding and early evapo-concentration stage. The OTUs richness consistently decreased after the early evapo-concentration stage, with lowest OTUs richness observed for the early flooding stage (mean richness = 2). The fungal diversity showed a varying trend when compared to OTUs richness, where the highest diversity index was observed for the flooding stage (mean diversity index = 1.11) whereas the lowest was observed

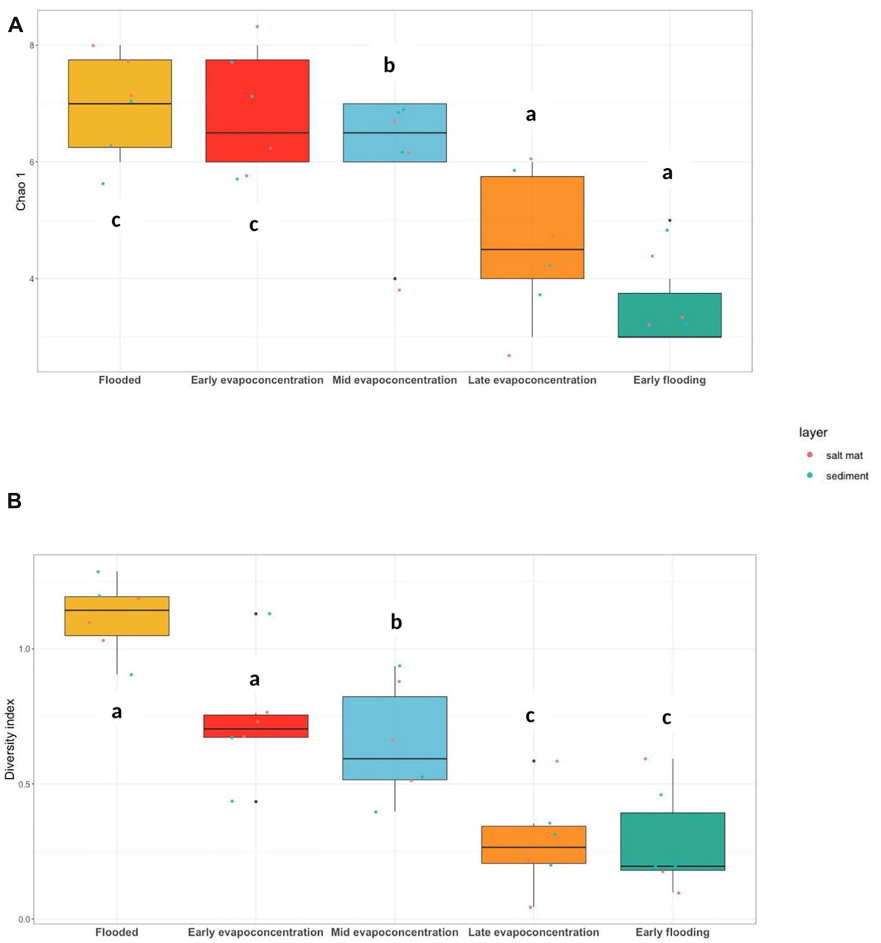


FIGURE 4 | (A) ITS gene OTUs richness, Chao 1 (ANOVA, $p = 0.001$) and **(B)** Shannon-Weiner diversity index (ANOVA, $p < 0.001$) at phylum level. The letters show significant differences as obtained with *Tukey's post hoc* test.

for the extreme early flooding stage (mean diversity index = 0.28). A Tukey's *post hoc* test revealed that the flooding and early evapo-concentration stage were statistically similar to each other, whilst, the middle evapo-concentration and late evapo-concentration stages were statistically similar, and the early flooding stage was significantly different from all other stages. Additionally, a Tukey's *post hoc* test for OTUs richness showed that the flooding, early evapo-concentration and early flooding stages were significantly different from all other stages.

Interestingly, when the fungal community composition was visualized (**Supplementary Figure 5**), an increase in unidentified fungi belonging to the *Ascomycota* phylum was observed. This indicated that a large portion of the fungi living in Lake Magic are likely unidentified. The salt mat during the early flooding stage was, however, the most diverse when compared to other time points and was abundant with the genus *Cladosporium*. Variation in the composition of other genera including *Fusarium, Ulocladium,* and *Hostaea* was also seen.

Comparing the Bacterial and Fungal Communities

The dissimilarity between microbial communities at different lake stages was assessed using NMDS using Bray–Curtis dissimilarity indices made with the OTUs table at phylum level. The sediment and salt mat bacterial communities from the flooding, mid evapo-concentration and late evapo-concentration stages tightly clustered together (**Figure 5A**). In contrast, the salt mat and sediment communities under the early flooding stage clustered separately (ANOSIM, $R^2 = 0.49$, and $p = 0.001$). However, similar to alpha diversity, when an ANOSIM test was applied to determine the variability in the salt mat and sediment layers, no significant difference was observed (ANOSIM, $R^2 = 0.05$, and $p = 0.103$).

The NMDS analysis of fungal communities showed less clustering when compared to the bacterial community analysis (**Figure 5B**), apart from the early flooding stage, where beta diversity of the fungal community was statistically different (ANOSIM, $R^2 = 0.462$, and $p = 0.001$). However, no significant difference was found in the beta diversity of fungi in the sediment versus the salt mat layers.

Finally, it is interesting to directly compare the bacterial and archaeal communities' abundances and diversities at the different time points (**Supplementary Figures 2, 5**, respectively). We see that, in general, the bacterial community was both more abundant and more diverse, with the exception of the early-flooding time point in the salt mat sample, where the diversities were similar, but the archaeal community was more abundant.

Predicted Ecological Functions of the Bacterial Communities

The ecological functions of the bacterial communities were predicted using the FAPROTAX pipeline for the different stages of the lake. These analyses indicated aerobic chemoheterotrophy, iron oxidation and sulfur-related pathways as likely dominant functions within the lake's bacterial community (**Figure 6**). Functions related to carbon metabolism, such as methylotrophy and methanotrophy were predicted but were not abundant. Nitrogen-related functions included nitrification and ammonia oxidation and were observed to fluctuate between different lake stages. For instance, these data indicated that nitrification and ammonia oxidation-predicted functions in the lake were significantly higher within the salt mat during the early flooding stage (**Supplementary Figure 6**). Interestingly, the nitrogen-related predicted functions were also high in the sediment samples during the early evapo-concentration and mid evapo-concentration lake stage. Sulfur-related predicted functions were lower during the early flooding stage, but significantly increased in the sediment layer during all evapo-concentration stages. Sulfur-related predicted function was found to be highest in the sediment during the early evapo-concentration stage. Predicted functions related to iron oxidation and reduction were relatively low at all timepoints, but significantly increased during the early evapo-concentration and late evapo-concentration stage within the sediment layer. Finally, iron-based respiration predicted pathways fluctuated more frequently when compared to predicted sulfur and nitrogen-related functions during all lake stages (**Supplementary Figure 6**).

DISCUSSION

Acid saline lakes in Western Australia host unique microorganisms that can be studied to understand life under extreme conditions, novel biogeochemical processes and new biotechnological avenues (Benison et al., 2007). In this study, the bacterial and fungal microbial community dynamics were studied using a temporal approach to resolve how an extreme lake microbiome changes during different stress phases due to changes in the physico-chemical properties of the habitat.

Previous studies on acidic hypersaline lakes in Western Australia revealed a high level of bacterial diversity within Lake Magic and other Western Australia lakes water (Mormile et al., 2009; Zaikova et al., 2018), but have relied upon single time (lake stage) sampling points. Our results indicate that the alpha (α)-diversity of the bacterial and fungal microbial communities differed significantly between the less extreme (flooding) stage and the more extreme (late evapo-concentration) stage. These differences indicated that the lake microbiome diversity is driven to a large degree by the high salt and low pH conditions within the lake (Podell et al., 2014). It has previously been reported for hypersaline environments that the ionic concentration in these environments hinders the solubility of oxygen, and hence, the oxygen concentration is very low in acid saline lakes (Sherwood et al., 1991). However, the majority of microorganisms isolated from hypersaline environments are aerobic heterotrophs that are also capable of fermentation (Dyall-Smith, 2009). Our results are also in line with these previous findings. Moreover, our results show that the number of OTUs in the salt mat layer were consistently higher in all samples. These results are also similar to the results obtained by Aerts et al. (2019) for four acid saline lakes in Western Australia, and suggest that the microbial community residing in the salt mat likely has increased access to light, water

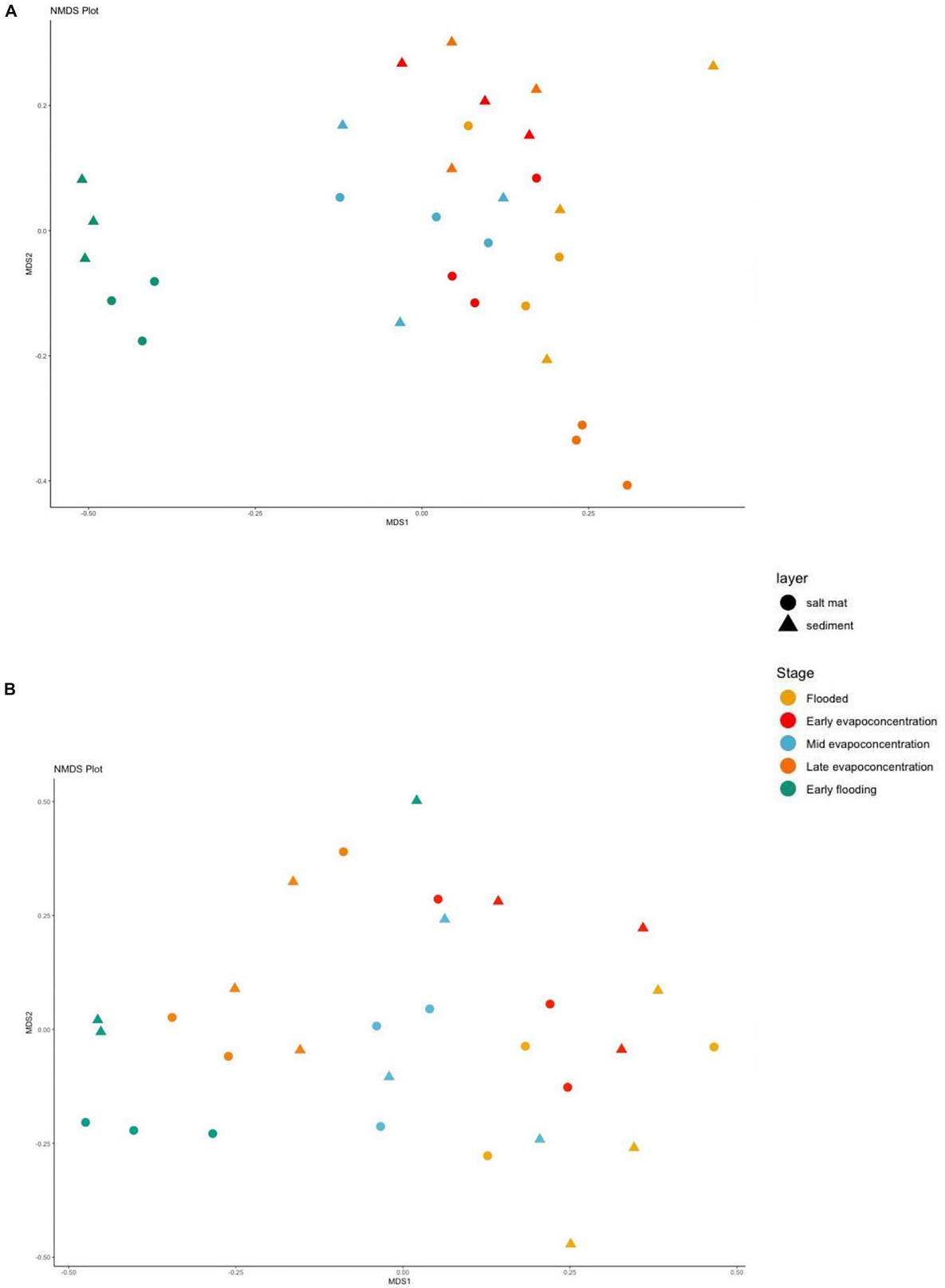


FIGURE 5 | Beta diversity shown in a Non-metric multidimensional scaling (NMDS) plot using Bray-Curtis dissimilarity indices for **(A)** 16S rRNA **(B)** ITS sequences for different time points and layers at phylum level. Samples are colored according to the time points and shaped according to the layer.

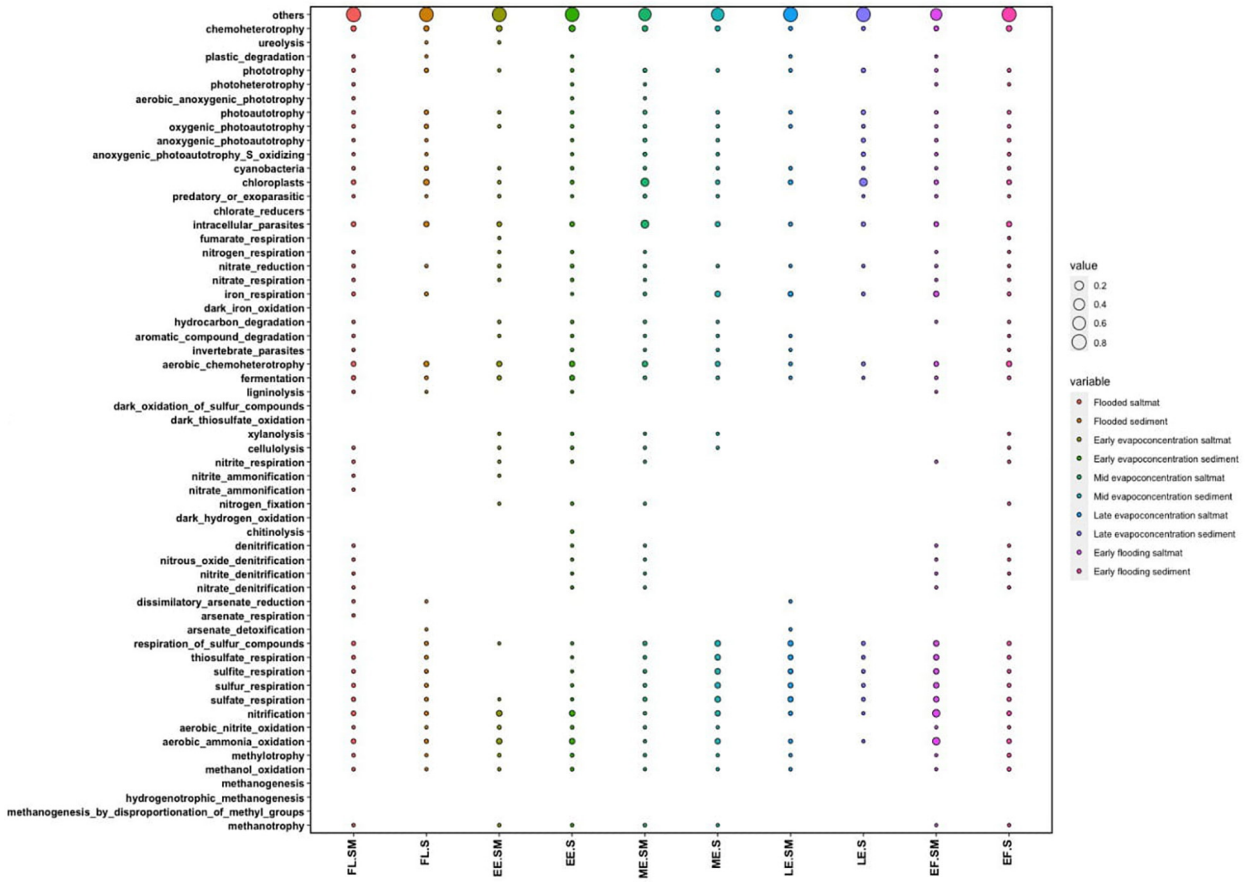


FIGURE 6 | Dot plot showing the FAPROTAX predicted functions of the bacterial community within the salt mat and sediment samples at different lake stages.

and oxygen. This is also explained by the availability of oxygen at the air/water interface, which exhibits an unequal distribution of oxygen between the salt mat and the sediment layer. Since the salt mat is directly in contact with the water column in the lake there is higher availability of oxygen locally (Podell et al., 2014). Comparing the microbial composition of the salt mat and sediment in Lake Magic also indicates that microorganisms more tolerant to high salt and pH conditions are selected within the salt mat, where the environmental conditions such as salinity and pH are harsher and more dynamic compared to those in the sediment layer.

The prokaryotic community at all stages of the lake cycle was dominated by halotolerant and acidophilic bacteria, whilst archaeal taxa were found in much lower abundances. Archaeal sequences were derived from only two phyla and these observations are in good agreement with Johnson et al. (2015) who observed low archaeal diversity within four acidic hypersaline lakes in Yilgarn Craton. It has been reported for Lake Tyrell in Victoria, Australia, that succession of different microorganisms is dependent upon the solutes present in the lakes (Podell et al., 2014). Hence, the variation in solute concentration in Lake Magic at different stages is likely driving the succession of archaea and bacteria. It could also be hypothesized that bacterial taxa outcompete archaeal taxa in these unique environments because of the dynamic nature of the Western Australian lakes, where environmental changes are frequent and, therefore, stress tolerant species are selected, as opposed to obligate archaeal extremophiles (Benison and Bowen, 2006). However, it should also be considered that low representation of archaeal sequences might be because of the primer sequences used in this study.

Organic carbon and nitrogen concentrations were found to be low at all lake stages and increased slightly during the flooded (FL) stages, in line with previous studies in similar lakes in Western Australia (Ruecker et al., 2016; Aerts et al., 2019). Similarly, phosphorus levels were also low in Lake Magic, as in other Western Australian lakes (Aerts et al., 2019), with phosphorus often being considered a limiting factor along with carbon and nitrogen for the growth of microorganisms (Elser et al., 2007). Microbiome analyses indicated an abundant community of heterotrophs within the lake and hence, low carbon, nitrogen and phosphorus are likely to be responsible for the low the biomass in these samples (Aerts et al., 2019). However, carbon and nitrogen levels were highest during the mid evapo-concentration and

these increases are likely due to photosynthetic inputs via blooming of the halotolerant alga *Dunaliella*. This is also reflected in the increased microbial diversity during the mid evapo-concentration stage. Blooming of algae commonly occurs during this lake stage and provides a source of photosynthetically derived nutrient input which causes increased diversity in the lake (Zaikova et al., 2018).

During the evapo-concentration stages iron concentrations increased which was also mirrored in the predicted microbiome functions of increased iron respiration activity within the bacterial community. Iron metabolism is considered an important function of the sediment inhabitants and is thought to be responsible for the increasing of pH in the sediment that is, becoming less acidic (Lu et al., 2016; Zaikova et al., 2018). The members of *Alicyclobacillus* genus are reported to be involved in iron oxidation along with archaea in other acid saline lakes in Western Australia (Johnson et al., 2015; Lu et al., 2016). The members of the genus are also capable of reducing iron (Yahya et al., 2008; Lu et al., 2010). Previous metagenomic studies of lake magic indicated that *Alicyclobacillus* was the most abundant genus within the sediment. Moreover, the genus *Acidiphilium* was also found to be abundant in the sediment samples (Zaikova et al., 2018) with members of this genus are involved in iron reduction (Weber et al., 2006; Sánchez-Andrea et al., 2011; Zaikova et al., 2018). These data indicated that the relative abundance of the members of *Alicyclobacillus* and *Acidiphilium* varied across different lake stages and is likely more complicated than first thought, in terms of lake biogeochemistry (Zaikova et al., 2018). *Alicyclobacillus* abundance was significantly higher in sediment samples during the flooding stage and was significantly increased during the early evapo-concentration stage within the salt mat samples. In contrast the genus *Acidiphilium* was more abundant within the salt mat samples during the later, late evapo-concentration stage. Interestingly the overall abundance of these genera decreased during the higher stress stages of the lake.

Previously it has been suggested that archaea dominate iron oxidation in acid saline extreme environment and that the role of bacteria is limited. Moreover, the iron activity is known to decrease by the presence of high salt in the environment (Lu et al., 2016). In the study, *Acidiphilium* spp. were isolated from an acid saline lake in Western Australia which formed long filamentous structures during low pH and high salinity conditions indicating that these species have developed a coping mechanism for extreme stress (Lu et al., 2016). Examining the functional profile of our samples it can be seen that predicted iron respiration increases significantly in the sediment during the early evapo-concentration and late evapo-concentration stage and is the lowest in the sediment during the flooding stage potentially because of the coping mechanism adapted by these species.

The activity of acidophiles is known to decrease significantly in presence of high concentration of chloride ions (Suzuki et al., 1999; Shiers et al., 2005). However, Lu et al. (2016) found that *Acidiphilium* spp. formed long filaments as a coping mechanism to high chloride ions within the Dalyup River in Western Australia. As iron metabolism is affected by the presence of chloride ions we hypothesize that this may explain the bacterial spp. of *Alicyclobacillus* and *Acidiphilium* being dominant during early evapo-concentration and flooding stages. It is also suggested that under the high chloride conditions, based upon community data, these species are responsible of increasing the pH in the sediment. Hence, we postulate from these results that microorganisms in Lake Magic adapt to stressful conditions of pH and salinity not only physiologically, but also play crucial roles in changing their external environment making it more habitable for themselves and other members of the lake.

One of the most abundant genera detected in these data was *Salinisphaera* (7%), in both the sediment and salt mat. *Salinisphaera* species are mesophilic, halotolerant and slightly acidophilic (pH range 5.0–7.5), surviving in a range of moderately acidic and saline conditions as well as high concentrations of metal ions. Species belonging to genus *Salinisphaera* have been isolated from a range of environments, including hydrothermal vents, solar salterns, brine from salt wells, seawater and marine fish surfaces (Antunes et al., 2003; Mormile et al., 2007; Crespo-Medina et al., 2009; Bae et al., 2010; Gi et al., 2010; Park et al., 2012; Zhang et al., 2012; Shimane et al., 2013). Critically, these species are able to metabolize both autotrophically and heterotrophically (Antunes et al., 2003, 2011; Crespo-Medina et al., 2009; Bae et al., 2010; Park et al., 2012; Zhang et al., 2012; Shimane et al., 2013). In addition, *Salinisphaera* species are also involved in the uptake of iron and siderophore production (Antunes et al., 2003). Molecular studies here indicated that the abundance of *Salinisphaera* was consistently high in salt mats during all lake stages, except during the FL stage, where it was more abundant within the sediment. *Salinisphaera* was previously reported to be the single dominant OTUs in the lake water (Zaikova et al., 2018) during evapo-concentration in Lake Magic. These findings together indicate that most of the *Salinisphaera* spp. reside in the water column of Lake Magic. The dramatic increase in its representation during the late evapo-concentration stage in the salt mat and sediment suggests that it is highly tolerant of extreme pH and acidic environmental conditions, including tolerance to heavy metal ions, allowing it to survive through the evapo-concentration stages.

Members of genus *Sulfurimonas* have been isolated from diverse environments such as hydrothermal vents, marine sediments and terrestrial habitats and are known to play an important role in chemoautotrophic processes (Han and Perner, 2015). The members of this genus can grow on a variety of electron donors and acceptors and, thus, are able to colonize disparate environments. These include different reduced sulfur compounds such as sulfide, elemental sulfur, sulfite, and thiosulfate (Han and Perner, 2015). Many members of the genus are also involved in nitrogen and hydrogen metabolism. In our samples, *Sulfurimonas* (1.2%) relative abundance continuously decreased in salt mats but increased in relative abundance within the sediments as the environmental conditions became more stressful. Ultimately, it decreased significantly during the most extreme (late evapo-concentration) lake stages but still maintained a low representation in

the sediment. When examining the functional profile data, predicted sulfur functions significantly increased in the sediment samples during the evapo-concentration stages, suggesting that *Sulfurimonas* spp. play a crucial role in cycling key nutrients in the sediment. Since the members of the genus are able to survive chemolithoautotrophically using various electron acceptors and donors (Campbell et al., 2006; Grote et al., 2008), it suggests that these species are capable of adapting to the changing environmental conditions of Lake Magic. A similar trend of abundance was seen for *Flavobacterium, Bacillus, and Syntrophobacter*, where their abundances increased in the sediment during the evapo-concentration stages and we postulate that they play a crucial role in niche construction by interacting with other dominant taxa as the stress increases.

The fungal diversity was seen to decrease as the environmental conditions became more acidic and saline. A single phylum, *Ascomycota* (76.8%), became dominant as the lake became dry and hypersaline. Most of the members of this phylum are unidentified, and hence, most of the fungi residing in Lake Magic are either novel, or the sequence length of the marker gene used in this study is not sufficient to be able to characterize these members. It can be deduced that these fungi are tolerant of the acidic hypersaline conditions of the lake. In a previous study on Lake Magic microbiology, the water samples were found to be highly diverse in terms of eukaryotic community composition, being abundant (∼98.5%) in Ascomycota. *Aspergillus* and *Penicillium* were the most abundant genera in Lake Magic water. However, the results for sediment samples were quite similar to our results (Zaikova et al., 2018). The presence of the halotolerant algae *Dunaliella* during evapo-concentration stages is a major contributor to carbon content in the lake. It can be deduced from the data that as *Dunaliella* increase in the lake (during middle evapo-concentration) the diversity in the salt mat increases rapidly. Interestingly, this difference in the diversity of lake water and sediment column indicates the exchange between the two compartments in terms of nutrition. The water column is more abundant in oxygen and has access to sunlight whilst, the higher composition of eukaryotes in the water column which contribute to the carbon and nitrogen content in the salt mat and the sediment.

In conclusion, the paper describes the temporal diversity of life living in a unique poly-extreme environment. Specifically, we studied the microbiome of one of the most extreme environments on earth, having a pH of about 1.5 and salinity seven times that of sea water. These environments are highly under-studied. Moreover, the temporal dynamics of poly-extreme environments have not previously been investigated. The findings in this paper highlight many potential survival strategies of microbes living in a poly-extremophile environment. The results of the study also point out the role of microbial interactions as a survival strategy of microbes in such an environment (Zengler and Zaramela, 2018). Finally, we found previously unknown bacterial species that are found in these environments and need to be investigated further.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI under the project ID PRJNA721776.

AUTHOR CONTRIBUTIONS

N-U-HG conducted field sampling, carried out the research, analyzed, and wrote the manuscript. MW and AW designed the study and wrote the manuscript. All authors reviewed and edited the manuscript.

FUNDING

This work was supported by International Postgraduate Research Scholarship (IPRS) by University of Western Australia awarded to N-U-HG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.649594/full#supplementary-material

REFERENCES

Aerts, J. W., van Spanning, R. J. M., Flahaut, J., Molenaar, D., Bland, P. A., Ehrenfreund, P., et al. (2019). Microbial communities in sediments from four mildly acidic ephemeral salt lakes in the yilgarn craton (Australia) – terrestrial analogs to ancient mars. *Front. Microbiol.* 10:779.

Antunes, A., Eder, W., Fareleira, P., Santos, H., and Huber, R. (2003). Salinisphaera shabanensis gen. nov., sp. nov., a novel, moderately halophilic bacterium from the brine-seawater interface of the Shaban deep. Red sea. *Extremophiles* 7, 29–34. doi: 10.1007/s00792-002-0292-5

Antunes, A., Ngugi, D. K., and Stingl, U. (2011). Microbiology of the Red sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* 3, 416–433. doi: 10.1111/j.1758-2229.2011.00264.x

Bae, G. D., Hwang, C. Y., Kim, H. M., and Cho, B. C. (2010). Salinisphaera dokdonensis sp. nov., isolated from surface seawater. *Int. J. Syst. Evol. Microbiol.* 60, 680–685. doi: 10.1099/ijs.0.010058-0

Benison, K. C., and Bowen, B. B. (2006). Acid saline lake systems give clues about past environments and the search for life on Mars. *Icarus* 183, 225–229. doi: 10.1016/j.icarus.2006.02.018

Benison, K. C., Beitler Bowen, B., Oboh-Ikuenobe, F. E., Jagniecki, E. A., Laclair, D. A., Story, S. L., et al. (2007). Sedimentology of acid saline lakes in Southern Western Australia: newly described processes and products of an extreme environment. *J. Sediment. Res.* 77, 366–388. doi: 10.2110/jsr.2007.038

Bowen, B. B., and Benison, K. C. (2009). Geochemical characteristics of naturally acid and alkaline saline lakes in southern Western Australia. *Appl. Geochem.* 24, 268–284. doi: 10.1016/j.apgeochem.2008.11.013

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27, 325–349. doi: 10.2307/1942268

Campbell, B. J., Engel, A. S., Porter, M. L., and Takai, K. (2006). The versatile ε-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* 4, 458–468. doi: 10.1038/nrmicro1414

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.

Chénard, C., Wijaya, W., Vaulot, D., Lopes, dos Santos, A., Martin, P., et al. (2019). Temporal dynamics of bacteria, Archaea and protists in equatorial coastal waters. *Sci. Rep.* 9:16390.

Conner, A. J., and Benison, K. C. (2013). Acidophilic halophilic microorganisms in fluid inclusions in halite from lake magic, Western Australia. *Astrobiology* 13, 850–860. doi: 10.1089/ast.2012.0956

Crespo-Medina, M., Chatziefthimiou, A., Cruz-Matos, R., Pérez-Rodríguez, I., Barkay, T., Lutz, R. A., et al. (2009). Salinisphaera hydrothermalis sp. nov., a mesophilic, halotolerant, facultatively autotrophic, thiosulfate-oxidizing gammaproteobacterium from deep-sea hydrothermal vents, and emended description of the genus Salinisphaera. *Int. J. Syst. Evol. Microbiol.* 59, 1497–1503. doi: 10.1099/ijs.0.005058-0

Cruaud, P., Vigneron, A., Fradette, M. S., Dorea, C. C., Culley, A. I., Rodriguez, M. J., et al. (2019). Annual protist community dynamics in a freshwater ecosystem undergoing contrasted climatic conditions: the saint-charles river (Canada). *Front. Microbiol.* 10:2359.

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

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x

Dopson, M., Holmes, D. S., Lazcano, M., McCredden, T. J., Bryan, C. G., Mulroney, K. T., et al. (2017). Multiple osmotic stress responses in Acidihalobacter prosperus result in tolerance to chloride ions. *Front. Microbiol.* 7:2132.

Dyall-Smith, M. (2009). *'The Halohandbook. Protocols for Haloarchaeal Genetics', Version 7.2*.

Elser, J. J., Bracken, M. E. S., Cleland, E. E., Gruner, D. S., Harpole, W. S., Hillebrand, H., et al. (2007). Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol. Lett.* 10, 1135–1142. doi: 10.1111/j.1461-0248.2007.01113.x

Gardes, M., and Bruns, T. D. (1993). ITS primers with enhanced specificity for Basidomycetes-application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2, 113–118. doi: 10.1111/j.1365-294x.1993.tb00005.x

Gi, D. B., Chung, Y. H., Hye, M. K., and Byung, C. C. (2010). Salinisphaera dokdonensis sp. nov., isolated from surface seawater. *Int. J. Syst. Evol. Microbiol.* 60, 680–685.

Grote, J., Jost, G., Labrenz, M., Herndl, G. J., and Jürgens, K. (2008). Epsilonproteobacteria represent the major portion of chemoautotrophic bacteria in sulfidic waters of pelagic redoxclines of the baltic and black seas. *Appl. Environ. Microbiol.* 74, 7546–7551. doi: 10.1128/aem.01186-08

Han, Y., and Perner, M. (2015). The globally widespread genus Sulfurimonas: versatile energy metabolisms and adaptations to redox clines. *Front. Microbiol.* 6:989.

Heidelberg, K. B., Nelson, W. C., Holm, J. B., Eisenkolb, N., Andrade, K., and Emerson, J. B. (2013). Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell, Australia. *Front. Microbiol.* 4:115.

Johnson, S. S., Chevrette, M. G., Ehlmann, B. L., and Benison, K. C. (2015). Insights from the metagenome of an acid salt lake: the role of biology in an extreme depositional environment. *PLoS One* 10:e0122869. doi: 10.1371/journal.pone.0122869

Ju, F., and Zhang, T. (2014). Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* 9, 683–695. doi: 10.1038/ismej.2014.162

Louca, S., Parfrey, L. W., and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science* 353, 1272–1277. doi: 10.1126/science.aaf4507

Lu, S., Gischkat, S., Reiche, M., Akob, D. M., Hallberg, K. B., and Küsel, K. (2010). Ecophysiology of Fe-cycling bacteria in acidic sediments. *Appl. Environ. Microbiol.* 76, 8174–8183. doi: 10.1128/aem.01931-10

Lu, S., Peiffer, S., Lazar, C. S., Oldham, C., Neu, T. R., Ciobota, V., et al. (2016). Extremophile microbiomes in acidic and hypersaline river sediments of Western Australia. *Environ. Microbiol. Rep.* 8, 58–67. doi: 10.1111/1758-2229.12351

Mormile, M. R., Hong, B. Y., and Benison, K. C. (2009). Molecular analysis of the microbial communities of Mars Analog Lakes in Western Australia. *Astrobiology* 9, 919–930. doi: 10.1089/ast.2008.0293

Mormile, M. R., Hong, B., Adams, N. T., Benison, K. C., and Oboh-Ikuenobe, F. (2007). "Characterization of a moderately halo-acidophilic bacterium isolated from Lake Brown, Western Australia," in *Proceedings of the Instruments, Methods, and Mission for Astrobiology* (San Diego, CA).

Nagarkar, M., Countway, P. D., Du, Yoo, Y., Daniels, E., Poulton, N. J., et al. (2018). Temporal dynamics of eukaryotic microbial diversity at a coastal Pacific site. *ISME J.* 12, 2278–2291. doi: 10.1038/s41396-018-0172-3

Park, S. J., Cha, I. T., Kim, S. J., Shin, K. S., Hong, Y., Roh, D. H., et al. (2012). Salinisphaera orenii sp. nov., isolated from a solar saltern. *Int. J. Syst. Evol. Microbiol.* 62, 1877–1883. doi: 10.1099/ijs.0.028647-0

Podell, S., Emerson, J. B., Jones, C. M., Ugalde, J. A., Welch, S., Heidelberg, K. B., et al. (2014). Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *ISME J.* 8, 979–990. doi: 10.1038/ismej.2013.221

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nuc. Acids. Res.* 41, D590–D596.

R Core Team (2014). *R: a language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ruecker, A., Schröder, C., Byrne, J., Weigold, P., Behrens, S., and Kappler, A. (2016). Geochemistry and mineralogy of Western Australian Salt Lake sediments: implications for *Meridiani planum* on Mars. *Astrobiology* 16, 525–538. doi: 10.1089/ast.2015.1429

Sánchez-Andrea, I., Rodríguez, N., Amils, R., and Sanz, J. L. (2011). Microbial diversity in anaerobic sediments at Río Tinto, a naturally acidic environment with a high heavy metal content. *Appl. Environ. Microbiol.* 77, 6085–6093. doi: 10.1128/aem.00654-11

Sherwood, J. E., Stagnitti, F., Kokkinn, M. J., and Williams, W. D. (1991). Dissolved oxygen concentrations in hypersaline waters. *Limnol. Oceanogr.* 36, 235–250. doi: 10.4319/lo.1991.36.2.0235

Shiers, D. W., Blight, K. R., and Ralph, D. E. (2005). Sodium sulphate and sodium chloride effects on batch culture of iron oxidising bacteria. *Hydrometallurgy* 80, 75–82. doi: 10.1016/j.hydromet.2005.07.001

Shimane, Y., Tsuruwaka, Y., Miyazaki, M., Mori, K., Minegishi, H., Echigo, A., et al. (2013). Salinisphaera japonica sp. nov., a moderately halophilic bacterium isolated from the surface of a deep-sea fish, *Malacocottus gibber*, and emended description of the genus Salinisphaera. *Int. J. Syst. Evol. Microbiol.* 63, 2180–2185. doi: 10.1099/ijs.0.047845-0

Suzuki, I., Lee, D., Mackay, B., Harahuc, L., and Oh, J. K. (1999). Effect of various ions, pH, and osmotic pressure on oxidation of elemental sulfur by *Thiobacillus thiooxidans*. *Appl. Environ. Microbiol.* 65, 5163–5168. doi: 10.1128/aem.65.11.5163-5168.1999

Weber, K. A., Achenbach, L. A., and Coates, J. D. (2006). Microorganisms pumping iron: Anaerobic microbial iron oxidation and reduction. *Nat. Rev. Microbiol.* 4, 752–764. doi: 10.1038/nrmicro1490

White, T. J., Bruns, T. D., Lee, S. B., and Taylor, J. W. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *PCR Protocols: A Guide to Methods and Applications,* eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (New York, NY: Academic Press), 315–322.

Whiteley, A. S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., et al. (2012). Microbial 16S rRNA Ion tag and community metagenome sequencing using the Ion torrent (PGM) platform. *J. Microbiol. Methods* 91, 80–88. doi: 10.1016/j.mimet.2012.07.008

Wickham, H. (2011). ggplot2. Wiley. Interdiscip. *Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147

Yahya, A., Hallberg, K. B., and Johnson, D. B. (2008). Iron and carbon metabolism by a mineral-oxidizing Alicyclobacillus-like bacterium. *Arch. Microbiol.* 189, 305–312. doi: 10.1007/s00203-007-0319-5

Zakrzewski, M., Proietti, C., Ellis, J. J., Hasan, S., Brion, M. J., Berger, B., et al. (2017). Calypso: a user friendly web-server for mining and visualizing microbiome-environment interactions. *Bioinformatics* 33, 782–283. doi: 10.1093/bioinformatics/btw725

Zaikova, E., Benison, K. C., Mormile, M. R., and Johnson, S. S. (2018). Microbial communities and their predicted metabolic functions in a desiccating acid salt lake. *Extremophiles* 22, 367–379. doi: 10.1007/s00792-018-1000-4

Zengler, K., and Zaramela, L. S. (2018). The social network of microorganisms — how auxotrophies shape complex communities. *Nat. Rev. Microbiol.* 16, 383–390. doi: 10.1038/s41579-018-0004-5

Zhang, Y. J., Tang, S. K., Shi, R., Klenk, H. P., Chen, C., Yang, L. L., et al. (2012). Salinisphaera halophila sp. nov., a moderately halophilic bacterium isolated from brine of a salt well. *Int. J. Syst. Evol. Microbiol.* 62, 2174–2179. doi: 10.1099/ijs.0.035584-0

Conflict of Interest: 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 © 2021 Ghori, Wise and Whiteley. 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: 25 June 2021
doi: 10.3389/fmicb.2021.639995

Interplay of Various Evolutionary Modes in Genome Diversification and Adaptive Evolution of the Family *Sulfolobaceae*

Rachana Banerjee[1†], Narendrakumar M. Chaudhari[1†], Abhishake Lahiri[1,2], Anupam Gautam[1,3†], Debaleena Bhowmik[1,2], Chitra Dutta[1‡], Sujay Chattopadhyay[4], Daniel H. Huson[5,6] and Sandip Paul[1,2†]*

[1] Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata, India, [2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India, [3] Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research, Kolkata, India, [4] JIS Institute of Advanced Studies and Research, JIS University, Kolkata, India, [5] Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany, [6] Cluster of Excellence: Controlling Microbes to Fight Infection, Tübingen, Germany

Sulfolobaceae family, comprising diverse thermoacidophilic and aerobic sulfur-metabolizing Archaea from various geographical locations, offers an ideal opportunity to infer the evolutionary dynamics across the members of this family. Comparative pan-genomics coupled with evolutionary analyses has revealed asymmetric genome evolution within the *Sulfolobaceae* family. The trend of genome streamlining followed by periods of differential gene gains resulted in an overall genome expansion in some species of this family, whereas there was reduction in others. Among the core genes, both *Sulfolobus islandicus* and *Saccharolobus solfataricus* showed a considerable fraction of positively selected genes and also higher frequencies of gene acquisition. In contrast, *Sulfolobus acidocaldarius* genomes experienced substantial amount of gene loss and strong purifying selection as manifested by relatively lower genome size and higher genome conservation. Central carbohydrate metabolism and sulfur metabolism coevolved with the genome diversification pattern of this archaeal family. The autotrophic CO_2 fixation with three significant positively selected enzymes from *S. islandicus* and *S. solfataricus* was found to be more imperative than heterotrophic CO_2 fixation for *Sulfolobaceae*. Overall, our analysis provides an insight into the interplay of various genomic adaptation strategies including gene gain–loss, mutation, and selection influencing genome diversification of *Sulfolobaceae* at various taxonomic levels and geographical locations.

Keywords: thermoacidophilic archaea, pan-genome, genome evolution, positive selection, metabolic pathways

OPEN ACCESS

Edited by:
Rafael R. de la Haba, University of Seville, Spain

Reviewed by:
Changyi Zhang, University of Illinois at Urbana–Champaign, United States
Flavia Figueira Aburjaile, Federal University of Minas Gerais, Brazil

****Correspondence:***
Sandip Paul
websandip@gmail.com; sandippaul@iicb.res.in

†Present address:
Rachana Banerjee and Sandip Paul, JIS Institute of Advanced Studies and Research, JIS University, Kolkata, India
Narendrakumar M. Chaudhari, Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University Jena, Jena, Germany
Anupam Gautam, Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany; International Max Planck Research School "From Molecules to Organisms," Max Planck Institute for Developmental Biology, Tübingen, Germany

‡Former Chief Scientist

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

Received: *10 December 2020*
Accepted: *06 May 2021*
Published: *25 June 2021*

Citation:
Banerjee R, Chaudhari NM, Lahiri A, Gautam A, Bhowmik D, Dutta C, Chattopadhyay S, Huson DH and Paul S (2021) Interplay of Various Evolutionary Modes in Genome Diversification and Adaptive Evolution of the Family Sulfolobaceae. Front. Microbiol. 12:639995. doi: 10.3389/fmicb.2021.639995

INTRODUCTION

The microbial genome divergence is a dynamic evolutionary process of ecological and genetic differentiation. Various mechanisms that both create and maintain the phenotypic diversity in genomes include gene gain–loss, mutation, natural selection, and genetic drift (Loewe and Hill, 2010). Among these, the gene gain–loss is considered to be one of the important factors affecting the microbial evolution, and further gene gain–loss along with habitat isolation are proposed to be the key points of ecological speciation in microbes (Polz et al., 2013). Several other studies have

also shown that ecological divergence and natural selection may play important roles in genome diversification (Vos, 2011; Shapiro et al., 2016). Geographical isolation along with genetic drift also impacts the biogeographic patterns of microorganisms and is responsible for microbial biodiversity (Martiny et al., 2006; Andam et al., 2016).

A considerable fraction of the genetic diversity is likely to be considered as regulators of adaptive features specific for that particular habitat. Deciphering these adaptive signals encrypted in the genomes of any organism is important from both evolutionary and ecological perspectives. The complex nature of these signals due to the multilayer interactions either with cohabitants or with environment makes the task extremely challenging. In this context, one important module of research is the evolutionary dynamics of extremophiles that offer insights into how their genome-wide variations, linked to structural and functional profiles, are shaped by extreme environmental attributes (Horikoshi and Grant, 1998; Seckbach, 2000). Extremophiles can withstand and flourish in environments that are otherwise considered to be too extreme to support any terrestrial life forms. Extreme environments are generally characterized by extreme temperatures, pressures, pH, and salinity (Rothschild and Mancinelli, 2001). The archaeal domain, being a substantial part of the extremophiles, has specifically evolved and acquired adaptations to sustain such extreme environments (Reed et al., 2013). Availability of high-quality, large-volume, and low-cost genomic data has augmented the study to deal with the global properties of the genomes and helped to reveal the various adaptive mechanisms to survive in extreme environments. During the past decade, several studies have reported on the influence of genomic attributes including relative synonymous codon usage bias, dinucleotide abundance, purine loading, and horizontal gene transfers (HGTs) on sustaining extreme conditions (Lynn et al., 2002; Das et al., 2006; Paul et al., 2008; Dutta and Paul, 2012; van Wolferen et al., 2013).

One powerful approach to unraveling the molecular mechanisms of adaptation is the comparative genomics study with a large number of complete genome sequences. In this regard, the pan-genome approach has been found to be valuable in the comparative analysis for a set of microbial genomes (Lefébure and Stanhope, 2007; Tettelin et al., 2008). Conventionally, the entire collection of genes present in diverse strains of a particular species is termed as "pan-genome" (Tettelin et al., 2008; Brockhurst et al., 2019). On the one hand, site-specific recombination circulated through mobile genetic elements and gene gain–loss events are associated with the flexible or accessory genes and thereby offer genomic and metabolic diversity in a population; the core genomic part, on the other hand, is linked with vertical transmission and homologous recombination, providing a stable genomic and metabolic backbone for the population. As the core clusters embody a set of highly conserved genes, phylogenetic analysis on the basis of core gene clusters provides information on the evolutionary history of its constituent organisms (Lefébure and Stanhope, 2007; Chattopadhyay et al., 2012a; Freschi et al., 2019). Thus, comparative pan-genomic study coupled with evolutionary analyses could be used as a framework to investigate the genome diversification associated with life under extreme conditions.

In the present study, we have used a comprehensive set of extremophilic Archaea from the family *Sulfolobaceae*. The family consists of thermoacidophilic organisms, with optimal growth temperatures ranging from 65 to 90°C and optimum growth at pH values of approximately 2. Another notable feature is the aerobic sulfur oxidation, which is restricted to the members of this family among Archaea (Friedrich et al., 2005; Wu et al., 2021). Several studies illustrate *Sulfolobus*, *Saccharolobus* (*Saccharolobus solfataricus* previously *Sulfolobus solfataricus*) *Acidianus*, and *Metallosphaera* as the principal components and enriched consortium of the microbial communities from hot springs (Kozubal et al., 2008; Sakai and Kurosawa, 2018). The genus *Sulfolobus* has enormous importance from biotechnological viewpoint as members of this genus are a source of new enzymes, biomaterials, and metabolic pathways and also efficient in oxidative desulfurization of sour crude oil and have a role in decreasing the cost of processing of petroleum (Benvegnu et al., 2009; Bräsen et al., 2014; Besse et al., 2015; Littlechild, 2015; Nazari et al., 2017; Quehenberger et al., 2017). Also, the most commonly studied sulfur-oxidizing genera *Sulfolobus*/*Saccharolobus* comprised species displaying a wide variety of genomic characteristics. Despite being biologically quite similar and dwelling in comparable environments, there are obvious differences between *Sulfolobus islandicus* and *Sulfolobus acidocaldarius* strains (Mao and Grogan, 2017). For instance, among *S. islandicus* isolates, through a comparative genomic study, it has been observed that geographical isolation plays a significant role in its increased genomic diversity between populations (Reno et al., 2009). Several studies indicate that genomes diverge genetically and metabolically according to the geographic locations, supporting the fact that both core and accessory parts are restricted because of discontinuous habitat (Whitaker, 2003; Mao and Grogan, 2017). On the other hand, for another species of this genus *S. acidocaldarius*, the genome is remarkably conserved across different geographical locations, and the evolutionary basis for this inconsistency in genome diversity of *S. acidocaldarius* with *S. islandicus* is yet to be understood. Previous works have discussed several factors that account for the high preservation among geographically different *S. acidocaldarius* strains, for example, propagule dispersal, differential ability to colonize and replace local populations, resistant nature toward mobile elements, etc. (Brügger et al., 2002; Chen et al., 2005; Redder and Garrett, 2006; Anderson et al., 2017; Mao and Grogan, 2017; Wagner et al., 2017).

Thus, the complete genomes of the isolates from the family *Sulfolobaceae* offer an excellent opportunity to grasp the role of HGT, recombination, and mutation in their differential genomic and metabolic diversification in presence or absence of selection pressures. Toward this, we performed a comparative pan-genomic and subsequent evolutionary analysis for publicly available complete genomes of the members of family *Sulfolobaceae* and deciphered the contingent contributions of gene gain–loss, mutation, and recombination in shaping up the asymmetric evolution of this archaeal family. We believe that our study sheds light on the understanding of fundamental

evolutionary processes along with the selective pressures that have shaped genomic and metabolic diversification of the family *Sulfolobaceae*.

MATERIALS AND METHODS

Sequence Retrieval and Annotation

Fully assembled chromosomes of 30 organisms belonging to the family *Sulfolobaceae* including genera *Sulfolobus* (15 strains), *Saccharolobus* (6 strains), *Metallosphaera* (7 strains), and *Acidianus* (2 strains) from different geographical locations were considered from the NCBI GenBank resource[1] (**Supplementary Table 1**). We extracted the genome annotations available in NCBI for all the 30 genomes used in the present study. We also reannotated the genomes using the RAST annotation server (Aziz et al., 2008) and compared them with the annotation extracted from NCBI. Both the annotations were found to be 99.01% ± 0.68% identical. We used NCBI-provided annotations for further analysis.

Pan-Genome Construction

For pan-genomic profile construction, we used all the annotated protein coding genes (total 76,369) from 30 genomes under study (**Supplementary Table 1**). We implemented CD-HIT Suite (Fu et al., 2012) for clustering of all the amino acid sequences by considering 50% sequence identity and alignment coverage as orthologous clustering criteria (Edgar, 2010; Li et al., 2012; Steinegger and Söding, 2018; Tonkin-Hill et al., 2020) to construct strain-wise gene presence/absence–abundance pan-matrix. We also used OrthoMCL (Li et al., 2003) for the identification of orthologous groups applying 50% sequence identity and alignment coverage and implemented GET_HOMOLOGUES software package (Contreras-Moreira and Vinuesa, 2013) for detailed pan-genome analysis along with the reconstruction of orthologous gene families. Both pan-matrices (derived from CD-HIT and OrthoMCL) were then used to identify and extract core/mosaic/unique clusters. In order to maintain uniqueness for core gene sets, in case of duplicates, only one sequence from each organism is considered in each gene cluster. In this way, we removed all paralogous and duplicate sequences. Then, we selected as core only those orthologous gene clusters detected as core gene families by both CD-HIT and OrthoMCL algorithms. In order to perform different evolutionary analyses, we extracted the nucleotide sequence of core clusters.

Apart from carrying out comparative genomics at the intergenus level considering four different genera of the same family *Sulfolobaceae*, we also carried out pan-genomic studies at species level separately for four species containing multiple strains, viz., *S. acidocaldarius*, *S. islandicus*, *S. solfataricus*, and *Metallosphaera sedula*. For reconstruction of pan-genomes for these four species separately, we followed the same steps as mentioned previously for gene clustering of the entire *Sulfolobaceae* family.

[1]http://www.ncbi.nlm.nih.gov/genbank/

Study of Recombination

We identified probable recombination events for each of the core clusters using the Recombination Detection Program (RDP4) default settings (Martin et al., 2015). This software includes an array of eight different recombination-detection algorithms. These algorithms are from three different methods: phylogenetic (BOOTSCAN, RDP, and SISCAN), substitution (GENECONV, MAXCHI, CHIMAERA, and LARD), and distance comparison (PHYLPRO). We assigned a gene region to be tentatively affected by homologous recombination event only by finding significant ($P < 0.05$) results in at least three out of these different recombination detection algorithms in RDP4.

Phylogenetic Study

We used ClustalW (Larkin et al., 2007) to align amino acid sequences of core genes and FASconCAT (Kück and Meusemann, 2010) to concatenate the alignments. Two core phylogenies were created based on the variable regions of the concatenated alignments of all core genes as well as non-recombinant core genes, without considering the gaps by using Jones–Taylor–Thornton substitution model (Jones et al., 1992) under maximum-likelihood method of MEGA 7 (Kumar et al., 2016). The pan-genome phylogeny was constructed by performing a parsimony analysis in MEGA 7 by converting the gene presence/absence–abundance pan-matrix into a binary matrix and finally to hypothetical sequence by replacing 0 and 1 with A and T, respectively. All the phylogenetic profiles created in our study were statistically supported through 1,000 bootstrap iterations.

Gene Gain–Loss Identification

In order to identify gene gains and losses, we used the COUNT (Csûös, 2010) software, applying asymmetric Wagner parsimony with gain penalty 1.6 and loss penalty 1. We considered the pan-genome phylogeny as the evolutionary trajectory of the genomes and the gene presence/absence–abundance pan-matrix as input in COUNT for gene gain–loss calculation. To confirm the COUNT results using a different approach, we implemented Sankoff's parsimony algorithm (Sankoff, 1975). For each gene, ancestral presence (1) and absence (0) states were computed using a gain penalty of 1.6 and loss penalty of 1.0, favoring late occurrences of changes during trace-back. Based on Sankoff's algorithm, for each node, the number of genes gained (state change from 0 to 1) and lost (state change from 1 to 0), and the number of genes present were computed in a tree traversal. The script implementing Sankoff's parsimony algorithm was written in Python 3.8.0 and available on request.

Functional Annotation and Metabolic Pathway Prediction

Functional annotations of selected genes were performed using Archaeal Clusters of Orthologous Genes (arCOGs) database (Makarova et al., 2015) (2014 update). We used the best BlastP hits (Camacho et al., 2009) for each representative protein sequence against arCOGs database. In order to retain the best hits, sequence identity cutoff of 50% with at least 50%

query length coverage and *e*-value cutoff of 1.0E-5 were used. Functional annotations of selected genes for Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) assignments and prediction of KEGG pathways were performed using KEGG Automatic Annotation Server using Bidirectional Best Hits method with default parameters (Moriya et al., 2007). A χ^2 test with Yates' correction and two-tailed *P* values has been used to check the significant differences between the gain and loss for each arCOG and KEGG categories for a particular species (**Supplementary Table 2**). The *P* value stands for probability and quantifies how likely it is that any observed difference between groups is due to chance.

Comparative Analysis of Metabolic Pathways

Literature survey and the information from the *Sulfolobus* Systems Biology (SulfoSYS) Project[2] (Zaparty et al., 2010) were used to categorize 71 genes for Central Carbohydrate Metabolism (CCM). BlastP (50% sequence identity and alignment coverage) search of these 71 CCM genes against *Sulfolobaceae* unfiltered pan-genome yielded 2,329 homologs from 105 gene clusters with 71 different KO terms (**Supplementary Table 3**).

We identified a total of 135 and 12 genes for "Sulfur metabolism" and "Nitrogen metabolism," respectively, from KEGG Pathway Database (Kanehisa and Goto, 2000). Homology search (as followed for CCM pathway) yielded 32 gene clusters (459 genes, 17 KO terms) and 14 gene clusters (191 genes, 11 KO terms) for sulfur and nitrogen metabolism, respectively (**Supplementary Tables 4, 5**). The abundance matrices for the three pathways were then converted into three single linkage clustering along with heatmaps by ClustVis web tool on the basis of Ward method using Manhattan distance (Metsalu and Vilo, 2015).

Evolutionary Rate Calculation

We calculated the evolutionary rate dN/dS (ω) by considering all possible pairs (225,765) of 30 *Sulfolobaceae* genomes for each of the 519 core protein sequence clusters, using the "yn00" program (Yang and Nielsen, 2000) from Phylogenetic Analysis by Maximum Likelihood package. Synonymous sites and ω values were considered to be saturated or unreliable, when, $dS < 0.1$, or $dN/dS \geq 99$, for a pair of sequences. Therefore, we discarded those pairs of sequences before calculating the average of ω values. Next, the average dN, dS, and dN/dS were calculated for all possible combinations at species and strain levels from 519 core genes.

Identification of Positive Selection by the Branch-Site Method and Log-Likelihood Ratio Test

A total of 2,412 gene pairs (having $dN > dS$) comprising 354 protein sequence clusters were chosen as positively selected gene pairs out of the 225,765 gene pairs from the core group and subjected to further analysis through branch-site method (BSM) and log-likelihood ratio test (LRT) (Nielsen and Yang, 1998) for testing the statistical significance of positive selection along lineages. The lnL values of the 354 clusters were calculated for both the selection model and null model. The LRT values for each of the clusters were then calculated as follows: $\Delta LRT = 2 \times (lnL1 - lnL0)$, where lnL1 represents the lnL values resulting from the selection model, and lnL0 represents the same from the null model (Vishnu et al., 2015). A significant result with the branch-site model using χ^2 test ($df = 1$, $P < 0.05$) pointed toward the fact that positive selection has affected a particular cluster during a specific evolutionary time.

RESULTS AND DISCUSSION

Genome Diversification and Differential Patterns of Gene Gain–Loss Events

The core genome phylogeny (**Supplementary Figure 1A**) constructed for the *Sulfolobaceae* family depicts the genome divergence pattern similar to that of the pan-genome tree (**Supplementary Figure 1B**) representing the balance of recent gene gains and losses in various ecosystems. Further weighted parsimony-based gene gain–loss estimation detected a set of 2,354 genes in the genome of the common ancestor for all members of the *Sulfolobaceae* family (**Figure 1**). There was an initial phase of genome shrinkage as the four genera began to diversify and the genome streamlining continued for most of the species groups. The internal or terminal nodes that were represented by either single strain or multiple strains of a particular species experienced an overall level of gene gain with one exception of *S. acidocaldarius*. At the strain level, this gene gain–loss dynamics varied a lot, where a majority of strains of a particular species experienced a large number of gene gains–losses at the terminal nodes, suggesting the acceleration of these events during the recent diversification of each strain. In general, terminal nodes were characterized by much higher frequency of gene gains than losses. In contrast, a significant amount of gene deletion was observed in case of two *S. acidocaldarius* strains (DSM 639 and N8) isolated from different geographic locations and *S. solfataricus* strain P2 from Italy. It should be noted that SULB and SULC resulted from extensive passage for an adaptive laboratory experiment from SULA; thus, they were practically the same strain with few genetic changes between them (McCarthy et al., 2015). The derivative strains for *S. solfataricus* and *M. sedula* showed primarily the gene loss events, implying restricted gene flow due to laboratory conditions and loss of function via mutation or deletion (Ai et al., 2016). Again, for all four genera, considerable amount of genome size variation of constituent species can be described by the differential phenomenon of gene acquisition and loss. Overall, the analysis outlined a dynamic gene gain–loss model through the genome diversification of *Sulfolobaceae* family. While *S. islandicus*, *S. solfataricus*, *Sulfolobus* sp. A20, *Sulfolobus tokodaii*, and *Acidianus manzaensis* experienced a genome expansion, *S. acidocaldarius*, *Metallosphaera cuprina*, *M. sedula*, and *Acidianus hospitalis* lost genes in course of their genome diversification. Specifically for the multistrain species, this result

[2]http://www.sulfosys.com/ccm-pathway-map.html

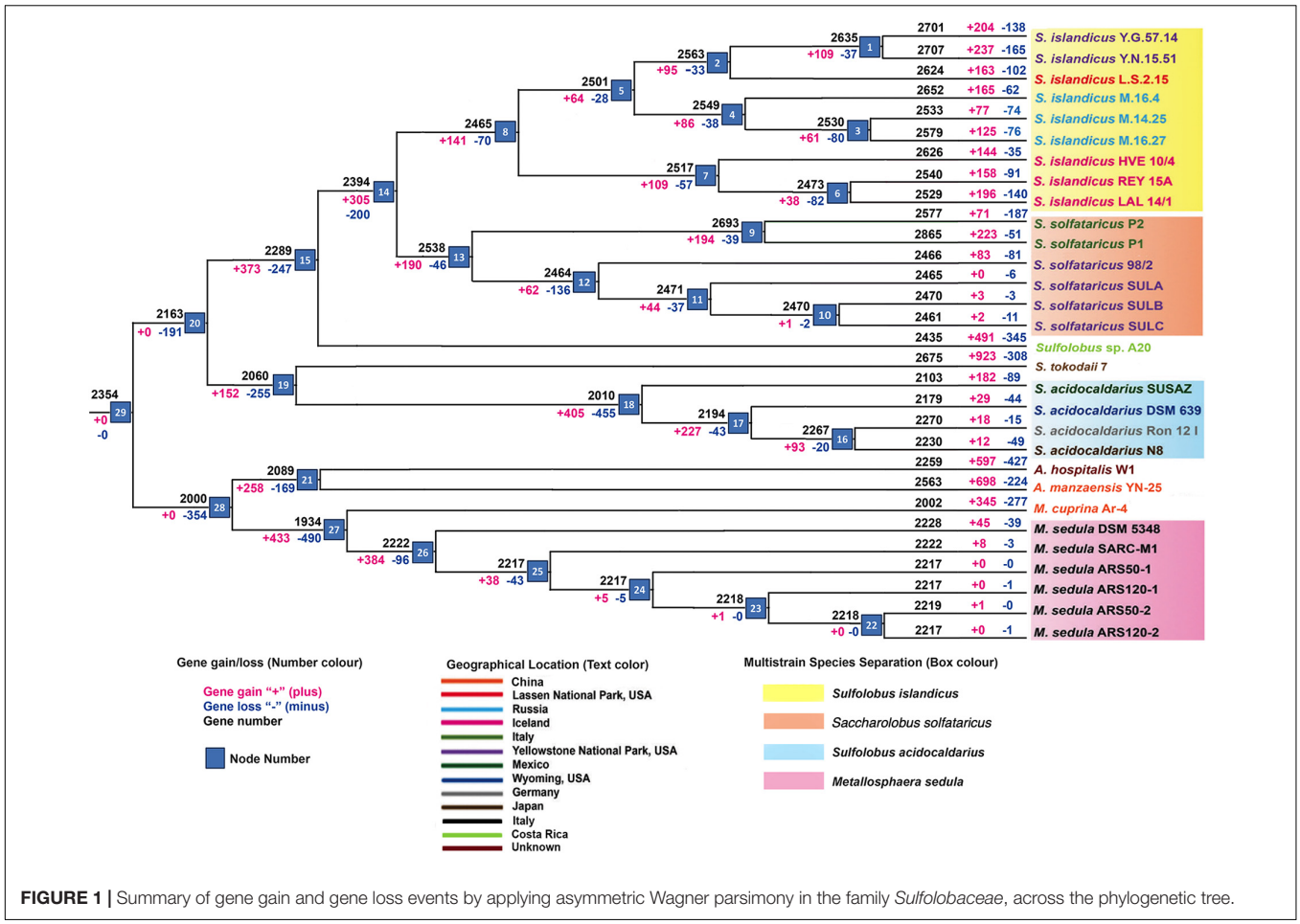


FIGURE 1 | Summary of gene gain and gene loss events by applying asymmetric Wagner parsimony in the family *Sulfolobaceae*, across the phylogenetic tree.

corroborated well with the genome size of respective species (average genome sizes are 2.65 ± 0.12, 2.81 ± 0.16, 2.17 ± 0.08, and 2.14 ± 0.13 Mbs for the strains of *S. islandicus*, *S. solfataricus*, *S. acidocaldarius*, and *M. sedula*, respectively). We also identified geographical location-wise gene gains and losses for *S. islandicus* and *S. solfataricus*. In case of *S. islandicus*, population specific to all three major geographical locations experienced net gene gain, whereas subpopulations from Russia experienced less amount of net loss and YNP subpopulation within the United States experienced higher gene gains. Similar kinds of population- and subpopulation-level gene gain–loss dynamics for *S. islandicus* except for Russian subpopulation were reported in an earlier pan-genome–based work (Reno et al., 2009).

It is to be noted here that a similar kind of gene gain–loss pattern can also be seen when we used Sankoff's algorithm (**Supplementary Figure 2**). The initial phase of genome shrinkage and subsequent expansion along with dynamic gene gain–loss can also be seen.

The gene sharing pattern within the different members of the family *Sulfolobaceae* as depicted in **Figure 1** was found to be in complete corroboration with the gene sharing pattern of the multistrain species, viz., *S. islandicus*, *S. solfataricus*, *S. acidocaldarius*, and *M. sedula*, when the pan-genomes of these species were reconstructed separately (**Supplementary Figures 3–6**). For example, 2,158 genes are core in nature in *M. sedula*–specific pan-genome (**Supplementary Figure 6**), and this number is close to the gene number calculated for *M. sedula* taxa–specific node at gene gain–loss tree (node 26 in **Figure 1**). We can find a smaller number of gene gain–loss after that node for *M. sedula* genomes. In these four pan-genome analyses, the unique genes for different member of any pan-genomic group are also corroborating well with the net gene gain–loss of the respective organism in the terminal node (**Figure 1**).

Previous studies about the genomic evolution of Archaea, bacteria, and eukaryotes estimated rich ancestral gene content and a prevailing trend of genome streamlining in subsequent lineage adaptation and diversification (Hottes et al., 2013). Some of these models also showed a period of gene gains as they adapt to new environments followed by a longer period of genome streamlining as they diversified (Richards et al., 2014). Our data indicated that in case of *Sulfolobaceae* genome diversification, the trend of genome streamlining was followed by periods of differential gene gains, resulting in an overall genome expansion in some species, whereas there was reduction in others.

Based on a probabilistic model of gene gain and loss, Csuros and Miklos (2009) used the posterior probabilities for ancestral state reconstruction in Archaea, and described gene loss as

a major force shaping the archaeal genomes throughout their evolutionary history. Despite this inherent trend, they found that few species experienced overall gene gain overcoming the continuous loss, and *S. solfataricus* was among them, although to a lesser extent than other group of Archaea. Also, they estimated that both the ancestral state node of two genomes of the genus *Sulfolobus* and the terminal node of *S. acidocaldarius* experienced a higher extent of gene loss than gain, whereas the reverse was true for the terminal node of *S. solfataricus*. Thus, our results were similar in part to that study, depicting the general trend of asymmetric gene gains–losses in different species of the family *Sulfolobaceae*. The trend of either overall gene gain or loss was not specific to any particular genus but rather to a particular species or two very closely related species. Any population may successfully thrive into a new environment by inventing or gaining major physiological or metabolic ways. The genomes further diversify by the process of genome streamlining, whereas the advantageous gains or inventions may be retained by population to thrive better.

The process of genome streamlining happens as the genomes diversify by genetic drift given the isolated and disconnected nature of the population characterized by the extreme habitat they thrive in (Whitaker, 2003). Population under significant energy stress may also favor loss of dispensable functions and specialization (Valentine, 2007). Another possibility for genome streamlining is the population-size effects due to the physical isolation (Lynch, 2006), despite the slightly deleterious loss of function. Also, the gene gain events are found to be adaptive in most of the cases (Sela et al., 2016).

Functional Concurrence for Gene Gain–Loss Events and Metabolic Pathways Diversification

We next looked into the impact of gene gain and loss events on cellular functions by examining their representation in arCOGs and KEGG pathway categories from multistrain species, viz., *S. islandicus*, *S. solfataricus*, *S. acidocaldarius*, and *M. sedula*. As these species contain multiple strains, more robust gene gain–loss estimation is possible at the respective species node. Interestingly, the representation from arCOG functional category "Metabolism" was found to be significantly higher ($P < 0.001$) in loss events than the gain ones for all four species. We found significantly higher representation in gene loss than gain for the "Cellular Processes and Signaling" category only for *S. acidocaldarius* (**Figure 2A**). Further analysis on the KEGG pathway enzymes also followed a similar trend as in the case of arCOG analysis (**Supplementary Figure 7**). The functional concurrence in the loss events of "Metabolism"-related genes was further looked in more detail. We found that there were mostly replacements of genes for the categories "Carbohydrate metabolism," "Energy metabolism," "Amino acid metabolism," and "Metabolism of cofactors and vitamins" for all these four species (**Figure 2B**). Moreover, within these pathway categories, *S. acidocaldarius* showed significantly higher gene loss for "Carbohydrate metabolism" and "Energy metabolism," whereas there was significantly higher gene gain for "Metabolism of cofactors and vitamins."

The metabolic potential of different *Sulfolobaceae* genomes differs considerably. From their metabolic preferences, it was reported that *Sulfolobus/Saccharolobus* species are mixotrophs growing chemolithoautotrophically or heterotrophically (Keeling et al., 1998). With chemolithoautotrophic metabolic features, these genomes can acquire their necessary carbon from environmental carbon dioxide, as well as they can also use inorganic compounds such as nitrogen and sulfur as their source of energy (Tang et al., 2011). On the other hand, because of their heterotrophic metabolic feature, *Sulfolobus/Saccharolobus* genomes can obtain their energy from organic compounds (Stark et al., 2017). The use of substrates as carbon source varies in *S. solfataricus* and *S. acidocaldarius* strains; the first one can use a wide range of compounds as carbon source including polysaccharides, disaccharides, hexoses, pentoses, aldehydes, alcohols, sugar acids, amino acids, peptides, and tryptone, whereas the latter have a comparatively restricted range of substrate preferences than *S. solfataricus* (Dai et al., 2016; Stark et al., 2017). Another genus *Metallosphaera* contains genomes that are obligate aerobes (Wheaton et al., 2015). The members of this genus are facultative chemolithoautotrophs following autotrophic carbon dioxide fixation pathways with a wide range of inorganic substrate choices. *M. cuprina* is the only reported genome from this genus, which is heterotrophic and shows significant substrate preference for their growth (Wheaton et al., 2015). The genus *Acidianus*, whose members are facultative anaerobes, are also capable of chemolithoautotrophy and can reduce and oxidize elemental sulfur, depending on oxygen availability (Wheaton et al., 2015).

This metabolic dynamicity of different constituent species of family *Sulfolobaceae* may be attributed toward the acquisition or loss of genes mostly from different metabolic pathways in course of their genome divergence. In order to understand in more detail about the repertoire of metabolic pathway genes and their relation with genome divergence, we carried out detailed analysis about the three major metabolic pathways in *Sulfolobaceae*: CCM, sulfur metabolism, and nitrogen metabolism, by considering all the 30 organisms under study.

Central Carbohydrate Metabolism

Analyses of KEGG enzymes belonging to autotrophic CO_2 fixation pathways, i.e., reverse ribulose monophosphate pathway (RuMP), citric acid cycle (TCA) along with the glyoxylate shunt, 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle, dicarboxylate/4-hydroxybutyrate cycle (DC/HB), C3/C4 conversions, as well as the heterotrophic pathways such as Embden–Meyerhof–Parnas (EMP) pathway, branched Entner–Doudoroff (ED) pathway (both non-phosphorylative and semiphosphorylative mode), and glycogen and trehalose metabolism for *Sulfolobaceae* genomes, were performed to understand the impact of gene gain–loss on each of these pathways.

Almost all the enzymes in the reverse RuMP pathway were found to be the core in *Sulfolobaceae* family. Only exception was the ribose 5-phosphate isomerase A (RPI, K01807), which

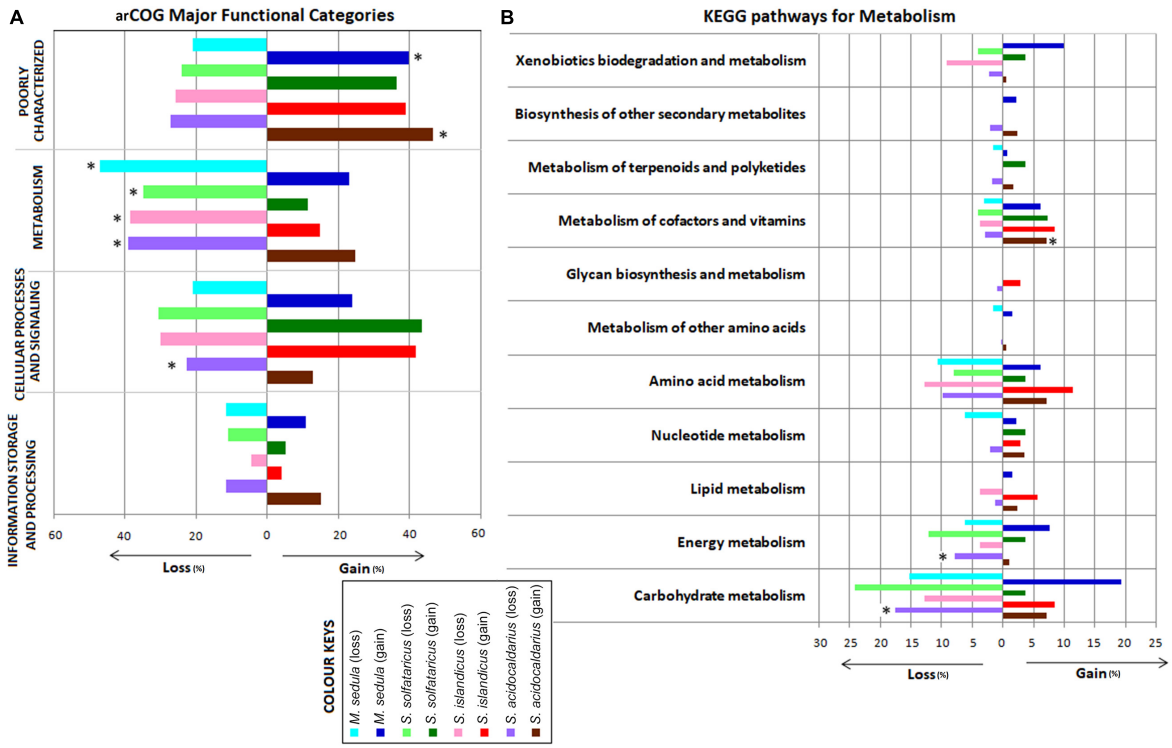


FIGURE 2 | Bar chart showing distribution of major functional categories of Archaeal Clusters of Orthologous Genes (arCOGs) at the species-level gene gain–loss **(A)** and distribution of the KEGG pathways for metabolism at the species-level gene gain–loss **(B)** of the family *Sulfolobaceae*. *Significant difference ($P < 0.05$) in that particular gain–loss event.

helps in the formation of ribose 5-phosphate from ribulose 5-phosphate (Ru5P), which was found to be lost from *S. solfataricus* strains SULA, SULB, SULC, and *S. tokodaii*. In case of the oxidative TCA cycle, all the enzymes were core in nature except for *A. hospitalis*, where the alpha and gamma subunits of pyruvate ferredoxin oxidoreductase (porA, K00169 and porG, K00172) were lost. The functional importance of these subunits would be important for further studies. The enzyme isocitrate lyase (aceA, K01637) was lost from several genomes both at species and genus levels: *S. tokodaii* and *S. solfataricus* strains from Yellowstone National Park and both the genera *Acidianus* and *Metallosphaera*. The enzyme malate synthase (aceB, K01638) that acts on glyoxylate to produce malate and enters into TCA cycle was found to be specifically gained by *M. sedula* genomes. But we have found that *M. sedula* genomes appear to be incapable of producing glyoxylate because of the absence of enzyme aceA in them. It is prominent from the present study that the glyoxylate shunt pathway is not core for *Sulfolobaceae*.

The enzymes required to carry out both HP/HB and DC/HB cycles were core for *Sulfolobaceae*, pointing toward the presence of these autotrophic carbon fixation pathways in this family. In case of *M. sedula* strains, two extra subunits of fumarate hydratase (fumA, K01677 and fumB, K01678) were found to be gained probably during the process of their speciation (**Supplementary Table 3** and **Supplementary Figure 8**).

Again all the enzymes involved in C3/C4 conversions were found to be core in nature, except phosphoenolpyruvate carboxykinase (PCK, K01596), which was specifically lost from *A. hospitalis*. C3/C4 conversions were also found to be a common property for the members of family *Sulfolobaceae* (**Supplementary Table 3** and **Supplementary Figure 8**).

It is well known that the genomes present in the *Sulfolobus* genus can metabolize glucose and galactose by the non-phosphorylative variant of the ED pathway (Dai et al., 2016). It was also reported that in *S. solfataricus* the ED pathway is imperative for glucose and galactose metabolism (Lamble et al., 2005). Here we found that the non-phosphorylative option of the ED pathway was entirely present in the genera *Sulfolobus/Saccharolobus*, except *S. tokodaii*, and was truncated or incomplete for the genera *Acidianus* and *Metallosphaera*. The incompleteness of this pathway was due to the loss of the key enzyme glucose dehydrogenase/aldose 1-dehydrogenase (ssgdh, K18125) by both genera *Acidianus* and *Metallosphaera* and species *S. tokodaii*. In addition, more gene losses were observed for genus *Acidianus* (KDG aldolase, kdpgA, K11395) and specifically for *A. hospitalis* (two subunits of glyceraldehyde dehydrogenase—cutA, K18020, and cutB, K18021). A semiphosphorylative ED pathway also was reported to be present in *S. solfataricus* as a substitute glucose metabolism pathway (Ahmed et al., 2005). As both the steps involving the enzymes ssgdh and kdpgA are common in non-phosphorylative and semiphosphorylative ED pathway, the semiphosphorylative ED pathway was also incomplete for the genera *Acidianus* and *Metallosphaera* and *S. tokodaii*. Moreover, enzymes, which

are exclusive for semiphosphorylative ED pathway such as KDG kinase (sskdgK, K18126) and glyceraldehyde-3-phosphate dehydrogenase (gapN, K18978), were lost from the genus *Acidianus*.

We further found complete EMP pathway for gluconeogenesis to be present in family *Sulfolobaceae*, except the interconversion of fructose 6-phosphate and glucose 6-phosphate. The enzyme glucose-6-phosphate isomerase (PGI, K15916) involved in the aforementioned conversion showed gains in species *S. islandicus*, *S. solfataricus*, and *Sulfolobus* sp. A20 and genus *Acidianus* and subsequent loss by few members of species *S. islandicus* and *S. solfataricus*. Glucose 1-phosphate formed in EMP pathway is converted to glycogen by starch synthase (glgA, K00703) that was found to be present for glycogen and trehalose metabolism in all genomes. This glycogen is again reversibly converted to glucose 1-phosphate by the enzyme glycogen phosphorylase (PYG, K00688) that was found to be lost from strain *S. islandicus* Y.N.15.51; species *S. acidocaldarius* and *Sulfolobus* sp. A20 and genera *Acidianus* and *Metallosphaera* signifying the absence of the glycogen to glucose 1-phosphate conversion for these genomes. Glycogen is also converted to glucose by glucoamylase (SGA1, K01178), a core enzyme for *Sulfolobaceae*. Glycogen, produced as a carbon storage compound (Quehenberger et al., 2017), is also a source for trehalose derived from the TreY/TreZ pathway in *Sulfolobus* (Maruta et al., 1996). Our study found that the enzyme (1–>4)-alpha-D-glucan 1-alpha-D-glucosylmutase (TreY, K06044) was gained by the genus *Acidianus*, signifying trehalose formation via the TreY pathway in them. On the other hand, the enzyme maltooligosyltrehalose trehalohydrolase (TreZ, K01236) was lost from *M. sedula*; thus, they did not have any of the TreY and TreZ pathways. Another unique trehalose synthesizing pathway is the unidirectional TreT pathway found in the hyperthermophilic phylum *Crenarchaeota*. We found trehalose synthase (TreT, K13057) to be the core pathway for *Sulfolobaceae* (**Supplementary Table 3** and **Supplementary Figure 8**).

In conclusion, the enzymes involved in the four main autotrophic CO_2 fixation pathways were almost core in nature for the family *Sulfolobaceae*. Furthermore, the TCA cycle, except the glyoxylate shunt, was also core for them. But the enzymes involved in the heterotrophic cycles were mostly lost from the genera *Acidianus* and *Metallosphaera* and the strain *S. tokodaii*.

Further, we looked into the chemolithotrophic metabolic pathways for *Sulfolobaceae* genomes, where they use inorganic compounds such as sulfur and nitrogen as the source of energy.

Sulfur Metabolism

In order to get an insight into the relationship of sulfur metabolism capability and evolutionary divergence of *Sulfolobaceae*, hierarchical clustering based on both abundance and presence/absence of sulfur metabolic pathways genes was performed (**Supplementary Table 4** and **Supplementary Figure 9**). Three enzymes, viz., cysteine synthase (cysK, K01738), cystathionine gamma-synthase (metB, K01739), and sulfide quinone oxidoreductase (SQR, K17218), were present in all 30 *Sulfolobaceae* genomes with multiple copies in some cases. SQR can catalyze the oxidation of hydrogen sulfide into polysulfide (Brito et al., 2009), whereas cysK and metB help to produce cysteine from sulfide, which is further used for cysteine and methionine metabolism (Ravanel et al., 1998). The enzymes sulfite reductase (NADPH) hemoprotein beta-component (cysI, K00381), phosphoadenosine phosphosulfate reductase (cysH, K00390), sulfite dehydrogenase (soeB, K21308), sulfate adenylyltransferase (sat, K00958), thiosulfate/3-mercaptopyruvate sulfurtransferase (TST, K01011) were found to be present in most of the *Sulfolobaceae* genomes. cysI, cysH, and sat enzymes were lost by *A. hospitalis* and *M. cuprina*, whereas soeB enzyme was lost by *S. solfataricus* P1 and *A. manzaensis*. cysH was also lost from *Sulfolobus* sp. A20, and TST was lost from the genus *Acidianus*. Multiple copies of cysH were found for *S. acidocaldarius* strains N8, DSM 639, and Ron12/I and *S. islandicus* strains REY15A and L.S.2.15. In addition, multiple copies of TST were present for *M. sedula* strains and most of the *Sulfolobus/Saccharolobus* genomes, except *S. acidocaldarius*. All these aforementioned five enzymes help in the conversion of hydrogen sulfide into sulfite with the successive transformation of the sulfite into sulfate producing ATP by phosphorylation (Dai et al., 2016). Gene encoding sulfate adenylyltransferase subunit 1 (cysN, K00956) was gained solely in *A. hospitalis*. Another important enzyme sulfur oxygenase/reductase (SOR, K16952) for sulfur oxidation in Archaea was lost by all members of *Sulfolobus/Saccharolobus*, except *S. tokodaii* and genus *Metallosphaera* (Dai et al., 2016). Enzyme thiosulfate reductase (phsA, K08352) was gained only by genus *Acidianus*. Multiple copies of doxA (K16936) were present in *S. solfataricus* strains, whereas multiple copies of doxD (K16937) were found in almost all members of *Sulfolobaceae*, except *S. acidocaldarius* SUSAZ and *Sulfolobus* sp. A20. doxDA enzymes are homologs of the TQO (thiosulfate: quinone oxidoreductase), which plays a key role in the transformation of thiosulfate into tetrathionate (Liu et al., 2012). Subunits of the enzyme for sulfur reductase, viz., sulfur reductase molybdopterin subunit (sreA, K17219), sulfur reductase FeS subunit (sreB, K17220), and sulfur reductase membrane anchor (sreC, K17221), are having a significant part in the formation of hydrogen sulfide via reduction of elemental sulfur. These enzymes were primarily gained by *S. islandicus*, *S. solfataricus*, and *A. manzaensis*. Later sreA and sreC were lost from *S. solfataricus* strains isolated from Yellowstone National Park. The only enzyme that helps in the sulfite formation from methane–sulfonate is toluene monooxygenase system ferredoxin subunit (tmoC, K15762) that was found to be lost from the genus *Metallosphaera* and species *A. hospitalis*. Among the *Sulfolobus* genomes, tmoC was lost from all *S. acidocaldarius* strains, *S. tokodaii*, *S. islandicus* M.14.25, *S. islandicus* M.16.27, *S. islandicus* Y.G.57.14, and *S. islandicus* LAL14/1 (**Supplementary Table 4** and **Supplementary Figure 9**).

Nitrogen Metabolism

In case of nitrogen metabolic pathways enzymes, two subunits of carbamoyl-phosphate synthase (carB, K01955 and carA, K01956) were found in all the 30 *Sulfolobaceae* genomes (**Supplementary Table 5** and **Supplementary Figure 10**). Two other enzymes glutamate dehydrogenase (gdhA, K00261) and glutamine synthetase (glnA, K01915) had multiple copies in some

of the genomes. The above results point toward the fact that all the genomes of *Sulfolobaceae* had a common tendency of using ammonia instead of nitrate, urea, cyanate, or formamide, as a source of nitrogen for the glutamate, glutamine, and carbamoyl-phosphate synthesis.

The abundance-based cluster analysis showed that two *S. islandicus* strains each from Iceland (REY15A and LAL14/1) and Russia (M.14.25 and M.16.27) were clubbing together. These strains specifically gained four subunits of nitrate reductase, i.e., narG (alpha subunit, K00370), narH (beta subunit, K00371), narJ (delta subunit, K00373), and narI (gamma subunit, K00374), and thus, they may be capable of nitrate utilization. To be mentioned here, the nitrate/nitrite transporter (nark, K02575) enzyme, primarily present in all *Sulfolobaceae* genomes, was lost from genera *Sulfolobus/Saccharolobus*, species *A. hospitalis*, and *M. cuprina* and then independently gained by the aforementioned four *S. islandicus* strains. The enzyme formamidase (K01455) was only gained by *S. islandicus* HVE10/4, and cyanate lyase (cynS, K01725) was gained by *A. manzaensis* and *S. tokodaii*, suggesting a much wider range of nitrogen sources, such as cyanide and cyanate, for these genomes (**Supplementary Table 5** and **Supplementary Figure 10**).

The detailed analysis of gene gain–loss patterns among the three major metabolic pathways of *Sulfolobaceae* genomes showed that CCM and sulfur metabolism potential almost followed their core/pan-genome diversification pattern. As in case of their genome divergence, within any genus, these two metabolic pathway profiles varied considerably. The repertoires of CCM pathway enzymes were quite distinct for two species of genus *Acidianus*, depicting the impact of gene acquisition and loss in course of their genome divergence. Also, the effect of gene losses was more pronounced in these two pathway genes for all the strains of the species *S. acidocaldarius*, *S. tokodaii*, and *A. hospitalis*. On the other hand, genes involved in nitrogen metabolism were diversely distributed in *Sulfolobaceae* genomes, indicating several independent gene acquisition and loss events irrespective of genome diversification. In case of *S. solfataricus* and *S. acidocaldarius*, this pathway genes seemed to be more conserved as evidenced by their repertoire in isolates from various geographical regions, whereas this was not the scenario for *S. islandicus* genomes and genus *Acidianus*. Thus, our analyses suggested that the ability to use carbohydrate and inorganic sulfur in the members of the family *Sulfolobaceae* has coevolved with their genome diversification but not for nitrogen metabolism.

Core Genome Diversity and Differential Selective Pressure

Apart from the gene gain–loss events and metabolic pathway differentiation, we looked into the evolutionary divergence of core genes by detecting the evidence of potential recombination events and nucleotide diversity. We identified 30 core genes (6% of the total core genes) with recombination signal, along with putative major parents, minor parents, and recombinants in each of these genes. Core genome phylogenetic tree with non-recombinant core genes (489 genes) was found to be congruent (K tree score = 0.0015) (Soria-Carrasco et al., 2007) with that of previous core genome tree, implying undetectable

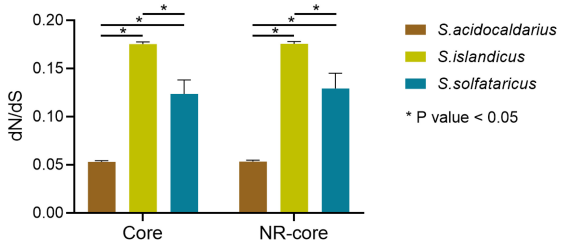


FIGURE 3 | Average dN/dS values for the core genes and non-recombinant core genes belonging to the species with multiple strains (*S. islandicus*, *S. solfataricus*, and *S. acidocaldarius*).

impact of recombination events on core genome divergence as a whole. As mobile genetic elements influence HGT as well as transposition, it is apparent that less susceptibility to the mobile genetic elements for *S. acidocaldarius* strains in comparison to *S. islandicus* and *S. solfataricus* strains resulted in higher preservation in *S. acidocaldarius* genomes (Brügger et al., 2002; Chen et al., 2005; Redder and Garrett, 2006; Quehenberger et al., 2017; Wagner et al., 2017). However, previous works demonstrated that to be deficient in mobile genetic elements does not result in the low genomic diversity in *S. acidocaldarius* strains. Instead, they proposed that significant disparity in the intrinsic evolutionary rates in the *Sulfolobus/Saccharolobus* genomes produced comparatively low diversity in the *S. acidocaldarius* strains (Anderson et al., 2017).

We further analyzed dN and dS values for multistrain species (*S. islandicus*, *S. solfataricus*, *S. acidocaldarius*, and *M. sedula*) for both 519 core genes and 489 non-recombinant core genes. The values of dN, dS, and dN/dS for *S. acidocaldarius* strains were significantly ($P < 0.05$) lower than that for *S. islandicus* and *S. solfataricus* showing their highly conserved nature (**Figure 3** and **Supplementary Table 6**). Similar trend was observed in non-recombinant core genes. We further incorporated five recently available complete genome sequences of *S. acidocaldarius* (NCBI reference Id: CP020360.1, CP020361.1, CP020362.1, CP020363.1, and CP020364.1) and identified the orthologs of 519 core genes from them and again calculated the dN/dS ratio. For these nine strains of *S. acidocaldarius*, the average dN/dS value of 519 core genes was found to be 0.053, which is again significantly ($P < 0.05$) lower than that of *S. islandicus* and *S. solfataricus*. The selection efficiency can plausibly be quantified by the core genome-wide dN/dS ratio, where a lower dN/dS ratio indicates a stronger purifying selection on an average. Thus, it can be inferred that a strong purifying selection was associated with *S. acidocaldarius* strains from different geographical locations. On the other hand, we found no changes at the nucleotide level of the core genes of the six *M. sedula* strains under study, signifying their absolute resemblance at the core-gene level probably due to their laboratory derived origin (Ai et al., 2015, 2016).

A previous study also reported lower nucleotide divergence in *S. acidocaldarius* genes than in *S. islandicus* (0.15% of *S. islandicus* value) (Mao and Grogan, 2017). Also, the study conducted by Anderson et al. (2017) identified higher conserved core genes from 47 *S. acidocaldarius* genomes (Anderson et al., 2017). This disparity in the extent of the conservation of core

genes for *S. acidocaldarius* was quite intriguing. One possible reason could be that the removal of any excess variation might have happened through very recent genetic bottleneck or selective sweep. However, an alternative possibility can be that any excess diversity in response to strongly positive selection pressures was not necessary for the maintenance of lifestyle, survival, and adaptation of *S. acidocaldarius* as a species. In essence, such a high level of conservation was needed owing to strong purifying selection as suggested in this study. Another hypothesis suggested that a more recent purifying event has taken place in the *S. acidocaldarius* population, or their emergence is a recent phenomenon than the *S. islandicus* populations (Mao and Grogan, 2017). Interestingly, in support of our findings, a recent study reported that smaller archaeal genomes experienced stronger purifying selection, as opposed to the common observations of positive selection dynamics in bacterial genomes (Lyu et al., 2017). In another study, epipelagic *Marinimicrobia* has been found to show properties consistent with the hypothesis regarding genome streamlining; i.e., small, compact genomes are correlated with stronger purifying selection pressure (Aylward and Martinez-Gutierrez, 2019). Importantly, within the genus *Sulfolobus*, *S. acidocaldarius* represented the smallest average genome size. This scenario can be explained by the existing theory suggesting strong purifying selection to be operative on prokaryotic genomes characterized by large effective population size so as to retain their compact genomes (Lynch, 2006).

Lineage and Functional Category–Specific Positive Selection

The branch-site codon model for testing positive selection determined 22 gene sets showing significant ($P < 0.05$; LRT test) adaptive evolution for different strains (**Supplementary Tables 7, 8A,B**). These 22 genes included 15 genes for *S. islandicus* strains, five genes for both the *S. islandicus* and *S. solfataricus* strains, one gene for both *S. solfataricus* and *S. acidocaldarius* strains, and one gene for *S. acidocaldarius* strains only (**Supplementary Tables 8A,B**). We then looked into the recombination events for these genes and found that three of them were affected by homologous recombination (**Supplementary Table 8A**). Interestingly, for all these gene sets, the positive selection signal was present specifically for *S. islandicus* strains. We found the *S. islandicus* strain (marked by "a" in **Supplementary Table 8A**) to appear as the major parent for one of the recombination events. This result implied that the changes within the positively selected gene pairs might be distributed via the process of homologous recombination.

All these genes were found to have defined functions and further GO enrichment analysis using Comparative GO web server[3] with *S. islandicus* strain L.S.2.15 as reference yielded significant ($P < 0.05$) enrichment of three molecular functional categories, viz., metal ion binding (GO id: 46872), ATP binding (GO id: 5524), and zinc ion binding (GO id: 8270). Of 22 genes, nine genes were from these three categories. Thus, the positively selected genes were not from random functional categories but were targeted toward particular functions. Among the functional categories of positively selected genes, superoxide dismutase is an up-regulated gene under oxygen-rich conditions in *S. solfataricus* (Sheng et al., 2014), helping the cells resist fatal oxidative stress (Culotta, 2001). Our study revealed that the positively selected changes in the gene pairs representing superoxide dismutase were distributed via the process of homologous recombination.

A comparative genomics study by Zaparty et al. (2010) reconstructed the CCM network where enzymes ribokinase, succinate dehydrogenase, RNA-binding protein, ATP-binding protein, AIR synthase, and different transcription regulating factors were found to be associated with that pathway in *S. solfataricus* (Zaparty et al., 2010). Notably, the present study identified some of these genes to be under positive selection pressures. The gene encoding ribokinase signifies the conversion of ribose 5-phosphate to ribose, a very important enzyme for autotrophic CO_2 fixation pathway, under positive selection for *S. islandicus* strains. To be noted here, another enzyme succinate semialdehyde reductase (K14465), involved in both HP/HB and DC/HB cycles, was also found to be under positive selection pressure for both *S. islandicus* and *S. solfataricus* strains. Again, the enzyme biotin carboxylase (K01964) was under positive selection pressure and found to be involved in both HP/HB cycle and C3/C4 conversions in these two species. All these three genes were from autotrophic CO_2 fixation pathways, thus signifying the functional importance of these positively selected genes in autotrophic pathways for the species *S. islandicus* and *S. solfataricus*. In addition, cystathionine gamma-synthase, which helps in the catalysis of the cleavage of cystathionine to generate cysteine, has been found to be under positive selection pressure for *S. islandicus* strains. In Archaea, the major source of sulfur is cysteine for the synthesis of sulfur-containing compounds (Borup and Ferry, 2000), and thus, the enzyme cystathionine gamma-synthase present within the sulfur metabolic pathway may have huge functional importance for the species *S. islandicus*.

In the study of microbial evolution, often the primary emphasis is given to the genome-specific genes acquired by HGT events, whereas the adaptive variations in the core genes get overlooked. Our analysis identified that in the course of genome diversification of the family *Sulfolobaceae*, a considerable amount of core genes representing specific protein functional categories evolved under positive selection pressures in certain species in an asymmetric way. Primarily, the core genes are expected to be conserved and to evolve under strong purifying selection pressures because of the housekeeping roles the respective encoded proteins play in metabolic pathways and physiological activities. Thus, this result is quite interesting, and in the recent past, it was found that the positive selection pressure on core genes was an important mechanism for adaptive evolution of several pathogenic bacteria (Chattopadhyay et al., 2012b). In our analysis, the enrichment of three important molecular functional categories (metal ion binding, ATP binding, and zinc ion binding), as well as the autotrophic CO_2 fixation pathway enzymes in positively selected genes, further signified the adaptive role of this selection pressure in Archaea. It is to be mentioned here that we could not detect any genes

[3]http://www.comparativego.com/

in genera *Acidianus* and *Metallosphaera* to be under positive selection pressure.

This result indicates that for *S. islandicus*, the maximum number of core genes was evolving under positive selection. This positive selection might have played an important role in the strain-level allopatric speciation and adaptation in *S. islandicus* in different geographical locations. Also, several core genes were under the influence of positive selection for both *S. islandicus* and *S. solfataricus* or their common ancestor, which again experienced a considerable amount of gene gains. Thus, both of these species might be under strong selection pressure in terms of acquiring new genes, as well as new sequence feature in their core genome components.

CONCLUSION

Comparative pan-genomics along with concomitant evolutionary analyses of 30 genomes of the family *Sulfolobaceae* revealed asymmetric genome evolution. Our analyses on various evolutionary mechanisms responsible for adaptive genome diversification pointed toward the asymmetric nature of these mechanisms shaping the member genomes of this family. This study enabled us to understand the sources of genome-level variations that exist in both species and genus levels within the family *Sulfolobaceae*. Pan-genomics coupled with gene gain–loss analysis explored the inherent pattern of genome streamlining, followed by waves of differential gene gains, which resulted as genome expansion in some species while reduction in others in comparison to the ancestral state. The considerable amount of genome size variation of the members of any genus may be attributable to these waves of gene acquisitions. At the species level, substantial gene gain was evident for *S. islandicus* and *S. solfataricus*, whereas there was loss for *S. acidocaldarius.* Analysis of these gene gain–loss patterns among the three major metabolic pathways of *Sulfolobaceae* genomes revealed that the CCM and sulfur metabolism potential of its members coevolved with the genome diversification pattern of this archaeal family. The autotrophic CO_2 fixation pathways, except for the glyoxylate shunt, were nearly conserved in the family *Sulfolobaceae*. But the heterotrophic cycles were mostly lost from both the genera *Acidianus* and *Metallosphaera* and the strain *S. tokodaii* 7. Therefore, the CCM pathway enzymes were relatively different for *Acidianus*, *Metallosphaera*, and *S. tokodaii*, whereas the enzymes involved in sulfur metabolism were essentially lost from the genus *Acidianus*, thereby suggesting the impact of gene gain–loss pattern in course of genome divergence of *Sulfolobaceae*.

The strong purifying selection operating on the relatively reduced genomes of *S. acidocaldarius* conformed to a much higher level of sequence conservation in this species than in others. A considerable amount of core genes of *Sulfolobaceae* was found to evolve under positive selection pressures in a lineage- and functional category-specific manner, as evidenced predominantly in the adaptive evolution of *S. islandicus*. While thriving in the similar niche, the lineages can experience contrasting evolutionary routes leading to differential genome adaptation trail as indicated in our study. The enrichment of important functional categories, viz., metal ion binding, ATP binding, and zinc ion binding in positively selected genes, as well as important CO_2 fixation pathway genes, further signified the adaptive role of these genes. The genera *Sulfolobus/Saccharolobus* has potential in biotechnological application due to their wide substrate specificity, less contamination possibility, thriving in extreme conditions, and no need for cooling arrangements (Abdel-Banat et al., 2010). Also, the extreme growth conditions of this genus characterized by high temperature and low pH are potentially beneficial for several industries by increasing the substrate solubility (Quehenberger et al., 2017). In this regard, it is important to further assess these positively selected genes for their functional significance by laboratory experiments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

RB conceptualized the project, performed the analysis, generated and organized the results. NMC, AL, AG, and DB performed required programming during analysis and data generation. CD and SC added thoughtful suggestions during the work and made critical reading of the manuscript. DHH performed the analysis of gene gain–loss events. SP conceived and coordinated the project and performed data analysis and interpretation. SP and RB wrote the manuscript with inputs from other authors. All the authors read and approved the final manuscript.

FUNDING

Financial support for RB from the DBT-RA Program in Biotechnology and Life Sciences is gratefully acknowledged. NMC and AL were supported by CSIR-Senior Research Fellowship. AG was supported by Research Fellowship from National Institute of Pharmaceutical Education and Research (NIPER), Kolkata and Ministry of Chemicals and Fertilizers, Department of Pharmaceuticals, Government of India. DB was supported by ICMR-Senior Research Fellowship program. SP was supported by Ramanujan Fellowship of Science and Engineering Research Board (SERB), Government of India.

ACKNOWLEDGMENTS

We thank Gagneet Kaur for her initial technical help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.639995/full#supplementary-material

REFERENCES

Abdel-Banat, B. M. A., Hoshida, H., Ano, A., Nonklang, S., and Akada, R. (2010). High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast? *Appl. Microbiol. Biotechnol.* 85, 861–867. doi: 10.1007/s00253-009-2248-5

Ahmed, H., Ettema, T. J. G., Tjaden, B., Geerling, A. C. M., van der Oost, J., and Siebers, B. (2005). The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* 390, 529–540. doi: 10.1042/BJ20041711

Ai, C., McCarthy, S., Schackwitz, W., Martin, J., Lipzen, A., and Blum, P. (2015). Complete genome sequences of evolved arsenate-resistant *Metallosphaera sedula* strains. *Genome Announc.* 3:e01142-15. doi: 10.1128/genomeA.01142-15

Ai, C., McCarthy, S., Eckrich, V., Rudrappa, D., Qiu, G., and Blum, P. (2016). Increased acid resistance of the archaeon, *Metallosphaera sedula* by adaptive laboratory evolution. *J. Ind. Microbiol. Biotechnol.* 43, 1455–1465. doi: 10.1007/s10295-016-1812-0

Andam, C. P., Doroghazi, J. R., Campbell, A. N., Kelly, P. J., Choudoir, M. J., and Buckley, D. H. (2016). A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio* 7:e02200-15. doi: 10.1128/mBio.02200-15

Anderson, R. E., Kouris, A., Seward, C. H., Campbell, K. M., and Whitaker, R. J. (2017). Structured populations of *Sulfolobus acidocaldarius* with susceptibility to mobile genetic elements. *Genome Biol. Evol.* 9, 1699–1710. doi: 10.1093/gbe/evx104

Aylward, F. O., and Martinez-Gutierrez, C. A. (2019). Strong purifying selection contributes to genome streamlining in epipelagic *Marinimicrobia*. *bioRxiv* [Preprint]. doi: 10.1101/653261 653261,

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

Benvegnu, T., Lemiègre, L., and Cammas-Marion, S. (2009). New generation of liposomes called archaeosomes based on natural or synthetic archaeal lipids as innovative formulations for drug delivery. *Recent Pat. Drug Deliv. Formul.* 3, 206–220. doi: 10.2174/187221109789105630

Besse, A., Peduzzi, J., Rebuffat, S., and Carre-Mlouka, A. (2015). Antimicrobial peptides and proteins in the face of extremes: lessons from archaeocins. *Biochimie* 118, 344–355. doi: 10.1016/j.biochi.2015.06.004

Borup, B., and Ferry, J. G. (2000). Cysteine biosynthesis in the Archaea: *Methanosarcina thermophila* utilizes O -acetylserine sulfhydrylase. *FEMS Microbiol. Lett.* 189, 205–210. doi: 10.1111/j.1574-6968.2000.tb09231.x

Bräsen, C., Esser, D., Rauch, B., and Siebers, B. (2014). Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* 78, 89–175. doi: 10.1128/MMBR.00041-13

Brito, J. A., Sousa, F. L., Stelter, M., Bandeiras, T. M., Vonrhein, C., Teixeira, M., et al. (2009). Structural and functional insights into sulfide: quinone oxidoreductase. *Biochemistry* 48, 5613–5622. doi: 10.1021/bi9003827

Brockhurst, M. A., Harrison, E., Hall, J. P., Richards, T., McNally, A., and MacLean, C. (2019). The ecology and evolution of pangenomes. *Curr. Biol.* 29, R1094–R1103. doi: 10.1016/j.cub.2019.08.012

Brügger, K., Redder, P., She, Q., Confalonieri, F., Zivanovic, Y., and Garrett, R. A. (2002). Mobile elements in archaeal genomes. *FEMS Microbiol. Lett.* 206, 131–141. doi: 10.1111/j.1574-6968.2002.tb10999.x

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Chattopadhyay, S., Paul, S., Kisiela, D. I., Linardopoulou, E. V., and Sokurenko, E. V. (2012a). Convergent molecular evolution of genomic cores in *Salmonella enterica* and *Escherichia coli*. *J. Bacteriol.* 194, 5002–5011. doi: 10.1128/JB.00552-12

Chattopadhyay, S., Tchesnokova, V., McVeigh, A., Kisiela, D. I., Dori, K., Navarro, A., et al. (2012b). Adaptive evolution of class 5 fimbrial genes in enterotoxigenic *Escherichia coli* and its functional consequences. *J. Biol. Chem.* 287, 6150–6158. doi: 10.1074/jbc.M111.303735

Chen, L., Brügger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., et al. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. *J. Bacteriol.* 187, 4992–4999. doi: 10.1128/JB.187.14.4992-4999.2005

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

Csûös, M. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26, 1910–1912. doi: 10.1093/bioinformatics/btq315

Csuros, M., and Miklos, I. (2009). Streamlining and large ancestral genomes in archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* 26, 2087–2095. doi: 10.1093/molbev/msp123

Culotta, V. C. (2001). Superoxide dismutase, oxidative stress, and cell metabolism. *Curr. Top. Cell. Regul.* 36, 117–132. doi: 10.1016/s0070-2137(01)80005-4

Dai, X., Wang, H., Zhang, Z., Li, K., Zhang, X., Mora-López, M., et al. (2016). Genome sequencing of *Sulfolobus* sp. A20 from Costa Rica and comparative analyses of the putative pathways of carbon, nitrogen, and sulfur metabolism in various *Sulfolobus* Strains. *Front. Microbiol.* 7:1902. doi: 10.3389/fmicb.2016.01902

Das, S., Paul, S., Bag, S. K., and Dutta, C. (2006). Analysis of *Nanoarchaeum equitans* genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. *BMC Genomics* 7:186. doi: 10.1186/1471-2164-7-186

Dutta, C., and Paul, S. (2012). Microbial lifestyle and genome signatures. *Curr. Genom.* 13, 153–162. doi: 10.2174/138920212799860698

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Freschi, L., Vincent, A. T., Jeukens, J., Emond-Rheault, J. G., Kukavica-Ibrulj, I., Dupont, M. J., et al. (2019). The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol. Evol.* 11, 109–120. doi: 10.1093/gbe/evy259

Friedrich, C. G., Bardischewsky, F., Rother, D., Quentmeier, A., and Fischer, J. (2005). Prokaryotic sulfur oxidation. *Curr. Opin. Microbiol.* 8, 253–259. doi: 10.1016/j.mib.2005.04.005

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

Horikoshi, K., and Grant, W. D. (1998). *Extremophiles: Microbial life in Extreme Environments*. New York, NY: Wiley-Liss.

Hottes, A. K., Freddolino, P. L., Khare, A., Donnell, Z. N., Liu, J. C., and Tavazoie, S. (2013). Bacterial adaptation through loss of function. *PLoS Genet.* 9:e1003617. doi: 10.1371/journal.pgen.1003617

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., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Keeling, P. J., Klenk, H.-P., Singh, R. K., Schenk, M. E., Sensen, C. W., Zillig, W., et al. (1998). *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. *Extremophiles* 2, 391–393. doi: 10.1007/s007920050083

Kozubal, M., Macur, R. E., Korf, S., Taylor, W. P., Ackerman, G. G., Nagy, A., et al. (2008). Isolation and distribution of a novel iron-oxidizing crenarchaeon from acidic geothermal springs in Yellowstone National Park. *Appl. Environ. Microbiol.* 74, 942–949. doi: 10.1128/AEM.01200-07

Kück, P., and Meusemann, K. (2010). FASconCAT: convenient handling of data matrices. *Mol. Phylogenet. Evol.* 56, 1115–1118. doi: 10.1016/j.ympev.2010.04.024

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

Lamble, H. J., Theodossis, A., Milburn, C. C., Taylor, G. L., Bull, S. D., Hough, D. W., et al. (2005). Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon *Sulfolobus solfataricus*. *FEBS Lett.* 579, 6865–6869. doi: 10.1016/j.febslet.2005.11.028

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

Lefébure, T., and Stanhope, M. J. (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* 8:R71. doi: 10.1186/gb-2007-8-5-r71

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

Li, W., Fu, L., Niu, B., Wu, S., and Wooley, J. (2012). Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief. Bioinform.* 13, 656–668. doi: 10.1093/bib/bbs035

Littlechild, J. A. (2015). Enzymes from extreme environments and their industrial applications. *Front. Bioeng. Biotechnol.* 3:161. doi: 10.3389/fbioe.2015.00161

Liu, Y., Beer, L. L., and Whitman, W. B. (2012). Sulfur metabolism in archaea reveals novel processes. *Environ. Microbiol.* 14, 2632–2644. doi: 10.1111/j.1462-2920.2012.02783.x

Loewe, L., and Hill, W. G. (2010). The population genetics of mutations: good, bad and indifferent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1153–1167. doi: 10.1098/rstb.2009.0317

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60, 327–349. doi: 10.1146/annurev.micro.60.080805.142300

Lynn, D. J., Singer, G. A., and Hickey, D. A. (2002). Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Res.* 30, 4272–4277. doi: 10.1093/nar/gkf546

Lyu, Z., Li, Z.-G., He, F., and Zhang, Z. (2017). An important role for purifying selection in archaeal genome evolution. *mSystems* 2:e00112-17. doi: 10.1128/mSystems.00112-17

Makarova, K., Wolf, Y., and Koonin, E. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life* 5, 818–840. doi: 10.3390/life5010818

Mao, D., and Grogan, D. W. (2017). How a genetically stable extremophile evolves: modes of genome diversification in the Archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 199:e00177-17. doi: 10.1128/JB.00177-17

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

Martiny, J. B., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341

Maruta, K., Mitsuzumi, H., Nakada, T., Kubota, M., Chaen, H., Fukuda, S., et al. (1996). Cloning and sequencing of a cluster of genes encoding novel enzymes of trehalose biosynthesis from thermophilic archaebacterium *Sulfolobus acidocaldarius*. *Biochim. Biophys. Acta* 1291, 177–181. doi: 10.1016/s0304-4165(96)00082-7

McCarthy, S., Gradnigo, J., Johnson, T., Payne, S., Lipzen, A., Martin, J., et al. (2015). Complete genome sequence of *Sulfolobus solfataricus* strain 98/2 and evolved derivatives. *Genome Announc.* 3:e00549-15.

Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43, W566–W570. doi: 10.1093/nar/gkv468

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321

Nazari, F., Kefayati, M. E., and Raheb, J. (2017). The study of biological technologies for the removal of sulfur compounds. *J. Sci. I. R. Iran* 28, 205–219. doi: 10.1007/s11274-015-1915-1

Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.

Paul, S., Bag, S. K., Das, S., Harvill, E. T., and Dutta, C. (2008). Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol.* 9:R70. doi: 10.1186/gb-2008-9-4-r70

Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* 29, 170–175. doi: 10.1016/j.tig.2012.12.006

Quehenberger, J., Shen, L., Albers, S.-V., Siebers, B., and Spadiut, O. (2017). *Sulfolobus* – a potential key organism in future biotechnology. *Front. Microbiol.* 8:2474. doi: 10.3389/fmicb.2017.02474

Ravanel, S., Gakière, B., Job, D., and Douce, R. (1998). Cystathionine gamma-synthase from *Arabidopsis thaliana*: purification and biochemical characterization of the recombinant enzyme overexpressed in *Escherichia coli*. *Biochem. J.* 331, 639–648. doi: 10.1042/bj3310639

Redder, P., and Garrett, R. A. (2006). Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J. Bacteriol.* 188, 4198–4206. doi: 10.1128/JB.00061-06

Reed, C. J., Lewis, H., Trejo, E., Winston, V., and Evilia, C. (2013). Protein adaptations in archaeal extremophiles. *Archaea* 2013:373275. doi: 10.1155/2013/373275

Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V., and Whitaker, R. J. (2009). Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8605–8610. doi: 10.1073/pnas.0808945106

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

Rothschild, L. J., and Mancinelli, R. L. (2001). Life in extreme environments. *Nature* 409, 1092–1101. doi: 10.1038/35059215

Sakai, H. D., and Kurosawa, N. (2018). *Saccharolobus caldissimus* gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of *Sulfolobus solfataricus* as *Saccharolobus solfataricus* comb. nov. *Int. J. Syst. Evol. Microbiol.* 68, 1271–1278. doi: 10.1099/ijsem.0.002665

Sankoff, D. (1975). Minimal mutation trees of sequences. *SIAM J. Appl. Math.* 28, 35–42.

Seckbach, J. (2000). *Journey to Diverse Microbial Worlds: Adaptation to Exotic Environments*. Dordrecht: Springer.

Sela, I., Wolf, Y. I., and Koonin, E. V. (2016). Theory of prokaryotic genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 113, 11399–11407. doi: 10.1073/pnas.1614083113

Shapiro, B. J., Leducq, J. B., and Mallet, J. (2016). What is speciation? *PLoS Genet.* 12:e1005860. doi: 10.1371/journal.pgen.1005860

Sheng, Y., Abreu, I. A., Cabelli, D. E., Maroney, M. J., Miller, A.-F., Teixeira, M., et al. (2014). Superoxide dismutases and superoxide reductases. *Chem. Rev.* 114, 3854–3918. doi: 10.1021/cr4005296

Soria-Carrasco, V., Talavera, G., Igea, J., and Castresana, J. (2007). The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. *Bioinformatics* 23, 2954–2956. doi: 10.1093/bioinformatics/btm466

Stark, H., Wolf, J., Albersmeier, A., Pham, T. K., Hofmann, J. D., Siebers, B., et al. (2017). Oxidative stickland reactions in an obligate aerobic organism – amino acid catabolism in the Crenarchaeon *Sulfolobus solfataricus*. *FEBS J.* 284, 2078–2095. doi: 10.1111/febs.14105

Steinegger, M., and Söding, J. (2018). Clustering huge protein sequence sets in linear time. *Nat. Commun.* 9:2542. doi: 10.1038/s41467-018-04964-5

Tang, K.-H., Tang, Y. J., and Blankenship, R. E. (2011). Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. *Front. Microbiol.* 2:165. doi: 10.3389/fmicb.2011.00165

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi: 10.1016/j.mib.2008.09.006

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21:180. doi: 10.1186/s13059-020-02090-4

Valentine, D. L. (2007). Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* 5, 316–323. doi: 10.1038/nrmicro1619

van Wolferen, M., Ajon, M., Driessen, A. J., and Albers, S. V. (2013). How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles* 17, 545–563. doi: 10.1007/s00792-013-0552-6

Vishnu, U. S., Sankarasubramanian, J., Sridhar, J., Gunasekaran, P., and Rajendhran, J. (2015). Identification of recombination and positively selected genes in *Brucella*. *Indian J. Med. Microbiol.* 55, 384–391. doi: 10.1007/s12088-015-0545-5

Vos, M. (2011). A species concept for bacteria based on adaptive divergence. *Trends Microbiol.* 19, 1–7. doi: 10.1016/j.tim.2010.10.003

Wagner, A., Whitaker, R. J., Krause, D. J., Heilers, J.-H., Van Wolferen, M., Van Der Does, C., et al. (2017). Mechanisms of gene flow in archaea. *Nat. Rev. Microbiol.* 15, 492–501. doi: 10.1038/nrmicro.2017.41

Wheaton, G., Counts, J., Mukherjee, A., Kruh, J., and Kelly, R. (2015). The confluence of heavy metal biooxidation and heavy metal resistance: implications for bioleaching by extreme thermoacidophiles. *Minerals* 5, 397–451. doi: 10.3390/min5030397

Whitaker, R. J. (2003). Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science* 301, 976–978.

Wu, B., Liu, F., Fang, W., Yang, T., Chen, G. H., He, Z., et al. (2021). Microbial sulfur metabolism and environmental implications. *Sci. Total Environ.* 778:146085. doi: 10.1016/j.scitotenv.2021.146085

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43. doi: 10.1093/oxfordjournals.molbev.a026236

Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A., et al. (2010). "Hot standards" for the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* 14, 119–142. doi: 10.1007/s00792-009-0280-0

Conflict of Interest: 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 © 2021 Banerjee, Chaudhari, Lahiri, Gautam, Bhowmik, Dutta, Chattopadhyay, Huson and Paul. 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: 16 July 2021
doi: 10.3389/fmicb.2021.679245

Physiological and Genomic Characterization of a Hyperthermophilic Archaeon *Archaeoglobus neptunius* sp. nov. Isolated From a Deep-Sea Hydrothermal Vent Warrants the Reclassification of the Genus *Archaeoglobus*

Galina Slobodkina[1], Maxime Allioux[2], Alexander Merkel[1], Marie-Anne Cambon-Bonavita[2], Karine Alain[2], Mohamed Jebbar[2] and Alexander Slobodkin[1]*

[1]*Winogradsky Institute of Microbiology, Federal Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia,* [2]*Laboratoire de Microbiologie des Environnements Extrêmes LM2E, Univ Brest, CNRS, IFREMER, IRP 1211 MicrobSea, UMR 6197, Plouzané, France*

OPEN ACCESS

Edited by:
Rafael R. de la Haba, University of Sevilla, Spain

Reviewed by:
Chaomin Sun, Institute of Oceanology, Chinese Academy of Sciences (CAS), China
Tamara N. Nazina, Winogradsky Institute of Microbiology, Russian Academy of Sciences (RAS), Russia
Ida Helene Steen, University of Bergen, Norway

***Correspondence:**
Galina Slobodkina
gslobodkina@mail.ru

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

Received: *11 March 2021*
Accepted: *30 June 2021*
Published: *16 July 2021*

Citation:
Slobodkina G, Allioux M, Merkel A, Cambon-Bonavita M-A, Alain K, Jebbar M and Slobodkin A (2021) Physiological and Genomic Characterization of a Hyperthermophilic Archaeon Archaeoglobus neptunius sp. nov. Isolated From a Deep-Sea Hydrothermal Vent Warrants the Reclassification of the Genus Archaeoglobus. Front. Microbiol. 12:679245. doi: 10.3389/fmicb.2021.679245

Hyperthermophilic archaea of the genus *Archaeoglobus* are the subject of many fundamental and biotechnological researches. Despite their significance, the class *Archaeoglobi* is currently represented by only eight species obtained as axenic cultures and taxonomically characterized. Here, we report the isolation and characterization of a new species of *Archaeoglobus* from a deep-sea hydrothermal vent (Mid-Atlantic Ridge, TAG) for which the name *Archaeoglobus neptunius* sp. nov. is proposed. The type strain is SE56^T (=DSM 110954^T = VKM B-3474^T). The cells of the novel isolate are motile irregular cocci growing at 50–85°C, pH 5.5–7.5, and NaCl concentrations of 1.5–4.5% (w/v). Strain SE56^T grows lithoautotrophically with H_2 as an electron donor, sulfite or thiosulfate as an electron acceptor, and CO_2/HCO_3^- as a carbon source. It is also capable of chemoorganotrophic growth by reduction of sulfate, sulfite, or thiosulfate. The genome of the new isolate consists of a 2,115,826 bp chromosome with an overall G + C content of 46.0 mol%. The whole-genome annotation confirms the key metabolic features of the novel isolate demonstrated experimentally. Genome contains a complete set of genes involved in CO_2 fixation *via* reductive acetyl-CoA pathway, gluconeogenesis, hydrogen and fatty acids oxidation, sulfate reduction, and flagellar motility. The phylogenomic reconstruction based on 122 conserved single-copy archaeal proteins supported by average nucleotide identity (ANI), average amino acid identity (AAI), and alignment fraction (AF) values, indicates a polyphyletic origin of the species currently included into the genus *Archaeoglobus*, warranting its reclassification.

Keywords: sulfate reduction, chemolithoautotroph, anaerobe, hyperthermophile, *Archaea*

INTRODUCTION

Over several decades since their discovery, thermophilic archaea have been the subject of an increasing number of studies related to microbial ecology, biogeochemistry, origin of life, and evolution of the biosphere (Offre et al., 2013). It was long thought that hyperthermophilic archaea represented the deepest and shortest phylogenetic branches of the tree of life and could be considered the first living organisms on Earth since most of them have a chemolithoautotrophic mode of nutrition (Stetter, 2006; Martin and Sousa, 2016; Weiss et al., 2016), but this hypothesis is now being questioned. Indeed, numerous archaeal lineages, with new complete genomes or high-quality *metagenomes-assembled genomes* (MAGs) and *single-amplified genomes* (SAGs), have been discovered thanks to remarkable advances in sequencing techniques and in data processing capabilities, allowing the construction of robust phylogenetic trees based on several genetic markers, which demonstrate that the root is far from being resolved (Baker et al., 2020). Dissimilatory sulfate reduction is a microbial process with significant ecological and biogeochemical implication (Rabus et al., 2015). It is mainly related to the marine environments because of the high sulfate concentrations. In the marine ecosystems with temperatures above 80°C, the only conclusively proven actors of this process are the representatives of the genus *Archaeoglobus* that was established by Stetter et al. (1987) and Stetter (1988). The type strain, *A. fulgidus* VC-16^T, is one of the best studied *Archaea*, as it was the first described hyperthermophilic sulfate-reducing archaea, and one of the first archaea whose genome has been sequenced (Stetter, 1988; Klenk et al., 1997). In addition to physiological properties given in the original description, new metabolic features were found based on genome sequencing data analysis, e.g., growth coupled with carbon monoxide oxidation and sulfate reduction or acetogenesis (Henstra et al., 2007); the ability for anaerobic oxidation of fatty acids, *n*-alkenes, and *n*-alkanes (Khelifi et al., 2010, 2014); and growth with (per)chlorate reduction combining biotic and abiotic reactions (Liebensteiner et al., 2013). Along with the type strain, *A. fulgidus* 7324 has been isolated from North Sea oilfield waters and its genome has been sequenced (Beeder et al., 1994; Birkeland et al., 2017).

Archaeoglobus fulgidus became a source and a model organism for the study of structure and function of many thermostable enzymes. A good example is *A. fulgidus* ferritin (AfFt) whose unique structure and properties make it useful in clinical therapy (Palombarini et al., 2020). *Archaeoglobus fulgidus* is also a promising candidate to clean up oil-contaminated environments at low cost and with high efficiency due to its high resistance to anaerobic conditions, high temperatures, and high salts concentrations, and its ability to degrade alkanes (Park and Park, 2018).

At the time of writing, the genus *Archaeoglobus* includes five species with validly published names: *A. fulgidus* (Stetter, 1988), *A. profundus* (Burggraf et al., 1990), *A. veneficus* (Huber et al., 1997), *A. infectus* (Mori et al., 2008). and *A. sulfaticallidus* (Steinsbu et al., 2010). Together with related genera, *Geoglobus* and *Ferroglobus*, they compose the family *Archaeoglobaceae* belonging to the order *Archaeoglobales*, class *Archaeoglobi* (Parte et al., 2020; http://www.bacterio.net/index.html). It is notable that class *Archaeoglobi* is represented by only eight species obtained as axenic cultures and taxonomically characterized in detail. All known representatives of *Archaeoglobi* are strict anaerobes and hyperthermophiles isolated from marine hydrothermal systems and off-shore oil reservoirs. With the exception of *A. infectus* and *A. profundus*, all species of the order *Archaeoglobales* are capable of chemolithoautotrophic growth. Members of this order differ significantly in terms of the electron acceptors they use. All known species of the genus *Archaeoglobus* are able to grow by reduction of sulfite and thiosulfate. *Archaeoglobus fulgidus*, *A. profundus*, and *A. sulfaticallidus* can also reduce sulfate with organic carbon source, but only *A. sulfaticallidus* is capable of lithoautotrophic growth with sulfate as a terminal electron acceptor. Species of the genus *Geoglobus* are obligate iron-reducers (Kashefi et al., 2002; Slobodkina et al., 2009). The only representative of the genus *Ferroglobus*, *F. placidus,* can use nitrate, thiosulfate, and Fe^{3+} as electron acceptors (Hafenbradl et al., 1996; Tor et al., 2001).

Over the last decade, genomic data have increased exponentially and revealed many phylogenetic inconsistencies that require revision and reclassification of existing prokaryotic taxonomy and nomenclature. These taxonomic issues were recently addressed by proposing a standardized taxonomy for *Bacteria* and *Archaea* referred to as the Genome Taxonomy DataBase (GTDB; Parks et al., 2018; https://gtdb.ecogenomic.org/), based on average nucleotide criteria to delineate species and on sets of protein markers to infer a taxonomic position and define genomic clusters. *Archaea* were recognized as a separate domain in 1977 on the basis of small subunit (SSU) ribosomal RNA (rRNA) gene sequences analysis (Woese and Fox, 1977) and this phylogenetic marker was used for their classification afterward. Therefore, compared to *Bacteria*, which classification has long been relied on phenotypic properties, the *Archaea* are less affected by historical misclassifications (Zuo et al., 2015). The *Archaeoglobus* species were classified into the genus based mostly on phylogenetic analyses of 16S rRNA gene sequences. According to the GTDB, the genus *Archaeoglobus* encompasses four distinct clusters of genomic sequences. In this paper, we describe a novel thermophilic and facultative lithoautotrophic strain SE56^T, belonging to the genus *Archaeoglobus*. To further understand the physiological features and genome characteristics of this novel species, the whole-genome of strain SE56^T was sequenced and analyzed. An advanced phylogenomic reconstruction based on 122 conserved single-copy archaeal protein markers and supported by *Average Nucleotide Identity* (ANI), *Alignment Fraction* (AF), and *Average Amino Acid Identity* (AAI) values clearly indicated

Abbreviations: AAI, Average amino acid identity; AF, Alignment fraction; ANI, Average nucleotide identity; bp, Base pair; CDS, Coding DNA sequences; COGs, Clusters of orthologous groups; CRISPR, Clustered regularly interspaced short palindromic repeats; dDDH, digital DNA–DNA hybridization; DNA, deoxyribonucleic acid; EMBL, European Molecular Biology Laboratory; IMG/M, Integrated microbial genomes and microbiomes; GI, Genomic islands; GTDB, Genome taxonomy database; NCBI, National Center for Biotechnology Information; PGAP, Prokaryotic genome annotation pipeline; rRNA, Ribosomal ribonucleic acid; TCA, Tricarboxylic acid cycle.

a polyphyletic origin of the species currently affiliated to the genus *Archaeoglobus*.

MATERIALS AND METHODS

Origin of the Strain

Strain SE56^T was isolated from a hydrothermal diffuser fragment (from the sample BIC2-PL1917-11-PBT1-02) collected in 2018 during the BICOSE 2 oceanographic cruise,[1] using the manned operated submersible Nautile, at the TAG vent field (26°13'69"N 44°82'61"W, 3,625 m water depth) on the Mid-Atlantic Ridge. Onboard, the samples were immediately subsampled under sterile conditions. Small rock fragments were placed into 50 ml glass flasks with *in situ* seawater under N_2 flow, closed with a rubber stopper and aluminum cap, and were stored at 4°C.

Media and Cultivation

The medium for enrichment and isolation contained (per liter distilled water): 0.33 g NH_4Cl, 0.33 g KCl, 0.33 g $CaCl_2$•$6H_2O$, 0.33 g KH_2PO_4, 18.0 g NaCl, 4.33 g $MgCl_2$•$6H_2O$, 2.0 g $NaHCO_3$, 0.001 g resazurin, 0.5 g Na_2S·$9H_2O$, 1 ml trace element solution (Slobodkin et al., 2012) and 1 ml vitamin solution (Wolin et al., 1963). The isolation medium had a pH of 6.5–6.8 (measured at 25°C). The medium was dispensed in 10 ml portions into 17 ml Hungate tubes and heat-sterilized at 121°C for 30 min; the gas phase consisted of H_2/CO_2 (20: 80 v/v). Sodium sulfite from a sterile anoxic stock solution was added as an electron acceptor to a final concentration of 5 mM. All transfers and sampling of cultures were performed with syringes. All incubations were at 80°C unless otherwise noted.

Phenotypic Characterization

Cell growth was determined after acridine orange staining, by direct cell counting using a phase-contrast microscope (Olympus CX-41, Olympus Corp., Japan). The morphological properties of strain SE56^T were observed using light microscopy and transmission electron microscopy (JEM-100, JEOL, Japan). For Gram-staining cells from exponential and stationary phases of growth were used (Gerhardt et al., 1994). Determination of temperature, pH, and salinity ranges for growth was carried out as described previously (Slobodkina et al., 2016) in the same medium used for isolation, with pyruvate (10 mM) instead of molecular hydrogen. The influence of temperature on growth of strain SE56^T was determined in the range 40–95°C with an interval of 5°C. The effect of pH on growth was investigated at 80°C in the range 4.0–9.5 with a 0.5 pH unit interval. The salt requirement was tested at 80°C and pH 6.5 in the range 0–7% NaCl (w/v) with an interval of 0.5% (w/v). All experiments were performed in duplicate, whereas testing the effect of pH on growth was performed in triplicate. Soluble electron donors and acceptors were added from sterile anoxic stock solutions before inoculation. Elemental sulfur was added directly into each test tube with liquid medium prior to tindallisation. Fe(III) was provided in the form of amorphous iron(III) oxide (ferrihydrite) at about 90 mmol Fe(III) L^{-1}. The ferrihydrite was synthesized by titrating a solution of $FeCl_3$, with 10% (w/v) NaOH to pH 9.0. The effect of O_2 on growth of strain SE56^T was tested in 60 ml flasks containing 10 ml of the medium The flasks were sealed with a rubber stopper and aluminum screw cap. To check microaerobic growth, anoxic medium with pyruvate (10 mM) and without an electron acceptor, reducing agent, and resazurin was used. The headspace of the flask was filled with CO_2 (100%). Various amounts of sterile air were injected by syringes in the headspace to obtain oxygen concentrations of 0.5, 2.0, or 5.0% (v/v). The ability of the strain to aerobic growth was determined in the same medium prepared in air. The strain was regarded as utilizing the added electron acceptors and donors if growth was sustained after at least three subsequent transfers into fresh medium.

Genome Sequencing, Assembly, and Analysis

For genomic DNA extraction, the strain was cultivated at 80°C with pyruvate (10 mM) and sodium sulfite (5 mM) as electron donor and acceptor, respectively. Cells were harvested in the late exponential phase of growth by centrifugation at 4°C (15,000 × *g*, 20 min). The DNA was extracted using a FastDNA™ Spin Kit (MP Biomedicals, United States) according to the manufacturer's instructions. The whole genome was sequenced using the Illumina nanoMiSeq technology (Fasteris, Switzerland; 2 × 150 paired-reads, Nano V2 chemistry). Clean reads from short-reads sequencing were assembled and circularized using SPAdes v1.13.1 assembly pipeline (Bankevich et al., 2012; https://github.com/ablab/spades). Average coverage was calculated using BBtools (BBMap – Bushnell B., v38.70 – sourceforge.net/projects/bbmap/). Contamination and completeness were assessed by CheckM v1.1.2 (Parks et al., 2015; https://github.com/Ecogenomics/CheckM). Gene search and annotation were performed by means of the Rapid Annotation using Subsystem Technology (RAST)/SEED v2.0 pipeline (Overbeek et al., 2014; https://rast.nmpdr.org/), the Integrated Microbial genomes (IMGs)/M v.5.0 analysis system (Chen et al., 2019; https://img.jgi.doe.gov/), and National Center of Biotechnology Information (NCBI) integrated prokaryotic genome annotation pipeline (PAGP; Tatusova et al., 2016). Identification and classification of the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system was performed by the CRISPRCas Finder web server (Couvin et al., 2018; https://crispr.i2bc.paris-saclay.fr/Server/). The prediction of laterally transferred gene clusters [genomic islands (GI)] was performed with the IslandViewer4 web server (Bertelli et al., 2017; http://www.pathogenomics.sfu.ca/islandviewer/). Genome and gene clusters visualization was made with the CGView program (Grant and Stothard, 2008; https://paulstothard.github.io/cgview/) and Gene Graphics web application (Harrison et al., 2018; https://katlabs.cc/genegraphics/app). Analysis of the clusters of orthologous genes (COG) functional categories was performed with the eggNOG-Mapper v2 (Huerta-Cepas et al., 2017; http://eggnog-mapper.embl.de/). Hydrogenase classification has been checked using the HydDB webtool (Søndergaard et al., 2016; https://services.birc.au.dk/hyddb/).

[1]http://dx.doi.org/10.17600/18000004

Phylogenetic and Phylogenomic Analyses

For phylogenetic analysis, the 16S rRNA gene sequence of the isolate (retrieved from the whole-genome sequencing) was compared with other sequences in GenBank (Benson et al., 1999) using the BLAST program (Altschul et al., 1990; http://www.ncbi.nlm.nih.gov/BLAST/) and by means of the EzBioCloud server (Yoon et al., 2017; https://www.ezbiocloud.net/) to identify its closest relatives. Alignment with a representative set of related 16S rRNA gene sequences from the GenBank database was carried out with the CLUSTALW program implemented in the MEGA software package v7.0 (Kumar et al., 2016). Evolutionary analysis and phylogenetic tree reconstruction used the Maximum-Likelihood algorithm based on the Tamura–Nei model and the Neighbor-Joining methods (Saitou and Nei, 1987; Tamura and Nei, 1993) provided by MEGA v7.0. ANI were calculated using three methods: (i) OrthoANIu, the orthologous ANI algorithm using the USEARCH program (Edgar, 2010) provided by the EzBioCloud ANI calculator[2]; (ii) ANIb, the algorithm using BLASTN (Goris et al., 2007) provided by JSpeciesWS Online Service (Richter et al., 2016; http://jspecies.ribohost.com/jspeciesws/#analyse); and (iii) gANI obtained by the Microbial Species Identifier (MiSI) method (Varghese et al., 2015) using ANIcalculator implemented in the IMG/M system.[3] The alignment fraction (AF) values were also obtained *via* the Pairwise ANI tool implemented in the IMG/M online service. Digital DNA–DNA hybridizations (dDDH) were performed using the genome-to-genome distance (GGDC) method with the GGDC 2.0 blast+ model provided by the Genome-to-Genome Distance Calculator (Meier-Kolthoff et al., 2013; https://www.dsmz.de/). The AAI between the selected genomes was calculated using the aai.rb script from the enveomics collection.[4] The list of 122 archaeal marker genes used for phylogenetic inference was taken from the GTDB (Parks et al., 2018). These marker genes were extracted from genomes using GTDB-Tk v1.3.0 (Chaumeil et al., 2019; https://github.com/Ecogenomics/GTDBTk), aligned using mafft v7.475 (Katoh and Standley, 2013), trimmed using trimAl 1.2 (Capella-Gutiérrez et al., 2009), and concatenated. The phylogenomic tree was built using PhyML v3.3.2 (Guindon et al., 2010) and the Bayesian like transformation of approximate likelihood-ratio test for branches (Anisimova et al., 2011). LG was selected as the best substitution model by the SMS algorithm (Le and Gascuel, 2008; Lefort et al., 2017).

RESULTS AND DISCUSSION

Isolation of the Strain SE56^T

An enrichment culture was initiated by inoculation of the mixture of small rock fragments and sea water [at 10% (w/v)] into anoxically prepared, bicarbonate-buffered, sterile liquid medium with molecular hydrogen as an electron donor, sulfite as an electron acceptor, and CO_2/HCO_3^- as a carbon source. After incubation in the dark for 10 days at 80°C, growth of irregular-shaped cells was observed. After three subsequent transfers and following serial 10-fold dilutions in the same medium, only one morphological type was observed in the highest growth-positive dilution (10^{-6}). A pure culture of strain SE56^T was obtained by performing multiple dilution-to extinction series in the same medium. Purity of the isolate was confirmed by routine microscopic examination on various media, and sequencing of the 16S rRNA gene and genome.

16S rRNA Gene Phylogenetic Analysis

Phylogenetic analysis based on comparison of 16S rRNA gene sequences showed that the strain SE56^T belonged to the genus *Archaeoglobus* (**Figure 1**; **Supplementary Figure S1**) and had the highest sequence similarity to *A fulgidus* VC-16^T (98.6%). Sequence similarities between the 16S rRNA gene sequence of the isolate and those of other representatives in the order *Archaeoglobales* were 96.3–94.4%.

Cell Morphology, Physiology, and Metabolism

Cells of strain SE56^T were motile irregular cocci (approximately 0.6–0.8 μm in diameter), usually occurring singly (**Figure 2**). Cells stained Gram-negative and showed blue-green fluorescence when exposed to UV under a fluorescence microscope.

The temperature range for the growth of strain SE56^T was 50–85°C, with an optimum at 80°C. No growth was detected at 90°C or above, neither at 45°C or below, after 3 weeks of incubation. The pH range for growth was 5.5–7.5, with an optimum at pH 6.5. No growth was observed at pH 5.0 or below or at pH 8.0 or above. Growth of the isolate was observed at NaCl concentrations ranging from 1.5 to 4.5% (w/v), with an optimum at 2.0–2.5% (w/v). No growth was evident neither at 5.0% NaCl (w/v) or above, nor at 1.0% NaCl (w/v) and below (**Supplementary Figure S2**). The doubling time under optimal conditions was about 2 h. Yeast extract was not necessary for growth and did not stimulate it.

Growth with potential organic substrates was tested with sulfite as a terminal electron acceptor. The best growth (final cell density 2–4•10^8 cells ml^{-1}) was observed with pyruvate (10 mM), but growth was not observed after 7 days of incubation in the same conditions with the addition of elemental sulfur (5 g L^{-1}). Acetate propionate, butyrate, fumarate, succinate (10 mM each), stearate (1 mM), peptone, tryptone, and yeast extract (2 g L^{-1} each) also sustained growth of the isolate (final cell density 5–9•10^7 cells ml^{-1}). No growth was observed with formate, methanol, ethanol, isopropanol, glycerol, citrate, lactate, malate (10 mM each), fructose, glucose, sucrose, maltose, xylose, starch, and alginate (2 g L^{-1} each) during 4 weeks of incubation under optimal growth conditions.

Potential electron acceptors were tested with molecular hydrogen [H_2/CO_2 (20:80, v/v)], pyruvate (10 mM), or in simultaneous presence of hydrogen and pyruvate as energy and/or carbon sources. Strain SE56^T could grow lithoautotrophically with molecular hydrogen as an electron donor, sulfite (5 mM) or thiosulfate (15 mM) as a terminal

[2]https://www.ezbiocloud.net/tools/ani
[3]https://img.jgi.doe.gov/
[4]https://github.com/lmrodriguezr/enveomics

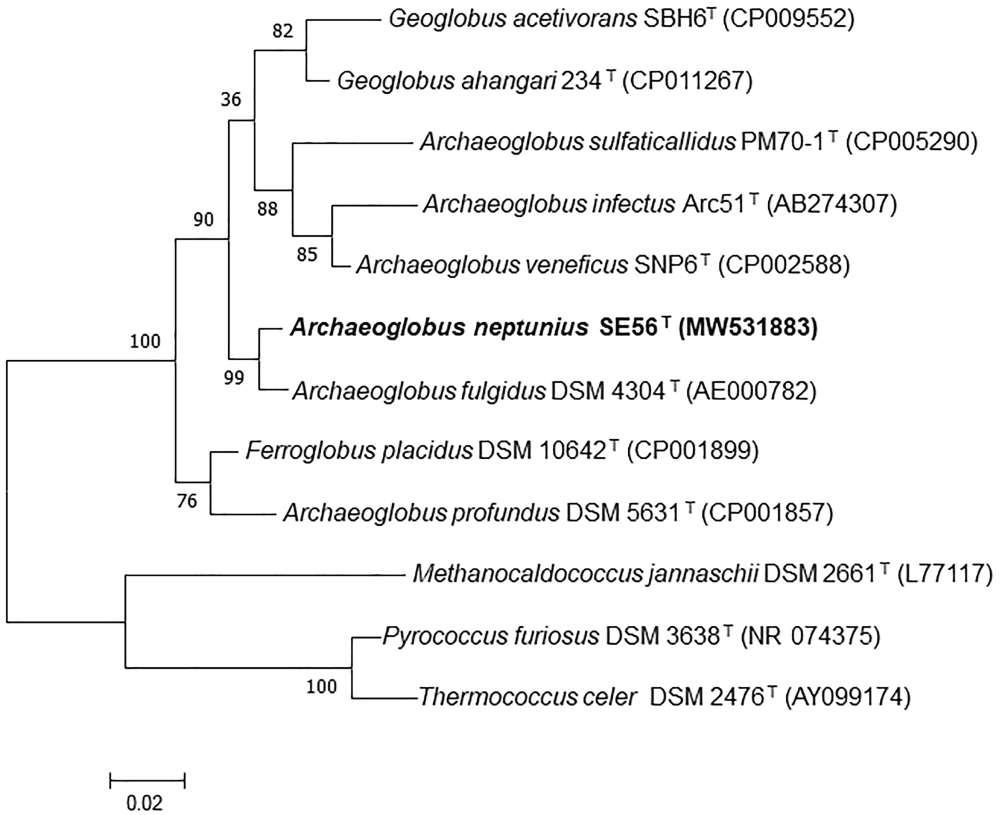


FIGURE 1 | Maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences showing the position of the strain SE56^T among the other members of the class *Archaeoglobi*. Bootstrap values based on 1,000 replicates are shown at branch nodes. *Methanocaldococcus jannaschii* DSM 2661^T, *Pyrococcus furiosus* DSM 3638^T, and *Thermococcus celer* DSM 2476^T were used as outgroup. Bar, 0.02 substitutions per nucleotide position.

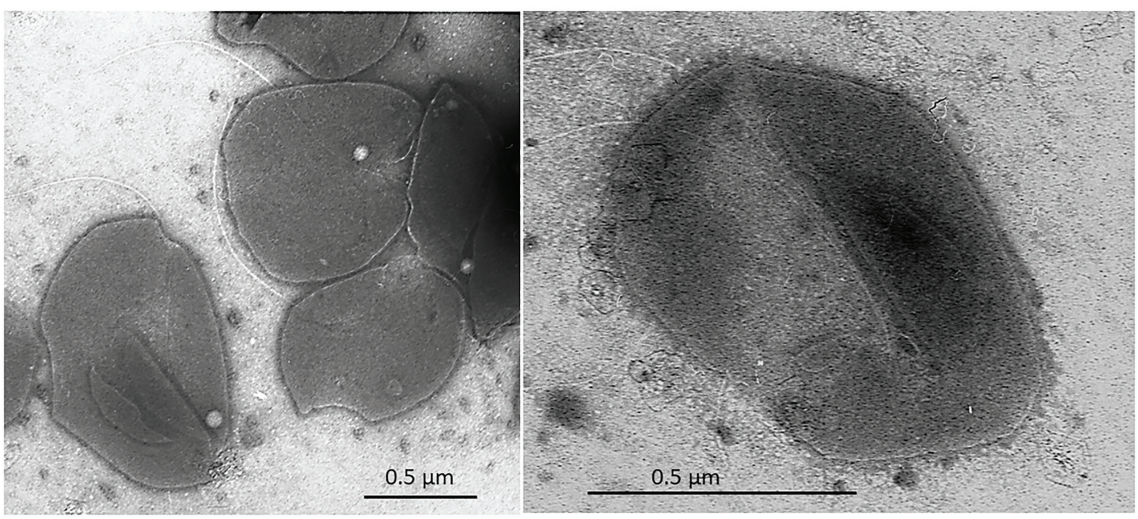


FIGURE 2 | Transmission electron micrograph (TEM) of negatively stained cells of *Archaeoglobus neptunius* SE56^T sp. nov. shows an overall cell morphology. Bar, 0.5 µm.

electron acceptor, and CO_2/HCO_3^- as carbon source. Under lithoautrotrophic conditions with sulfate (15 mM) as a terminal electron acceptor, no growth was observed. Growth with sulfate (15 mM), sulfite (5 mM), or thiosulfate (15 mM) was observed with pyruvate (10 mM) or with hydrogen combined with pyruvate. Elemental sulfur (5 g L^{-1}), fumarate, nitrate

(10 mM each), nitrite (2.5 mM), chlorate, perchlorate (5 mM each), and ferrihydrite [90 mmol Fe(III) L^{-1}] did not support growth under any tested conditions. Strain SE56^T did not grow aerobically or when oxygen (0.5, 2.0, or 5.0%, v/v) was added to the gas phase. No growth was observed with pyruvate (10 mM), yeast extract, or peptone (2 g L^{-1} each) in the absence of an electron acceptor.

The novel isolate shares general features with the other members of the genus, such as morphology, growth conditions, capacity to reduce oxidized sulfur compounds, and inability to grow through iron or nitrate reduction. The differential characteristics of strain SE56^T and other species of the genus *Archaeoglobus* are summarized in **Table 1**.

Genome Statistics

The draft genome of strain SE56^T was assembled into 32 contigs with genome size of 2,115,826 bp and a N_{50} value of 138,848 bp. Final assembly coverage was 153.537×. Based on CheckM analysis of single copy-core genes, the estimate of genome completeness was 100% and the estimate of contamination was 0%. The genomic DNA G + C content was 46.0 mol%. Annotation with prokaryotic genome annotation pipeline (PGAP) resulted in prediction of 2,471 genes, 2,386 of which are protein-coding sequences (CDSs) that cover about 96.5% of the entire genome. Genome also contained one copy of 5, 16, and 23S rRNA genes and 47 tRNA genes for all 20 standard amino acids. Two CRISPR loci were found (**Figure 3**).

Majority of the CDSs (72.3%) could be assigned to at least one COG group. The main predicted COG categories (encompassing more than 100 CDSs) were energy production and conversion (11.73%); amino acid transport and metabolism (8.60%); translation, ribosomal structure, and biogenesis (8.60%); transcription (6.73%); lipid transport and metabolism (5.87%); and coenzyme transport and metabolism (5.72%) and replication, recombination, and repair (5.36%).

The prediction of laterally transferred genes showed that the genome of strain SE56^T possesses four GI with a total length of 112.5 kb (**Figure 3**). Few genes were found encoding for enzymes related to archaeal chromosome replication, segregation, and cell division, such as ribonucleotide reductase, DNA helicase, transcription factor B, chromosome partitioning protein ParA, and cell division protein FtsH. Genes for 5'-nucleotidase SurE, AMP phosphorylase, and ribose-5-phosphate isomerase A were present as well, while the majority of CDSs located on the genomic islands encoded for hypothetical proteins (**Supplementary Table S1**).

Central Carbon Metabolism

Strain SE56^T is capable of autotrophic growth using CO_2 as a carbon source. Like all *Archaeoglobi* species except *A. profundus*, strain SE56^T has the complete set of genes for the CO_2 fixation *via* reductive acetyl-CoA (Wood-Ljungdahl) pathway (**Table 2**; Klenk et al., 1997; von Jan et al., 2010; Anderson et al., 2011; Stokke et al., 2013; Manzella et al., 2015; Mardanov et al., 2015). The genome also contains a gene coding for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO, WP_202319886). This enzyme belongs to form III RubisCO, which does not participate in autotrophy in *Archaea* (Berg, 2011). Many other enzymes of the Calvin-Benson cycle are missing. The key enzymes of other known microbial carbon fixation pathways, namely, ATP-citrate lyase and citryl-CoA lyase (two variants of the reductive tricarboxylic acid cycle), malonyl-CoA reductase (3-hydroxypropionate bi-cycle), and most of enzymes participating in 3-hydroxypropionate/4-hydroxybutyrate and dicarboxylate/4-hydroxybutyrate cycles are also absent.

TABLE 1 | Differential characteristics of strain SE56^T and the type stains of the species of the genus *Archaeoglobus*.

	***A. neptunius* SE56^T**	***A. fulgidus* VC-16^T**	***A. profundus* AV18^T**	***A. veneficus* SNP6^T**	***A. infectus* Arc51^T**	***A. sulfaticallidus* PM70-1^T**
Location of hydrothermal vent, water depth	Mid-Atlantic Ridge, TAG vent field, 3,625 m	Shallow-sea hydrothermal vent, Volcano, Italy	Guaymas Basin, Mexico, 2,000 m	Mid-Atlantic Ridge, "Snake Pit" site, 3,500 m	Pacific Ocean, Suiyo Seamount, Izu-Bonin Arc, 1,380 m	Pacific Ocean, Juan de Fuca Ridge, 2,658 m
Cell size, μm	0.6–0.8	0.1–1.0	1.3	0.5–1.2	0.5–1.0	0.4–2.2
Growth conditions						
Temperature (opt; °C)	50–85 (80)	60–95 (83)	65–90 (82)	65–85 (80)	60–75 (75)	60–80 (75)
pH (opt)	5.5–7.5 (6.5)	5.5–7.5 (ND)	4.5–7.5 (6.0)	6.5–8.0 (7.0)	6.5–7.0 (6.5)	6.3–7.6 (7.0)
NaCl (opt) % (w/v)	1.5–4.5 (2.0–2.5)	ND	0.9–3.6 (1.8)	0.5–4.0 (2.0)	1.0–4.0 (3.0)	0.5–3.5 (2.0)
Chemolithoautotrophic growth	+	+	–	+	–	+
SO_4^{2-} reduction	+	+	+	–	–	+
Electron donors						
Formate	–	+	–	+	–	–
Butyrate	+	+	–	ND	–	–
Pyruvate	+	+	–	+	–	+
Lactate	–	+	–	–	–	+
Fumarate	+	–	–	w	–	–
Yeast extract	+	+	–	+	–	–

Data for reference strains were taken from Stetter, 1988; Burggraf et al., 1990; Huber et al., 1997; Mori et al., 2008 and Steinsbu et al., 2010 (Archaeoglobus fulgidus VC-16^T, Archaeoglobus profundus AV18^T, Achaeoglobus veneficus SNP6^T, Archaeoglobus infectus Arc51^T and Archaeoglobus sulfaticallidus PM70-1^T, respectively). All strains were isolated from marine hydrothermal vents. +, Positive; (w), weakly positive; –, negative; and ND, not determined.

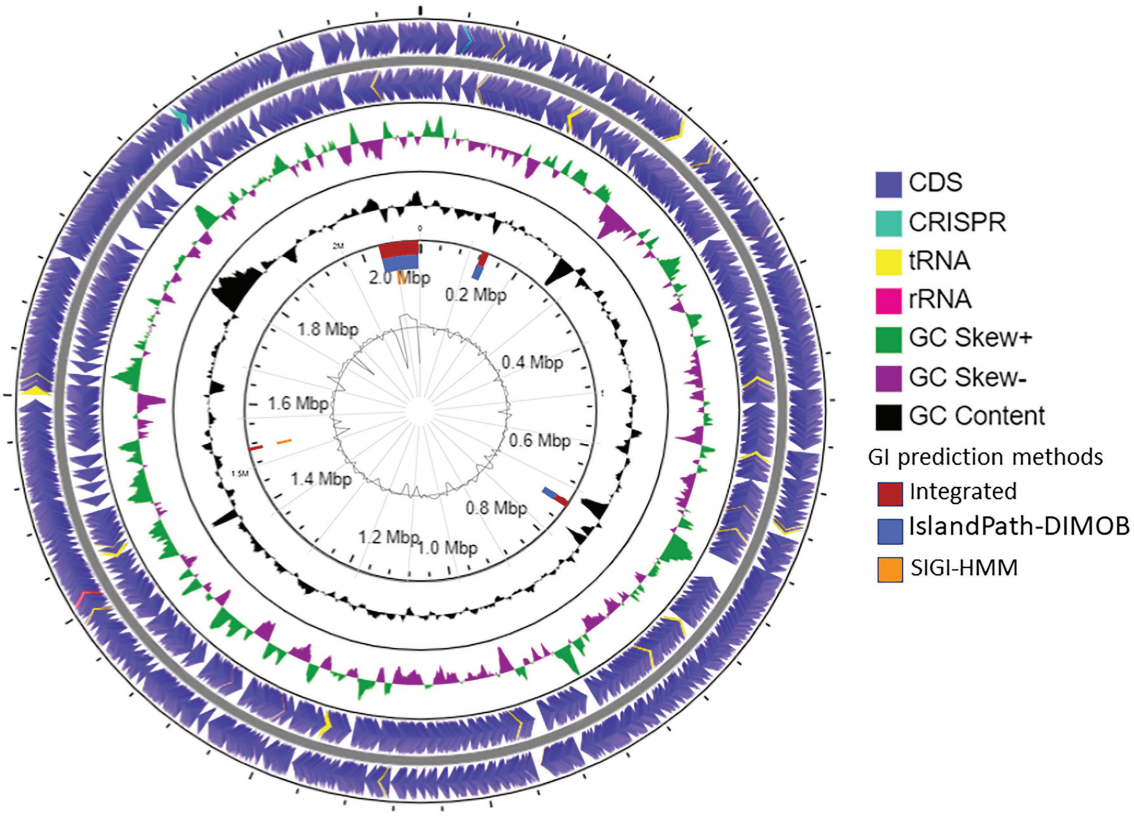


FIGURE 3 | Schematic representation of the genome of *A. neptunius* SE56^T. Labeling from the outside to the center is as follows: circle 1, genes on the forward strand; circle 2, genes on reverse strand [tRNAs yellow, ribosomal RNAs (rRNAs) lilac pink, and clustered regularly interspaced short palindromic repeats (CRISPR) cyan blue]; circle 3, G + C skew; circle 4, G + C content; circle 5, genomic islands shown as red, blue, and orange rectangles attributed, respectively, to integrated, IslandPath-DIMOB and SIGI-HMM prediction genomic islands methods; and circle 6, IslandViewer4 automatic calculation of the G + C content.

TABLE 2 | Gene sets encoding various physiological properties in genomes of type strains of *Archaeoglobi*.

	CO_2 fixation (W-L)	Gluconeogenesis (E-M)	Sulfate reduction	Fatty acids β-oxidation	Hydrogen oxidation	Flagella assembly	Chemotaxis
A. neptunius SE56^T	C	C	C	C	+	+	+
A. fulgidus VC-16^T	C	C	C	C	+	+	+
A. profundus AV 18^T	INC	C	C	INC	+	+	+
A. veneficus SNP6^T	C	C	C	INC	+	+	+
A. sulfaticallidus PM70-1^T	C	C	C	INC	+	+	–
G. ahangari 234^T	C	C	INC	C	+	+	–
G. acetivorans SBH6^T	C	C	INC	C	+	+	+
F. placidus DSM 10642^T	C	C	INC	C	+	+	+

W-L, Wood-Ljungdahl pathway; E-M, Embden-Meyerhof pathway; C, complete gene set for the pathway; INC, incomplete gene set for the pathway; +, genes present; and –, genes absent.

The genome of SE56^T contains most of the genes encoding the Embden-Meyerhof pathway enzymes (**Table 2**), with the exception of the enzymes that catalyze irreversible reactions of glycolysis, pyruvate kinase, and 6-phosphofructokinase. Instead, there are genes encoding fructose-bisphosphatase (WP_202319097) and phosphoenolpyruvate synthase (WP_202318379) that enable glucose production from pyruvate. Thus, the Embden-Meyerhof pathway in this organism operates in reverse direction, toward gluconeogenesis.

Although strain SE56^T lacks the key enzymes of the "standard" tricarboxylic acid cycle (TCA) cycle, aconitase and 2-oxoglutarate dehydrogenase complex, most of the genes involved in this cycle including citrate synthase (WP_202320798) are present. In this case, 2-oxoglutarate synthase may substitute the missing 2-oxoglutarate dehydrogenase. At least, two genes coding for 2-oxoglutarate synthase (*oorCBAD*) were found in the genome of strain SE56^T. As well as the majority of *Euryarchaeota* whose genomes are sequenced, the genome of strain SE56^T does not

encode orthologs of aconitase A or members of other aconitase families. A putative family of aconitases, aconitase X, has been proposed on the basis of comparative-genomic analysis (Makarova and Koonin, 2003). The predicted aconitase X consists of two proteins that function together as a TCA cycle enzyme catalyzing the citrate to isocitrate isomerization. Experimental validation of this prediction has not yet been received. Nevertheless, a complete oxidative TCA is declared to function in the cells of class *Archaeoglobi* members (*A. fulgidus*, *A. profundus*, *F. placidus*, *Geoglobus ahangari*, and *Geoglobus acetivorans*). In the genome of strain SE56^T, two genes (WP_202319516 and WP_202319760) were found encoding proteins that have 70% similarity with the proteins of the putative aconitase X from *A. fulgidus* (AF2333 and AF0055). Alternatively, incomplete TCA cycle can be used in the isolate SE56^T to provide intermediates for anabolic pathways. Acetyl-CoA formed in Wood-Ljungdahl pathway or from acetate can be converted to pyruvate by a pyruvate synthase POR (*porBADG*, WP_202318910–WP_202318913). Pyruvate is converted to oxaloacetate in an irreversible reaction catalyzed by pyruvate carboxylase (*pycAB*, WP_202320749, and WP_202320510). Further formation of the necessary biosynthetic intermediates, in particular, succinyl-CoA and 2-oxoglutarate, is enabled by the presence of corresponding enzymes. This pathway has been experimentally shown in methanogenic archaea (Fuchs and Stupperich, 1978; Shieh and Whitman, 1987). Each of *Archaeoglobi*'s available genome contains homologs of all the genes involved in this incomplete TCA cycle.

The capacity of strain SE56^T to utilize propionate is confirmed by a number of genes putatively encoding enzymes involved in propionate oxidation *via* the methylmalonyl-CoA pathway. This includes glutaconate CoA-transferase (WP_202318740 and WP_202318741) and a short *mmc* cluster (WP_202319622–WP_202319626) consisting of propionyl-CoA carboxylase, methylmalonyl-CoA epimerase and methylmalonyl-CoA mutase. Several enzymes such as succinyl-CoA synthase (SCS; WP_202320107 and WP_202320108), succinate dehydrogenase (SDH; WP_202319710, WP_202319660-, and WP_202319663), and fumarate hydratase (FHT; WP_202319682 and WP_202319683) can participate in both propionate oxidation and TCA cycle. The NAD-dependent malic enzyme (*maeA*, WP_202320482), pyruvate ferredoxin oxidoreductase (*porBADG*, WP_202318910–WP_202318913), and/or formate acetyltransferase (WP_202319769) can be involved in final steps of propionate oxidation to acetyl-CoA.

The genome contains a large number of genes coding for fatty acid utilization enzymes *via* β-oxidation pathway (**Table 2**), including long-chain acyl-CoA synthetases, acyl-CoA dehydrogenases, enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase, and acetyl-CoA acyltransferases. The genome also possesses 13 genes encoding acetyl-CoA C-acetyltransferase that are involved in fatty acid metabolism too, mainly in butyrate oxidation. This is consistent with the ability of strain SE56^T to utilize butyrate and stearate. *Archaeoglobus fulgidus*, as well as the species of the related genus *Geoglobus*, are known to use fatty acids as electron donors (Kashefi et al., 2002; Slobodkina et al., 2009; Khelifi et al., 2010). In the genomes of these organisms and in the genome of *Ferroglobus placidus*, a large number of genes encoding the core enzymes of fatty acid β-oxidation have been found (Klenk et al., 1997; Anderson et al., 2011; Manzella et al., 2015; Mardanov et al., 2015). In contrast, in the genome of *A. veneficus*, we found only one gene of long-chain acyl-CoA synthetase and two genes of acetyl-CoA acetyltransferase. The genome of *A. sulfaticallidus* which was reported as unable to utilize butyrate (Steinsbu et al., 2010), lacks genes of two of the key enzymes for β-oxidation: enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase. The same genes are missing in the genome of *A. profundus*, as previously reported (von Jan et al., 2010).

Our attempts to grow strain SE56^T with lactate as the electron donor and sulfite or sulfate as the electron acceptor were unsuccessful, although two known strains of *A. fulgidus* are capable of oxidizing lactate with sulfate. The genes for D-lactate and L-lactate dehydrogenases (*dld*, AF0394 and *lldD*, AF0807) were annotated in the genome of the type strain *A. fulgidus* (Klenk et al., 1997). The transcriptomic study proposed that lldD (AF0807) is part of a cluster and operates along with L-lactate permease (AF0806), monomeric dld (AF0808), and oligomeric LldEFG (AF0809–AF0811) lactate dehydrogenases (Hocking et al., 2014). This study also suggested to consider two more oxidoreductases (AF0507 and AF0867) as putative lactate dehydrogenases. In the genome of strain SE56^T, we have found close homologs (62–78%) of proteins encoded by AF0394, AF0507, and AF0867 and the homologs of the subunits of LldEFG lactate dehydrogenase. Yet, the homologs of L-lactate permease and lactate dehydrogenases lldD and dld, encoded by the genes AF0806–AF0808 were absent. Thus, the reason for the absence of growth of strain SE56^T with lactate is unclear. This may be because the missing enzymes are essential for lactate metabolism, or because the conditions for growth of strain SE56^T with lactate are not yet identified.

Hydrogen Oxidation

A cluster of genes encoding a membrane-bound hydrogen uptake [NiFe] hydrogenase (Vht), a cytoplasmic [NiFe] hydrogenase:heterodisulfide (Mvh:HdrABC) complex, and a set of hydrogenase maturation proteins (WP_202319836–WP_202319849) provides strain SE56^T the ability to utilize molecular hydrogen (**Table 2**). This cluster is syntenic to the one encoding the hydrogenases in *A. fulgidus* (AF1365–AF1381). Large catalytic subunits of Vht hydrogenase (vhtA) and Mvh hydrogenase (mvhA) of strain SE56^T have 83 and 63% of amino acids similarities with corresponding proteins in *A. fulgidus*. According to a model of hydrogen metabolism proposed for *A. fulgidus* (Hocking et al., 2014), Vht hydrogenase is required for energy conservation which is achieved by electron transfer from hydrogen to sulfite *via* DsrMK/HdrDE complexes, where a particular putative heterodisulfide reductase HdrDE (AF0755) may play an important role. We detected close homolog of this HdrDE (WP_202318846) with 75% amino acid similarity in strain SE56^T. Thus, the same model of hydrogenotrophy might be operative in *A. neptunius* SE56^T cells.

Reduction of Sulfate and Other Potential Electron Acceptors

Genome of strain SE56^T contains the complete set of genes involved in dissimilatory sulfate reduction (**Table 2**), including sulfate adenylyltransferase *sat* and adenylyl-sulfate reductase *aprAB* (WP_202320080 and WP_202320082–WP_202320083), manganese-dependent inorganic pyrophosphatase *ppaC* (WP_202318845), dissimilatory sulfite reductase *dsrABD* (WP_202320149–WP_202320151), and *dsrC* which is more distantly located (WP_202320007). Genes coding for the electron transfer complexes *dsrMKJOP* (WP_202318457–WP_202318460) and *QmoABC* (WP_202319400–WP_202319402) are also present. Growth through sulfate reduction is a distinctive characteristic among species of the genus *Archaeoglobus*. *Archaeoglobus veneficus* and *A. infectus* were reported to be incapable to grow with sulfate as an electron acceptor (Huber et al., 1997; Mori et al., 2008). Meanwhile, the genome of *A. veneficus* contains a complete set of genes for dissimilatory sulfate reduction (**Table 2**), and *A. infectus* has *aprA* and *dsrAB* genes. A number of microorganisms did not demonstrate the growth due to dissimilatory sulfate reduction, despite the presence of complete set of genes for this pathway in their genomes (Finster et al., 2013; Slobodkin and Slobodkina, 2019). The reason of this is unclear; probably, special cultivation conditions are required.

In genome of the type species of the genus *Archaeoglobus*, several molybdopterin-binding oxidoreductases have been found, including a gene cluster (AF0174–AF0176) to which the function of a nitrate reductase has been assigned (Klenk et al., 1997; Richardson et al., 2001). Later, growth related to perchlorate reduction was demonstrated in this strain, and a two-stage model was proposed for this pathway (Liebensteiner et al., 2013). Based on genomic and proteomic evidence, it has been shown that perchlorate is reduced to chlorite by the molybdopterin-binding oxidoreductase encoded by the gene cluster AF0174–AF0176. Tetrathionate reductase (AF0157−AF0159) and an enzyme coded by the genes AF2384–AF2386 have been proposed as candidates for the second step of perchlorate reduction. Later transcriptome analysis showed upregulation of genes AF2384−AF2386 during growth with thiosulfate *vs* sulfate indicating that corresponding enzyme is a thiosulfate reductase (Hocking et al., 2014). In the genome of strain SE56^T, we found a gene cluster (WP_202320693–WP_202320695) coding for proteins very similar to subunits of the perchlorate reductase of *A. fulgidus* (AF0174–AF0176). The amino acid similarity of the catalytic subunit (AF0176) to its homolog in strain SE56^T (WP_202320693) was 79%. However, the genome of SE56^T does not have close homologs to the genes of clusters AF0157–AF0159 and AF2384–AF2386, which could explain the inability of the strain to grow by (per)chlorate reduction. We did not find the close homologs to genes AF0157–AF0159 and AF2384–AF2386 in the genomes of the other *Archaeoglobi* representatives (**Figure 4**). Aside from the strain SE56^T, we found close homologs to the catalytic subunit AF0176 (79% amino acids similarity) in the genomes of *Ferroglobus placidus* DSM 10642^T (Ferp_0124) and *Geoglobus ahangari* 234^T (GAH_01285). In both genomes, they are part of clusters coding for subunits of molybdopterin oxidoreductases. In *G. ahangari*, this cluster (GAH_01285–01288) has been previously designated as a nitrate reductase NarGHIJ (Manzella et al., 2015). The genome of *F. placidus* additionally contains another cluster, referenced Ferp_0311–0314, encoding three subunits (NarGHI) of the Nar-type respiratory nitrate reductase and the chaperone NarJ, which could explain why nitrate respiration could be demonstrated experimentally (Hafenbradl et al., 1996; Anderson et al., 2011). Our analysis showed that the alpha subunit of the nitrate reductase has the highest similarity with the nitrate reductase subunit alpha of *Nitrospirae* bacterium (71% amino acids similarity) and *Deltaproteobacteria* (67–68%). The other *Archaeoglobi* do not have close homologs to this enzyme and are unable to carry out dissimilatory nitrate reduction. The ability of *F. placidus* to reduce nitrate seems to have been acquired by horizontal transfer and to correspond to a quite recent evolutionary event.

Motility and Chemotaxis

Strain SE56^T is motile and has a complete set of genes involved in proper assembly and function of flagella. The flagella

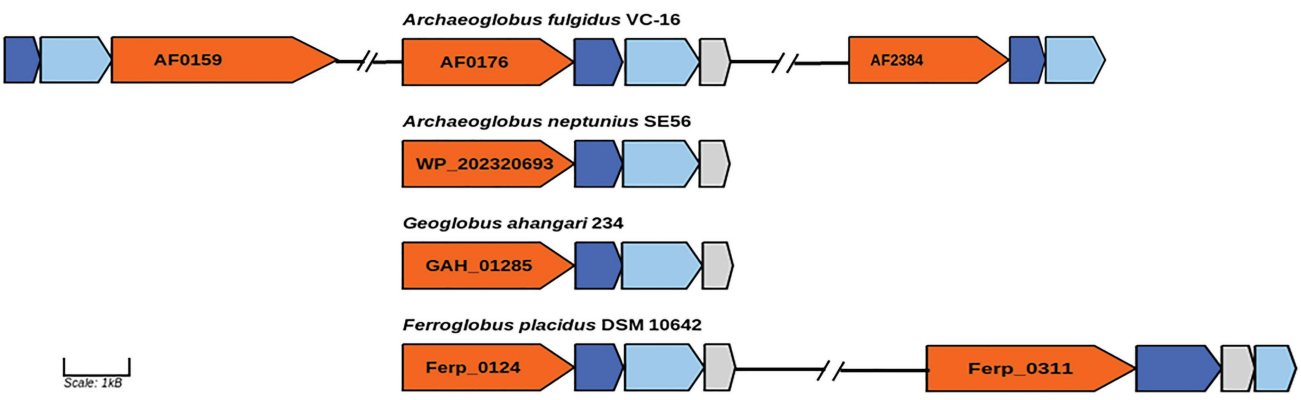


FIGURE 4 | Genes of molybdopterin oxidoreductases probably involved in perchlorate and nitrate reduction. Orange, molybdopterin oxidoreductase molybdopterin binding subunit; blue, molybdopterin-containing oxidoreductase family iron–sulfur binding subunit; light blue, molybdopterin-containing oxidoreductase family membrane subunit; and gray: molybdenum cofactor insertion protein. GenBank accession numbers of catalytic subunits of each enzyme are shown.

apparatus operon *flaB1-2B1-1D/EGFHIJ* (WP_202320707–WP_202320713) found in the genome of strain SE56^T is characteristic of *Euryarcheota* (Albers and Jarrell, 2015). The gene encoding preflagellin peptidase (FlaK; WP_202320792), essential for the flagella assembly, is not in the same cluster in the assembly used for this study. The isolate also has the chemotaxis system to direct the movement towards more favorable location, including an operon *cheWFYBACDR* (WP_202319867–WP_202319871) and genes encoding homologs of several methyl-accepting chemotaxis (MCP) proteins. Unlike *A. fulgidus*, *A. profundus* and *A. veneficus*, flagella- and chemotaxis-coding genes do not cluster together in the strain SE56^T. Among members of the genus *Archaeoglobus*, *A. profundus* and *A. sulfaticallidus* have been reported to be non-motile (Burggraf et al., 1990; Steinsbu et al., 2010). Our analysis of genome data of *A. sulfaticallidus* revealed the presence of a complete *fla* genes cluster as well as putative *flaK* gene (AGK60851–AGK60859; AGK61636). However, it lacks the chemotaxis system *che* genes that might explain the unobserved motility (**Table 2**). In contrast, *A. profundus* possesses both *fla* and *che* genes clusters which are virtually identical in content, order and orientation to corresponding genes of the motile *Archaeoglobi*. This discrepancy between genotype and phenotype has been noted earlier, and it was assumed that *A. profundus* can be motile in some specific conditions, or additional unknown factors are needed for motility (von Jan et al., 2010).

Phylogenomic Analysis

The 16S rRNA gene-based phylogenetic analysis showed that strain SE56^T belongs to the genus *Archaeoglobus*. Further clarification of the taxonomic position of strain SE56^T was carried out using genome-based methods, ANI, AF, dDDH and AAI. At the time of writing, all genome sequences of *Archaeoglobales* type strains were available in public databases, with the exception of the type strain of *A. infectus*. The properties and statistics of genome sequences are presented in **Table 3**. The pairwise OrthoANIu value between genomes of strain SE56^T and its closest relative organism, *A. fulgidus* VC-16^T (NC_000917.1), was 73.49%. The OrthoANIu values with other available species were below 68.5%. Similar results were obtained by means of ANIb and MiSI algorithms (**Table 3**). The AF value for strain SE56^T and *A. fulgidus* VC-16^T was 0.62, whereas these values for strain SE56^T and other representatives of the order *Archaeoglobales* were 0.21–0.30 (**Table 3**). The predicted *in silico* DNA–DNA hybridization (dDDH) values between strain SE56^T and members of the order *Archaeoglobales* varied from 31.7% (*A. profundus* DSM 5631^T) to 18.6% (*A. fulgidus* VC-16^T). All these values are lower than the proposed threshold criteria for prokaryotic species delineation (95–96% ANI, 0.6–0.7 AF and 70% dDDH; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Varghese et al., 2015).

Additionally, we studied the taxonomic position of strain SE56^T based on the alignment of a concatenated set of 122 single-copy partial amino acid sequences of conserved archaeal proteins. The phylogenomic reconstruction confirmed the position of strain SE56^T as a novel species within the genus

TABLE 3 | General genomic features and digital DNA–DNA hybridization (dDDH), average nucleotide identity (ANI), AAI, and alignment fraction (AF) values of strain *Archaeoglobus neptunius* SE56^T and the type strains of the related species.

	A. neptunius SE56^T	*A. fulgidus* VC-16^T	*A. profundus* AV18^T	*A. veneficus* SNP6^T	*A. sulfaticallidus* PM70-1^T	*G. ahangari* 234^T	*G. acetivorans* SBH6^T	*F. placidus* DSM 10642^T
Genome size (Mb)	2.12	2.18	1.56	1.9	2.08	1.77	1.86	2.2
G + C content (mol%)	46.0	48.6	42.0	47.0	43.2	53.1	46.8	44.1
CRISPR count	2	3	0	2	1	7	5	1
Number of genes	2,471	2,524	1846	2,180	2,304	2082	2,251	2,594
Protein coding genes	2,386	2,398	1764	2094	2,199	2001	2,177	2,487
Number of RNAs	52	51	52	51	56	51	53	54
dDDH (%)	100	18.6	31.7	26.8	23.3	22.2	20.4	23.7
ANIb (%)	100	72.58	67.06	67.94	68.0	67.43	67.8	67.19
OrthoANIu (%)	100	73.49	68.11	68.21	68.30	68.4	68.51	68.18
gANI (%)	100	74.03	68.52	68.92	69.26	69.13	69.14	68.62
AF	1.00	0.62	0.21	0.27	0.29	0.30	0.26	0.27
AAI (%)	100	74.88	57.63	59.23	59.0	58.91	57.86	57.84
Genbank ID	NZ_JAEKIW010000000	AE000782	CP001857	CP002588	CP005290	CP011267	CP009552	CP001899

Archaeoglobus. At the same time, it clearly indicated a polyphyletic origin of the species currently included into the genus *Archaeoglobus* (**Figure 5**). The AAI value between strain SE56^T and its closest relative, *A. fulgidus* VC-16^T, was 74.88%. The AAI values between the other members of the genus *Archaeoglobus* were in the range of 57–60% (**Figure 6**). Corresponding ANI and AF values were in the range of 68–70% and 0.21–0.29, respectively (**Supplementary Tables S2** and **S3**). At present, the AAI thresholds proposed for taxonomic delineation are 45–65% for the same family, 65–95% for the same genus and 95–100% for the same species (Konstantinidis et al., 2017). Other proposed genus demarcation criteria are 73.34–74.62% ANI and 0.308–0.354 AF (Barco et al., 2020). On this basis, strain SE56^T and *A. fulgidus* represent different species of one genus, while each of the other three known species of *Archaeoglobus* represents an individual genus, distinct from the first and distinct from each other. Among the other members of the class *Archaeoglobi*, only *G. ahangari* and *G. acetivorans* have AAI, ANI and AF values above the proposed genus threshold, that confirms the placement of these species into the same genus. Except strain SE56^T and *A. fulgidus*, as well as *Geoglobus* species, the AAI, ANI and AF values between all species of the *Archaeoglobi* class were in the range of 56–64%, 68–71%, and 0.21–0.39, respectively (**Figure 6**; **Supplementary Tables S2** and **S3**), proving that they belong to different genera of the same family.

We therefore conclude, that, according to phylogenomic differentiation supported by ANI, AF and AAI values, the genus *Archaeoglobus* should be divided into four distinct genera. The genus name *Archaeoglobus* should be retained for the monophyletic group containing the type species, *A. fulgidus*. We propose to consider the strain SE56^T as a new species of the genus *Archaeoglobus*, *Archaeoglobus neptunius* sp. nov. All four proposed genera belong to the family *Archaeoglobaceae* and share the main morphological, physiological and metabolic features. Representatives of these genera are hyperthermophilic strict anaerobes with respiratory type of metabolism that use sulfur compounds as the terminal electron acceptors. All known species were isolated from marine ecosystems. In addition to the phylogenetic distance, each of the proposed genera has the distinctive phenotypic properties (**Tables 1** and **2**). The genus *Archaeoglobus sensu stricto* differs from the other three genera in its ability to utilize fatty acids provided by the presence of all the genes necessary for β-oxidation of fatty acids. A key distinguishing feature of the genus represented by *A. profundus* is the inability to grow lithoautotrophically caused by the incompleteness of Wood-Ljungdahl pathway. The genus represented by

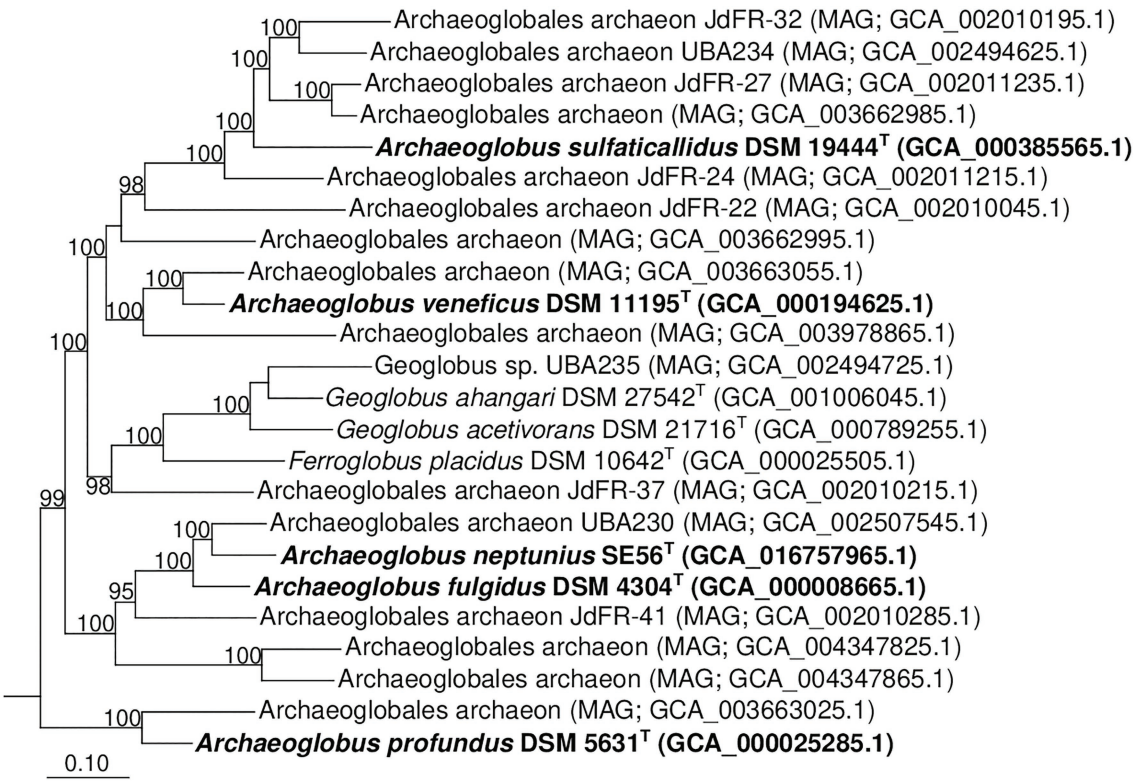


FIGURE 5 | Phylogenomic placement of *A. neptunius* and other species of family *Archaeoglobaceae* based on concatenated partial amino acid sequences of 122 archaeal conservative proteins [Parks et al., 2018; taxonomic designations correspond with Genome Taxonomy DataBase (GTDB)]. The tree was built using the PhyML 3.0 program (Guindon et al., 2010). For rooting, the tree sequences from 62 genomes of type strains of type species of genera of "*Halobacteriota*" phylum (correspond with GTDB) were taken. All assemblages of *Archaeoglobaceae* family according to GTDB Release 05-RS95 are shown on the tree. Bootstrap values above 90% are shown at the nodes. Bar, 0.10 changes per position.

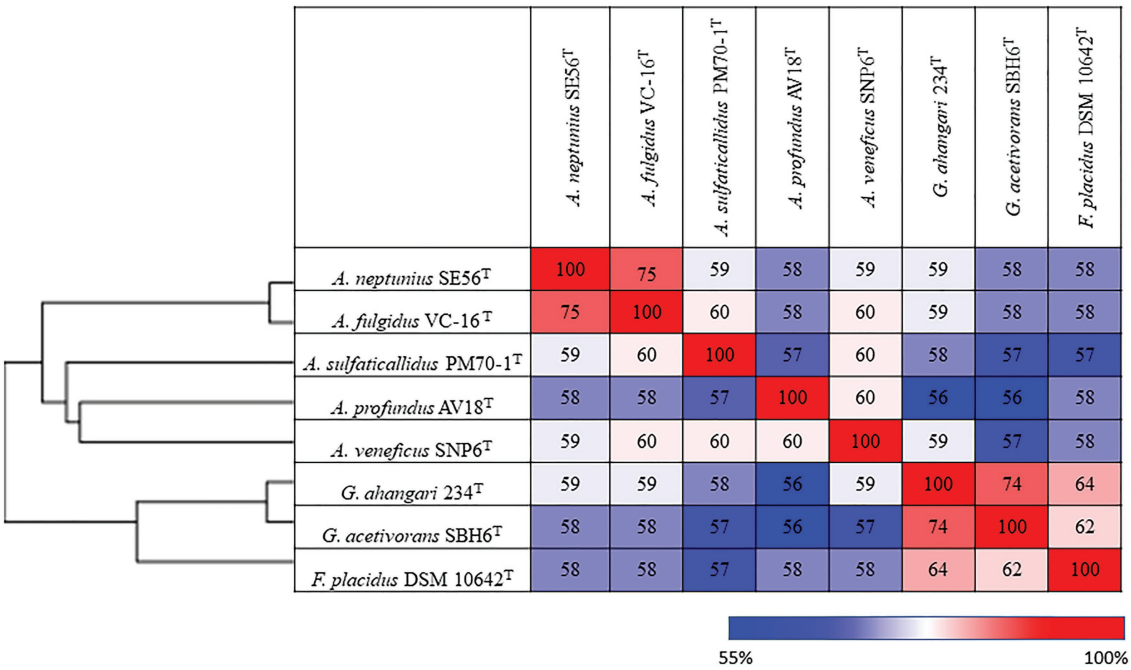


FIGURE 6 | Pairwise average amino acid identity (AAI) determined for members of the class *Archaeglobi*.

A. sulfaticallidus can use sulfate as an electron acceptor for lithoautotrophic growth with molecular hydrogen, whereas members of other genera can perform dissimilatory sulfate reduction only with organic carbon source. In addition, cells of *A. sulfaticallidus* are non-motile, lack genes encoding chemotaxis and have lower optimum growth temperature. The genus encompassing *A. veneficus* is distinguished by its inability to reduce sulfate although it has all the necessary genes. Further work, including the sequencing of the genome of *A. infectus* and core- and pan-genome analysis is required for the formal proposal of the reclassification of the genus *Archaeoglobus*.

Description of *Archaeoglobus neptunius* sp. nov.

Archaeoglobus neptunius (nep.tu'ni.us L. masc. adj. *neptunius*, pertaining to Neptunius (Roman god of the sea), referring to the habitat of the microorganism).

Cells are Gram-negative irregular motile cocci of 0.6–0.8 μm in diameter usually occurring singly. Blue-green fluorescence under UV light. Strictly anaerobic. Growth occurs at 50–85°C (optimum at 80°C), pH 5.5–7.5 (optimum pH 6.5) and NaCl concentrations of 1.5–4.5% [w/v; optimum 2.0–2.5% (w/v) NaCl]. With sulfite or thiosulfate as an electron acceptor, it grows chemolithoautotrophically with molecular hydrogen as an electron donor and CO_2/HCO_3^- as a carbon source. Chemoorganotrophic growth on acetate, propionate, butyrate, fumarate, pyruvate, succinate, peptone, tryptone and yeast extract with sulfate, sulfite or thiosulfate as electron acceptor. Growth is inhibited by elemental sulfur. Formate, methanol, ethanol, iso-propanol, glycerol, lactate, malate, citrate, fructose, glucose, sucrose, maltose, xylose, starch and alginate are not utilized. Elemental sulfur, fumarate, nitrate, nitrite, chlorate, perchlorate and Fe(III) are not used as electron acceptors.

The type strain is SE56^T (=DSM 110954^T = VKM B-3474^T), isolated from a deep-sea hydrothermal vent of the Mid-Atlantic Ridge (TAG vent field: 26°13'69"N 44°82'61"W, 3,625 m water depth). The genome consists of a 2,115,826 bp chromosome with an overall G + C content of 46.0 mol%.

GenBank/EMBL/DDBJ accession (16S rRNA gene): MW531883.

GenBank/EMBL/DDBJ accession (genome): NZ_JAEKIW010000000.

DATA AVAILABILITY STATEMENT

The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAEKIW000000000. The version described in this paper is version JAEKIW010000000. The BioSample data is available in the NCBI BioSample database (http://www.ncbi.nlm.nih.gov/biosample/) under the accession number SAMN17126929. The BioProject data is available in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) with BioProject ID: PRJNA686870. The 16S rRNA gene sequence of *A. neptunius* SE56T (retrieved from the whole-genome sequencing) has been deposited in GenBank/EMBL under accession number MW531883. The strain is available in the German Collection of Microorganisms and Cell Cultures (DSMZ) and All-Russian Collection of Microorganisms (VKM)

under the accession numbers DSM 110954T and VKM B-3474T.

AUTHOR CONTRIBUTIONS

AS, GS, and MJ: conceptualization. GS, MA, AM, and KA: investigation. M-AC-B: resources. GS: writing – original draft preparation. AS, KA, AM, M-AC-B, and MA: writing – review and editing. GS and AM: visualization. AS and MJ: supervision and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Russian Foundation for Basic Research, grant number 18-54-15008, by the Ministry of Science and Higher Education of the Russian Federation, by the French–Russian collaborative project (CNRS/RFBR) Neptune (PRC Russie 2017 n°281295), and by the IRP 1211 MicrobSea.

ACKNOWLEDGMENTS

The authors thank the captains and crews of the R/V *Pourquoi pas?* and the Nautile submersible team for their efficiency, as well as the chief scientists and scientific parties of the BICOSE 2 cruise (http://dx.doi.org/10.17600/18000004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.679245/full#supplementary-material

REFERENCES

Albers, S. V., and Jarrell, K. F. (2015). The archaellum: how *Archaea* swim. *Front. Microbiol.* 6:23. doi: 10.3389/fmicb.2015.00023

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

Anderson, I., Risso, C., Holmes, D., Lucas, S., Copeland, A., Lapidus, A., et al. (2011). Complete genome sequence of *Ferroglobus placidus* AEDII12DO. *Stand. Genomic Sci.* 5, 50–60. doi: 10.4056/sigs.2225018

Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* 60, 685–699. doi: 10.1093/sysbio/syr041

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of *Archaea*. *Nat. Microbiol.* 5, 887–900. doi: 10.1038/s41564-020-0715-z

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

Barco, R. A., Garrity, G. M., Scott, J. J., Amend, J. P., Nealson, K. H., and Emerson, D. (2020). A genus definition for *Bacteria* and *Archaea* based on a standard genome relatedness index. *MBio* 11, e02475–e02419. doi: 10.1128/mBio.02475-19

Beeder, J., Nilsen, R. K., Rosnes, J. T., Torsvik, T., and Lien, T. (1994). *Archaeoglobus fulgidus* isolated from hot North Sea oil field waters. *Appl. Environ. Microbiol.* 60, 1227–1231. doi: 10.1128/aem.60.4.1227-1231.1994

Benson, D. A., Boguski, M. S., Lipman, D. J., Oullette, B. F. F., Rapp, B. A., and Wheeler, D. L. (1999). GenBank. *Nucleic Acids Res.* 27, 12–17. doi: 10.1093/nar/27.1.12

Berg, I. A. (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl. Environ. Microbiol.* 77, 1925–1936. doi: 10.1128/AEM.02473-10

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343

Birkeland, N. K., Schönheit, P., Poghosyan, L., Fiebig, A., and Klenk, H. P. (2017). Complete genome sequence analysis of *Archaeoglobus fulgidus* strain 7324 (DSM 8774), a hyperthermophilic archaeal sulfate reducer from a North Sea oil field. *Stand. Genomic Sci.* 12:79. doi: 10.1186/s40793-017-0296-5

Burggraf, S., Jannasch, H. W., Nicolaus, B., and Stetter, K. O. (1990). *Archaeoglobus profundus* sp. nov., represents a new species within the sulfur-reducing *Archaebacteria*. *Syst. Appl. Microbiol.* 13, 24–28. doi: 10.1016/S0723-2020(11)80176-1

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, I.-M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, 666–677. doi: 10.1093/nar/gky901

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Finster, K. W., Kjeldsen, K. U., Kube, M., Reinhardt, R., Mussmann, M., Amann, R., et al. (2013). Complete genome sequence of *Desulfocapsa sulfexigens*, a marine deltaproteobacterium specialized in disproportionating inorganic sulfur compounds. *Stand. Genomic Sci.* 8, 58–68. doi: 10.4056/sigs.3777412

Fuchs, G., and Stupperich, E. (1978). Evidence for an incomplete reductive carboxylic acid cycle in Methanobacterium thermoautotrophicum. *Arch. Microbiol.* 118, 121–125. doi: 10.1007/BF00406084

Gerhardt, P., Murray, R. G. E., Wood, W. A., and Krieg, N. R. (1994). *Methods for General and Molecular Bacteriology*. Washington, DC: American Society for Microbiology.

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

Grant, J. R., and Stothard, P. (2008). The cgview server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36, W181–W184. doi: 10.1093/nar/gkn179

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

Hafenbradl, D., Keller, M., Dirmeier, R., Rachel, R., Roßnagel, P., Burggraf, S., et al. (1996). *Ferroglobus placidus* gen. Nov., sp. nov., a novel hyperthermophilic archaeum that oxidizes Fe^{2+} at neutral pH under anoxic conditions. *Arch. Microbiol.* 166, 308–314. doi: 10.1007/s002030050388

Harrison, K. J., de Crécy-Lagard, V., and Zallot, R. (2018). Gene graphics: a genomic neighborhood data visualization web application. *Bioinformatics* 34, 1406–1408. doi: 10.1093/bioinformatics/btx793

Henstra, A. M., Dijkema, C., and Stams, A. J. M. (2007). *Archaeoglobus fulgidus* couples CO oxidation to sulfate reduction and acetogenesis with transient formate accumulation. *Environ. Microbiol.* 9, 1836–1841. doi: 10.1111/j.1462-2920.2007.01306.x

Hocking, W. P., Stokke, R., Roalkvam, I., and Steen, I. H. (2014). Identification of key components in the energy metabolism of the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus* by transcriptome analyses. *Front. Microbiol.* 5:95. doi: 10.3389/fmicb.2014.00095

Huber, H., Jannasch, H., Rachel, R., Fuchs, T., and Stetter, K. O. (1997). *Archaeoglobus veneficus* sp. nov., a novel facultative chemolithoautotrophic hyperthermophilic sulfite reducer, isolated from abyssal black smokers. *Syst. Appl. Microbiol.* 20, 374–380. doi: 10.1016/S0723-2020(97)80005-7

Huerta-Cepas, J., Forslund, S. 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

Kashefi, K., Tor, J. M., Holmes, D. E., Gaw Van Praagh, C. V., Reysenbach, A.-L., and Lovley, D. R. (2002). *Geoglobus ahangari* gen. nov., sp. nov., a novel hyperthermophilic archaeon capable of oxidizing organic acids and growing autotrophically on hydrogen with Fe(III) serving as the sole electron acceptor. *Int. J. Syst. Evol. Microbiol.* 52, 719–728. doi: 10.1099/00207713-52-3-719

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

Khelifi, N., Amin Ali, O., Roche, P., Grossi, V., Brochier-Armanet, C., Valette, O., et al. (2014). Anaerobic oxidation of long-chain n-alkanes by the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus*. *ISME J.* 8, 3057–3060. doi: 10.1038/ismej.2014.58

Khelifi, N., Grossi, V., Hamdi, M., Dolla, A., Tholozan, J. L., Ollivier, B., et al. (2010). Anaerobic oxidation of fatty acids and alkenes by the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Appl. Environ. Microbiol.* 76, 3057–3060. doi: 10.1128/AEM.02810-09

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., et al. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390, 364–370. doi: 10.1038/37052

Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi: 10.1038/ismej.2017.113

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/molbe v/msw054

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

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Liebensteiner, M. G., Pinkse, M. W. H., Schaap, P. J., Stams, A. J. M., and Lomans, B. P. (2013). Archaeal (per)chlorate reduction at high temperature: an interplay of biotic and abiotic reactions. *Science* 340, 85–87. doi: 10.1126/science.1233957

Makarova, K. R., and Koonin, E. V. (2003). Filling a gap in the central metabolism of archaea: prediction of a novel aconitase by comparative-genomic analysis. *FEMS Microbiol. Lett.* 227, 17–23. doi: 10.1016/S0378-1097(03)00596-2

Manzella, M. P., Holmes, D. E., Rocheleau, J. M., Chung, A., Reguera, G., and Kashefi, K. (2015). The complete genome sequence and emendation of the hyperthermophilic, obligate iron-reducing archaeon *"Geoglobus ahangari"* strain 234^{T}. *Stand. Genomic Sci.* 10:77. doi: 10.1186/s40793-015-0035-8

Mardanov, A. V., Slobodkina, G. B., Slobodkin, A. I., Beletsky, A. V., Gavrilov, S. N., Kublanov, I. V., et al. (2015). The *Geoglobus acetivorans* genome: Fe(III) reduction, acetate utilization, autotrophic growth, and degradation of aromatic compounds in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 81, 1003–1012. doi: 10.1128/AEM.02705-14

Martin, W. F., and Sousa, F. L. (2016). Early microbial evolution: the age of anaerobes. *Cold Spring Harb. Perspect. Biol.* 8:a018127. doi: 10.1101/cshperspect.a018127

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 Bioinform.* 14:60. doi: 10.1186/1471-2105-14-60

Mori, K., Maruyama, A., Urabe, T., Suzuki, K., and Hanada, S. (2008). *Archaeoglobus infectus* sp. nov., a novel thermophilic, chemolithoheterotrophic archaeon isolated from a deep-sea rock collected at Suiyo Seamount, Izu-Bonin Arc, western Pacific Ocean. *Int. J. Syst. Evol. Microbiol.* 58, 810–816. doi: 10.1099/ijs.0.65422-0

Offre, P., Spang, A., and Schleper, C. (2013). *Archaea* in biogeochemical cycles. *Annu. Rev. Microbiol.* 67, 437–457. doi: 10.1146/annurev-micro-092412-155614

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

Palombarini, F., Di Fabio, E., Boffi, A., Macone, A., and Bonamore, A. (2020). Ferritin nanocages for protein delivery to tumor cells. *Molecules* 25:825. doi: 10.3390/molecules25040825

Park, C., and Park, W. (2018). Survival and energy producing strategies of alkane degraders under extreme conditions and their biotechnological potential. *Front. Microbiol.* 9:1081. doi: 10.3389/fmicb.2018.01081

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, R., Skarshewski, A., Chaumeil, P.-A., et al. (2018). Standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

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

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Rabus, R., Venceslau, S. S., Wöhlbrand, L., Voordouw, G., Wall, J. D., and Pereira, I. A. (2015). A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes. *Adv. Microb. Physiol.* 66, 55–321. doi: 10.1016/bs.ampbs.2015.05.002

Richardson, D. J., Berks, B. C., Russell, D. A., Spiro, S., and Taylor, C. J. (2001). Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. *Cell. Mol. Life Sci.* 58, 165–178. doi: 10.1007/PL00000845

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., Glöckner, F. O., 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

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

Shieh, J. S., and Whitman, W. B. (1987). Pathway of acetate assimilation in autotrophic and heterotrophic methanococci. *J. Bacteriol.* 169, 5327–5329. doi: 10.1128/jb.169.11.5327-5329.1987

Slobodkin, A. I., Reysenbach, A. L., Slobodkina, G. B., Baslerov, R. V., Kostrikina, N. A., Wagner, I. D., et al. (2012). *Thermosulfurimonas dismutans* gen. Nov., sp. nov., an extremely thermophilic sulfur-disproportionating bacterium from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 62, 2565–2571. doi: 10.1099/ijs.0.034397-0

Slobodkin, A. I., and Slobodkina, G. B. (2019). Diversity of sulfur-disproportionating microorganisms. *Microbiology* 88, 509–522. doi: 10.1134/S0026261719050138

Slobodkina, G. B., Baslerov, R. V., Novikov, A. A., Viryasov, M. B., Bonch-Osmolovskaya, E. A., and Slobodkin, A. I. (2016). *Inmirania thermothiophila* gen. Nov., sp. nov., a thermophilic, facultatively autotrophic, sulfur-oxidizing gammaproteobacterium isolated from shallow-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 66, 701–706. doi: 10.1099/ijsem.0.000773

Slobodkina, G. B., Kolganova, T. V., Querellou, J., Bonch-Osmolovskaya, E. A., and Slobodkin, A. I. (2009). *Geoglobus acetivorans* sp. nov., an iron(III)-reducing archaeon from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 59, 2880–2883. doi: 10.1099/ijs.0.011080-0

Søndergaard, D., Pedersen, C. N. S., and Greening, C. (2016). HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* 6:34212. doi: 10.1038/srep34212

Steinsbu, B. O., Thorseth, I. H., Nakagawa, S., Inagaki, F., Lever, M. A., Engelen, B., et al. (2010). *Archaeoglobus sulfaticallidus* sp. nov., a thermophilic and facultatively lithoautotrophic sulfate-reducer isolated from black rust exposed to hot ridge flank crustal fluids. *Int. J. Syst. Evol. Microbiol.* 60, 2745–2752. doi: 10.1099/ijs.0.016105-0

Stetter, K. O. (1988). *Archaeoglobus fulgidus* gen. Nov., sp. nov.: a new taxon of extremely thermophilic *Archaebacteria*. *Syst. Appl. Microbiol.* 10, 172–173. doi: 10.1016/S0723-2020(88)80032-8

Stetter, K. O. (2006). Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. B* 361, 1837–1843. doi: 10.1098/rstb.2006.1907

Stetter, K. O., Lauerer, G., Thomm, M., and Neuner, A. (1987). Isolation of extremely thermophilic sulfate reducers: evidence for a novel branch of archaebacteria. *Science* 236, 822–824. doi: 10.1126/science.236.4803.822

Stokke, R., Hocking, W. P., Steinsbu, B. O., and Steen, I. H. (2013). Complete genome sequence of the thermophilic and facultatively chemolithoautotrophic sulfate reducer *Archaeoglobus sulfaticallidus* strain PM70-1^{T}. *Genome Announc.* 1, 4–5. doi: 10.1128/genomeA.00406-13

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526. doi: 10.1093/oxfordjournals.molbev.a040023

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Tor, J. M., Kashefi, K., and Lovley, D. R. (2001). Acetate oxidation coupled to Fe(III) reduction in hyperthermophilic microorganisms. *Appl. Environ. Microbiol.* 67, 1363–1365. doi: 10.1128/AEM.67.3.1363-1365.2001

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., et al. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 43, 6761–6771. doi: 10.1093/nar/gkv657

von Jan, M., Lapidus, A., Del Rio, T. G., Copeland, A., Tice, H., Cheng, J. F., et al. (2010). Complete genome sequence of *Archaeoglobus profundus* type strain (AV18). *Stand. Genomic Sci.* 2, 327–346. doi: 10.4056/sigs.942153

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., and Nelson-Sathi, S., et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116. doi: 10.1038/nmicrobiol.2016.116

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090. doi:10.1073/pnas.74.11.5088

Wolin, E. A., Wolin, M. J., and Wolfe, R. S. (1963). Formation of methane by bacterial extracts. *J. Biol. Chem.* 238, 2882–2886. doi: 10.1016/S0021-9258(18)67912-8

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 assemblie. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Zuo, G., Xu, Z., and Hao, B. (2015). Phylogeny and taxonomy of *Archaea*: a comparison of the whole-genome-based CVTree approach with 16S rRNA sequence analysis. *Life* 5, 949–968. doi: 10.3390/life5010949

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

The reviewer TN declared a shared affiliation with one of the authors GS, AM, and AS to the handling editor at the time of the review.

Copyright © 2021 Slobodkina, Allioux, Merkel, Cambon-Bonavita, Alain, Jebbar and Slobodkin. 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: 27 August 2021
doi: 10.3389/fmicb.2021.701265

Effects of Simulated Microgravity on the Physiology of *Stenotrophomonas maltophilia* and Multiomic Analysis

Xiaolei Su[1,2†], Yinghua Guo[2,3†], Tingzheng Fang[1,2], Xuege Jiang[2], Dapeng Wang[1,2], Diangeng Li[4], Po Bai[5], Bin Zhang[6], Junfeng Wang[2] and Changting Liu[2*]*

[1]Medical School of Chinese PLA, Beijing, China, [2]Department of Respiratory and Critical Care Medicine, The Second Medical Center and National Clinical Research Center for Geriatric Disease, Chinese PLA General Hospital, Beijing, China, [3]College of Pulmonary and Critical Care Medicine, The Eighth Medical Center, Chinese PLA General Hospital, Beijing, China, [4]Department of Academic Research, Beijing Chaoyang Hospital Affiliated to Capital Medical University, Beijing, China, [5]Respiratory Diseases Department, PLA Rocket Force Characteristic Medical Center, Beijing, China, [6]Department of Respiratory and Critical Care Medicine, Binzhou Medical University Hospital, Binzhou, China

OPEN ACCESS

Edited by:
André Antunes, Macau University of Science and Technology, China

Reviewed by:
Khaled Y. Kamal, Texas A&M University, United States
Song Lin Chua, Hong Kong Polytechnic University, Hong Kong, SAR China

***Correspondence:**
Junfeng Wang
wangjf72@hotmail.com
Changting Liu
changtingliu1202@sohu.com

†These authors have contributed equally to this work

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

Received: *27 April 2021*
Accepted: *29 July 2021*
Published: *27 August 2021*

Citation:
Su X, Guo Y, Fang T, Jiang X, Wang D, Li D, Bai P, Zhang B, Wang J and Liu C (2021) Effects of Simulated Microgravity on the Physiology of Stenotrophomonas maltophilia and Multiomic Analysis. Front. Microbiol. 12:701265. doi: 10.3389/fmicb.2021.701265

Many studies have shown that the space environment plays a pivotal role in changing the characteristics of conditional pathogens, especially their pathogenicity and virulence. However, *Stenotrophomonas maltophilia*, a type of conditional pathogen that has shown to a gradual increase in clinical morbidity in recent years, has rarely been reported for its impact in space. In this study, *S. maltophilia* was exposed to a simulated microgravity (SMG) environment in high-aspect ratio rotating-wall vessel bioreactors for 14 days, while the control group was exposed to the same bioreactors in a normal gravity (NG) environment. Then, combined phenotypic, genomic, transcriptomic, and proteomic analyses were conducted to compare the influence of the SMG and NG on *S. maltophilia*. The results showed that *S. maltophilia* in simulated microgravity displayed an increased growth rate, enhanced biofilm formation ability, increased swimming motility, and metabolic alterations compared with those of *S. maltophilia* in normal gravity and the original strain of *S. maltophilia*. Clusters of Orthologous Groups (COG) annotation analysis indicated that the increased growth rate might be related to the upregulation of differentially expressed genes (DEGs) involved in energy metabolism and conversion, secondary metabolite biosynthesis, transport and catabolism, intracellular trafficking, secretion, and vesicular transport. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses showed that the increased motility might be associated the upregulation of differentially expressed proteins (DEPs) involved in locomotion, localization, biological adhesion, and binding, in accordance with the upregulated DEGs in cell motility according to COG classification, including pilP, pilM, flgE, flgG, and ronN. Additionally, the increased biofilm formation ability might be associated with the upregulation of DEPs involved in biofilm formation, the bacterial secretion system, biological adhesion, and cell adhesion, which were shown to be regulated by the differentially expressed genes (chpB, chpC, rpoN, pilA, pilG, pilH, and pilJ) through the integration of transcriptomic and proteomic analyses. These results suggested that simulated microgravity might increase the level of corresponding functional proteins by upregulating related genes to alter

physiological characteristics and modulate growth rate, motility, biofilm formation, and metabolism. In conclusion, this study is the first general analysis of the phenotypic, genomic, transcriptomic, and proteomic changes in *S. maltophilia* under simulated microgravity and provides some suggestions for future studies of space microbiology.

Keywords: simulated microgravity, *Stenotrophomonas maltophilia*, physiology, biofilm, mobility, multiomic analysis

INTRODUCTION

In recent decades, with the development of medicine and space technology, microbial space safety has become one of the hot spots in modern medical research. The space environment involves microgravity, high pressure, low temperature, cosmic radiation, and malnourished environments (Horneck et al., 2010). However, microbes can still survive under such extreme conditions (Venkateswaran et al., 2014). Some studies have found microbes on surfaces, air filters and even in air samples from the International Space Station (ISS), and the diversity and abundance of microbes have been investigated (Checinska et al., 2015). *Staphylococcus*, *Enterococcus faecalis*, *Bacillus*, *Propionibacterium*, and *Corynebacterium* have been identified as the most common species. Studies have showed that the space environment has some impacts on the immune system and motor system of humans, as well as on microorganisms in astronauts, such as growth rate, drug susceptibility, carbon source utilization and chemical sensitivity, biofilm formation ability, and motility (Taylor, 2015; Nadeau et al., 2016; Vaishampayan and Grohmann, 2019; Zhang et al., 2019). However, the impact of the space environment on microbes varies on different bacteria. Assuring that astronauts are sound in body and mind and have a good working capability is challenging. During spaceflight, the spaceship provides insulation from cosmic radiation and extreme temperature to some extent. Therefore, microgravity is a key influencing factor on astronauts and microbes.

Given the constraint on spaceflight missions and the uncertainties in outer space experiments, land-based, high-aspect rotating vessels (HARVs) have been developed by NASA (Wolf et al., 1992). HARVs simulate microgravity by rotating cultures on an axis perpendicular to gravity, and the weight of microbes can be affected by low shear forces, which can suspend bacteria in a constant free fall state (Radtke and Herbst-Kralovetz, 2012). Over the decades, HARVs have been widely used to explore the physiological states of microorganisms in a simulated microgravity (SMG) environment (Nickerson et al., 2000; Wilson et al., 2002a,b; Lynch et al., 2004; Lawal et al., 2010; Castro et al., 2011; Rasamiravaka and El Jaziri, 2016). In addition, HARVs provide a first-rank suspension growth environment that imitates some of the aspects of microgravity, such as a reduction in fluid shear, minimal turbulent motion, and short sedimentation of bacteria or host cells (Hammond and Hammond, 2001). Some experts have verified the effectiveness of HARVs through mathematical analyses (Lynch et al., 2006). In addition, the reduction in shear stress within HARVs promotes the formation of 3D tissue aggregates (Sawyer et al., 2008).

Stenotrophomonas maltophilia (*S. maltophilia*), which was first isolated in 1943, is an emerging Gram-negative multidrug-resistant organism (MDRO) that is highly relevant to respiratory infections in humans (Hugh et al., 1963). *Stenotrophomonas maltophilia* is an important nosocomial pathogen that is closely related to infections, particularly in immunocompromised people. Early research has suggested that *S. maltophilia*-associated infections are more common in conjunction with severe pneumonia, chronic obstructive pulmonary disease (COPD), bacteremia, biliary sepsis, bone and joint infections, eye infections, endocarditis, and meningitis (Papadakis et al., 1995; Fujita et al., 1996; Landrum et al., 2005; Lai et al., 2006; Takigawa et al., 2008; Yemisen et al., 2008). Currently, research on *S. maltophilia* mainly focuses on its drug resistance and clinical anti-infection treatments (Sanchez et al., 2009; Amin and Waters, 2016). However, the impact of the space environment or SMG on *S. maltophilia* has not been thoroughly investigated, and few research has explored whether and how the space environment or SMG changes its growth rate, phenotypic characteristics, antibiotic susceptibility, carbon source utilization and chemical sensitivity, biofilm formation ability, and motility. In this study, we studied the phenotypic variations of *S. maltophilia* and used combined genomic, transcriptomic, and proteomic analyses to reveal the impact of SMG on *S. maltophilia*. Our research results may provide new ideas for the prevention and treatment of infections in future spaceflight.

MATERIALS AND METHODS

Bacterial Strains and Culture Conditions

The original *S. maltophilia* strain, used in this study, was purchased from the China General Microbiological Culture Collection Center. The bacterial strain was grown at 37°C in Luria-Bertani (LB) liquid medium. The original strain of *S. maltophilia* (SMO) was cultured in a shake tube containing LB liquid medium at 37°C. *Stenotrophomonas maltophilia* in simulated microgravity (SMS) and *S. maltophilia* in normal gravity (SMN) were established by growing bacteria in continuous cultivation in high-aspect rotating vessel (HARV) bioreactors (Synthecon, Inc. Houston, Tex, United States). **Figure 1** shows the procedure of the SMG cultivation implemented by rotating the bioreactor on its axis perpendicular to gravity, while the normal gravity (NG) was implemented



FIGURE 1 | HARV bioreactors in the experimental setup. HARV system used to generate simulated microgravity (SMG) and normal gravity (NG) conditions.

by rotating the bioreactors on its axis parallel to gravity. The overnight culture of the original strain was adjusted to a turbidity of 1.0 and then inoculated at a dilution of 1:200 in HARV bioreactors. Bubbles in the reactor were carefully removed through the valves at both ends. After incubation in HARVs for 24 h at 37°C with a rotation of 25 rpm, the SMS and SMN bacterial suspensions were poured out and diluted to a turbidity of 1.0, added to new HARVs in a volume ratio of 1:200 filled with LB medium, and incubated at 37°C and 25 rpm for another 24 h. This step of bacterial inoculation in each group was repeated 14 times for 2 weeks (Wang et al., 2016b; Morrison and Nicholson, 2020).

Scanning Electron Microscopy

Morphological changes were also observed by scanning electron microscopy (SEM), and the three groups were washed three times with sterile phosphate-buffered saline (PBS; pH 7.4) and fixed with 2.5% glutaraldehyde at 4°C overnight. The samples were washed again with deionized water three times, and each solution was settled for 7 min and then dehydrated in a gradient series of ethanol (50, 70, 85, and 90%) for 15 min per step. The samples were subsequently immersed in 100% ethanol (three times for 10 min each) to prevent drying, and then, all the dehydrated samples were placed in a vacuum desiccator to critical point-dry. After coating the specimens with gold-palladium, the samples were observed using cold field emission scanning electron microscope (Hitachi SU8010, Japan; Haddad et al., 2021).

Growth Curves

Stenotrophomonas maltophilia in simulated microgravity and SMN were taken from HARVs every 2 h. Samples [10^6 colony forming unit (CFU)/ml] were inoculated into 96-well honeycomb microtiter plates, and SMO was also prepared at the same concentration. For each group, one 96-well plate was prepared with 200 µl of solution. The optical density at 600 nm was measured every 2 h for 24 h using Bioscreen C system and BioLink software (Bioscreen C, Finland). This experiment was performed in three replicates (Chopra et al., 2006; Zhang et al., 2019).

Carbon Source Utilization and Chemical Sensitivity Assay

Biolog GEN III plates (Biolog, Hayward, CA, United States) were used to test the ability of *S. maltophilia* to utilize 71 carbon sources and 23 chemical sensitivities in triplicate experiments. The test was performed according to the manufacturer's instructions. Briefly, we used a sterile cotton-tipped swab to remove a bacterial colony from the surface of a BUG1B agar plate, and then inoculated it into IF-A inoculum. The concentration of the bacterial colony was adjusted to 10^8 CFU/ml by a turbidimeter (Zhang et al., 2019). A 100 µl volume of inoculated IF-A was added to each of the 96 wells in the microplate. After sealing and incubating at 37°C for 24 h, "negative" (−), "positive" (+), and "weakly positive" (w) were scored according to visual similarity. The endpoint data of each well were measured according to the absorbance 590 nm in a microplate reader. This experiment was performed in three replicates.

Drug Susceptibility Testing

Drug susceptibility testing (DST) was performed using the disk diffusion method based on the CLSI M100 Ed30 document (CLSI, 2020): trimethoprim-sulfamethoxazole (SXT, 25 µg), levofloxacin (LEV, 5 µg), and minocycline (MNO, 30 µg). The entire surface of the Petri dish containing Mueller-Hinton agar (MHA) was covered with a specific inoculum (10^7–10^8 CFU/ml; Zhao et al., 2019), and then, the plate was dried for 15 min. The diameter of the zone of inhibition was measured after incubation for 18 h at 37°C. All drug susceptibility tests were performed in three replicates.

Qualitative Analysis of the Biofilm Formation Ability

After 2 weeks of cultivation under SMG and NG conditions, HARVs were stained with crystal violet to evaluate the biofilm formation ability of *S. maltophilia*. Briefly, after 2 weeks, the bacterial suspension was removed from the HARV bioreactors, washed gently with sterilized water twice and then stained with 0.1% crystal violet for 15 min at room temperature. Subsequently, the crystal violet dye was slowly removed through a sterile syringe, the reactor was washed with sterile deionized water three times, and then, pictures were taken (Wang et al., 2017).

Quantitative Analysis of the Biofilm Formation Ability

The SMS and SMN samples, which were cultivated for 14 days in the previous step of the qualitative measurement, were collected and diluted to 10^7 CFU/ml. The SMO sample was also diluted to 10^7 CFU/ml (Zhao et al., 2019). The three samples were diluted with LB medium at a volume ratio of 1:100, and added to a 24-well plate. Each well contained 1 ml LB medium. Then, the 24-well plate was sealed and fixed with rubber bands and placed in an oscillation incubator at 37°C and 200 rpm for 24 h. Next, planktonic bacteria were removed. After washing each well with deionized water twice, the plate was placed in a 50°C constant temperature incubator to dry for 15 min to fix the biofilm. About 2 ml of 0.1% crystal violet dye was added to each well at room temperature for 15 min to stain the adherent biofilm. Then, crystal violet was removed, and the biofilm was slowly washed twice with sterile deionized water, followed by the addition of 2 ml DMSO to fully digest the biofilm. Finally, the suspension was transferred to a 96-well plate to measure OD_{570}. The results are presented as the mean ± SD of three biological replicates. In addition, the SMS and SMN samples obtained from the cultivation after 14 days in HARV bioreactors were diluted to 10^7 CFU/ml and immediately subjected to CFU assay to identify the differences in viability (Shui et al., 2021).

Analysis of the Biofilm Formation Ability *via* Confocal Laser-Scanning Microscopy

The biofilm formation ability of the three groups was assessed using Filmtracer™ (LIVE/DEAD Biofilm Viability Kit, L10316, Molecular Probes, Invitrogen, CA, United States) in conjunction with confocal laser-scanning microscopy (CLSM; Leica TCS SP8, Berlin, Germany). LIVE/DEAD Biofilm Viability is a culture-independent method that depends on the cell membrane integrity as a viability measure; it utilizes a mixture of the SYTO9 green fluorescent nucleic acid stain and a red fluorescent nucleic acid stain and propidium iodide. Live and dead bacteria can be distinguished without waiting for growth plate results. The working solution of the fluorescent stains was prepared by adding 3 µl of the SYTO9 stain and 3 µl of the propidium iodide stain to 1 ml filter-sterilized water. The excitation/emission maxima for these dyes are approximately 480/500 nm for SYTO9 staining and 490/635 nm for propidium iodide. Three independent rounds of biofilm experiments were performed. Briefly, in each round of experiments, SMS, SMN, and SMO were inoculated in 35-mm confocal dishes (Solarbio, Beijing, China), cultivated at 37°C for 24 h, and then rinsed two times with sterile water to eliminate nonadherent bacteria. Then, the generated biofilms were carefully stained with 200 µl working solution stain for 30 min in the dark at room temperature. It was important to immediately add the stain before the biofilm dried. After removing the stain and washing two times with sterile water, images were acquired from random positions, and the number of images from each layer varied according to the thickness of the biofilm. The maximum biofilm thickness of the uniform position in the current field of view was set as the biofilm thickness (Jiang et al., 2021; Shui et al., 2021).

Motility Assay

To assess the motility of *S. maltophilia*, the SMS and SMN samples were taken from HARV bioreactors immediately after 14 days of SMG and NG cultivation. The difference in swimming motility was assessed using agar plates with slight modifications (Murray and Kazmierczak, 2006). Briefly, to monitor swimming, *S. maltophilia* diluted to 10^7 CFU/ml was point inoculated with a sterile inoculation needle onto plates containing 0.1% (w/v) Bacto agar, 10 g/L Tryptone (w/v), and 5 g/L Yeast extract (w/v). The inoculated samples were incubated for 48 h at 37°C, the plates were imaged, and the diameter of the zone of growth around the spot inoculation was measured. All the motility tests were performed in three replicates.

Genome Sequencing

Genome Sequencing and Assembly

DNA from the original strain of *S. maltophilia* was acquired using the conventional phenol–chloroform extraction method (Javadi et al., 2014). After the genomic DNA was segmented by Covaris, a genomic sequencing library was constructed. The whole genome was sequenced using the Illumina Hiseq and PacBio platforms according to standard quality controls, and raw reads were obtained. The raw reads were then filtered using the following steps: the adapter sequence was removed, bases other than A, G, C, and T at the 5'-end were removed, bases containing "N" reaches above 10% were removed; the ends of reads with lower sequencing quality (<Q 20) were trimmed; and the adapters and quality-trimmed small fragments less than 25 bp in length were removed. Self-correction was carried out using Falcon software (Version 0.3.0), and corrected reads were acquired. Quality assessment of the genome sequence obtained from the preliminary assembly determined whether the genome was contaminated, assessed the quality of sequencing, and assessed the condition of the genome. The bacterial genome scan map was assembled using SOAPdenove2 software, and multiple k-mer parameters were spliced after the sequences were optimized. Then, the optimal contig assembly results were obtained. The reads were compared to the contigs to form scaffolds according to the relationship between paired-ends and overlaps. The map of the complete bacterial genome was assembled using the canu and SPAdes. Overlaps of a certain

length at both ends of the final assembly sequence were cut off at one end. Finally, the complete chromosome sequence was obtained (Lu et al., 2019; Jiang et al., 2020; **Table 1**).

Genome Component Prediction

The genome component prediction included the prediction of coding genes, transfer RNA (tRNA), ribosome RNA (rRNA), and tandem repeat sequences. The detailed steps were performed as follows: (a) the related coding genes were predicted by Glimmer, GeneMarks and Prodigal software; (b) tRNA genes and rRNA genes were recognized by tRNAscan-SE software (Version V2.0) and Barrnap software, respectively. The nucleotide sequence information, anticodon information, and secondary structure information of the tRNA in each sample genome were obtained; and (c) Tandem repeat annotations were retrieved by the Tandem Repeats Finder. The type, location, and sequence information of all the rRNAs in the genome of each sample were obtained (Cao et al., 2020).

Gene Function Analysis

Six databases containing NR (Non-Redundant Protein Database), SwissProt, Pfam, COG (Clusters of Orthologous Groups), GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) were used for general gene function annotation. A whole-genome BLAST alignment was constructed using the six databases.

Comparative Genomic Analysis

The genomes of SMS and SMN were sequenced using high-throughput sequencing Illumina technology (Chi et al., 2021; Hao et al., 2021). In brief, DNA was extracted from samples of SMS and SMN and randomly fragmented, and DNA fragments of a certain length were recovered by electrophoresis. Then, DNA fragments assembled with adapters were amplified to prepare DNA clusters, and the clusters were sequenced (Gao et al., 2020; Zhang et al., 2021).

The sequences obtained using fastq software (Version 0.20.0) were filtered by removing reads that contained adapters and poly-N as well as those with a lower sequencing quality (<Q 20). The original SMO group was used as a reference. The SMS and SMN reads were mapped to that of the reference sequence using the mem algorithm of BWA software, and the coverage was determined using Picard-tools. Then, according to the comparison bam results, the sequencing depth and relative coverage were acquired. GATK was used to realign the reads near InDels to obtain the realigned bam, which could eliminate false positive single-nucleotide polymorphisms (SNPs). Then, snippy software (Version 4.6.0) was used to detect SNPs and small InDels. Finally, snpEff was used to annotate variant sites to determine the impact of mutations on the genome (Jiang et al., 2020).

Transcriptome Sequencing and Comparative Transcriptomic Analysis

Sequencing and Filtering

RNA was isolated immediately from bacterial cells incubated in semisolid LB medium. About 2 μg RNA was obtained from the two groups using a TruSeqTM Stranded Total RNA Library Prep Kit (Illumina, United States) following the manufacturer's instructions. Around 2 μg of RNA from each sample was used for RNA sample preparation after the concentration and purity of the RNA was detected by a Nanodrop 2020. After removing rRNAs, the purified mRNAs were fragmented into small pieces (~200 bp) using fragmentation buffer. The RNA fragments were used to produce first strain cDNA by reverse transcriptase PCR with random primers. Then, second strain cDNA was created using dUTP instead of dTTP. The sticky end of the double strand cDNA was linked to the adaptor and digested with the UNG enzyme. Then, the cDNA fragments were enriched *via* PCR amplification and quantified by TBS380 (Picogreen), and the clusters were sequenced using the Illumunia Hiseq platform (An et al., 2020).

Gene Expression Statistics and Differential Gene Expression Analysis

The quantitative expression software RSEM was used to quantitatively analyze the expression levels of genes and transcripts. The transcripts per million (TPM) and reads per kilobase million (FPKM) of each gene were calculated in accordance with gene length, and the regulation mechanism of genes were revealed by combining the sequence function information. According to the expression of genes in different samples, Venn, correlation and principal component analysis (PCA) analyses were performed. Venn analysis was used to display genes that were shared and uniquely expressed between samples. Correlation analysis of the biological replicate samples, on the one hand, checked whether the variation between biological replicates met the expectations of the experimental design, and on the other hand, provided a basic reference for the analysis of differentially expressed genes (DEGs). PCA could reduce the complexity of the data and determine the relationship and variation between samples. Finally, the read counts were mapped to the genes. Differential gene expression analysis was analyzed using DESeq2 software, and genes yielding a value of $p < 0.05$ and fold change (FC) >2 in two different samples were identified as DEGs (Zhang et al., 2020).

Functional Annotation and Enrichment Analysis

The identified DEGs were annotated on the basis of COG, GO, and KEGG functional annotation. The gene lists were enriched according to GO and KEGG enrichment analyses.

TABLE 1 | The assembly data of SMO.

Sequence type	Sequence topology	Sequence number	Plasmid No.	Total length (bp)	GC content (%)
Chromosome	Circular	1	0	4,779,890	66.37

GO enrichment analysis of the DEGs was carried out using Goatools software. GO terms with a value of $p < 0.05$ were identified as significantly enriched by DEGs. R scrip was used to estimate the meaningful statistical enrichment of the DEGs in KEGG pathways (Wang et al., 2020).

Proteomics Analysis

Polypeptide Preparation

Frozen bacterial samples stored in −80°C were transferred to shaking tubes, and lysis buffer was added to the samples and mixed well (Kan et al., 2020). A high-throughput tissue grinder was used and shaken three times for 40 s each time. The suspension was placed on ice for 30 min and sonicated every 5 min for 10 s. After incubating at 100°C for 10 min, the samples were refrigerated on ice, and centrifuged at 12,000 *g* for 20 min at 4°C. Precooled acetone was added to the supernatant at a ratio of 1:4 and precipitated overnight at −20°C. The next day, the mixture was centrifuged at 12,000 *g* for 20 min, and then, the supernatant was discarded. Next, 90% precooled acetone was added to the pellet and mixed well. This step was repeated twice. Then, the protein lysis solution was added to the precipitate, and after centrifugation at 12,000 *g* for 20 min at 4°C, the supernatant was obtained. The proteins were quantified by BCA and subsequently confirmed using SD-PAGE. After reductive alkylation and enzymatic hydrolysis, the processed peptides were labeled with Tandem Mass Tag™ (TMT™), reconstituted with UPLC loading buffer and subjected to high-pH liquid fractionation with a reversed-phase C18 column (Thermo, United States). Then, the purified peptides were subjected to an EASY-nLC 1,200 instrument (Thermo, United States) and Q_Exactive HF-X (Thermo, United States) to obtain raw data. The raw data were converted into MGF format by the Thermo Scientific tool Proteome Discoverer software (Version 2.4), and the exported MGF files were searched against the established database using the Sequest and Mascot modules. Eventually, analyses of the data and bioinformatics were performed according to the previous search results (Dai et al., 2020; Chen et al., 2021; Ye et al., 2021).

Functional Annotation and Differential Protein Analysis

According to the mass spectrometry results, all proteins and sequences were compared with information in databases, including GO, KEGG, COG, and subcellular location related databases, to obtain the annotation information of the proteins in each database and statistics on the annotations. Proteins with a fold change >1.2 or <0.83 and value of $p < 0.05$ in two different groups were defined as differentially expressed proteins (DEPs). Additional protein lists were assembled based on the DEPs, including GO, KEGG, and COG functional annotation and GO and KEGG pathway enrichment analyses.

STATISTICAL ANALYSIS

Differences in the growth rates among SMS, SMN, and SMO were evaluated by two-way ANOVA, followed by Tukey's multiple comparisons test. The biofilm formation ability and motility were evaluated by unpaired *t*-tests. GraphPad Prism (version 8.4.3) was used for statistical analysis. Differences with values of $p < 0.05$ were considered significant.

RESULTS

Morphology

Scanning electron microscopy was performed to observe the single cell morphology of SMO, SMN, and SMS. From observing the single bacteria, most of the three groups were found to be long rod shape. Under 5,000x SEM, the three groups showed entanglement and aggregation, and some shorter bacilli appeared sporadically. Under 20,000x SEM, it was found that SMO was full and smooth, but there were unequal mucus-like substances in SMN and SMS, and SMS had more mucus-like substances than SMN (**Figure 2**).

Growth Rate Assay

Growth rate curves for the three groups are shown in **Figure 3**. Compared with the SMN group, the SMS group exhibited a mild increased growth rate, especially after 6 h ($p < 0.0001$), 8 h ($p = 0.0004$), and 10 h ($p = 0.011$), while the SMO showed a similar growth curve to that of the SMN group.

Carbon Source Utilization Assays and Chemical Sensitivity and Drug Susceptibility Testing

Compared with SMO and SMN, SMS showed decreased carbon source utilization of D-arabitol (D3), myo-inositol (D4), D-aspartic acid (D8), L-pyroglutamic acid (E8), quinic acid (F8), and D-lactic acid methyl ester (G3), indicating that the metabolism pathways of SMS were inhibited in the SMG environment. In addition, the chemical sensitivity assay showed decreased tolerance to potassium tellurite (G12; **Table 2**). SMS, SMN, and SMO were all sensitive to SXT, LEV, and MNO. The drug susceptibility test results showed no statistically significant difference in the inhibitive zone diameter of the three antibiotics.

Biofilm Assay

The phenotypes of biofilm formation under both the SMG and NG conditions were analyzed after 2 weeks of cultivation by observing the films that generated and adhered to the gas-permeable membranes at the bottom of the bioreactors. The biofilm formed by *S. maltophilia* grown in the SMG environment were thicker than those grown in the NG environment (**Figure 4A**). The biofilm formation ability of the three groups was quantified by the OD_{570} values of crystal violet staining. The biofilm formation ability of SMS was more than twice that of SMN ($p = 0.0006$; **Figure 4B**). There was no significant difference between SMN and SMO. Besides, colony forming unit analysis of viable counts of bacteria was performed to identify the differences in viability. The result showed that there was

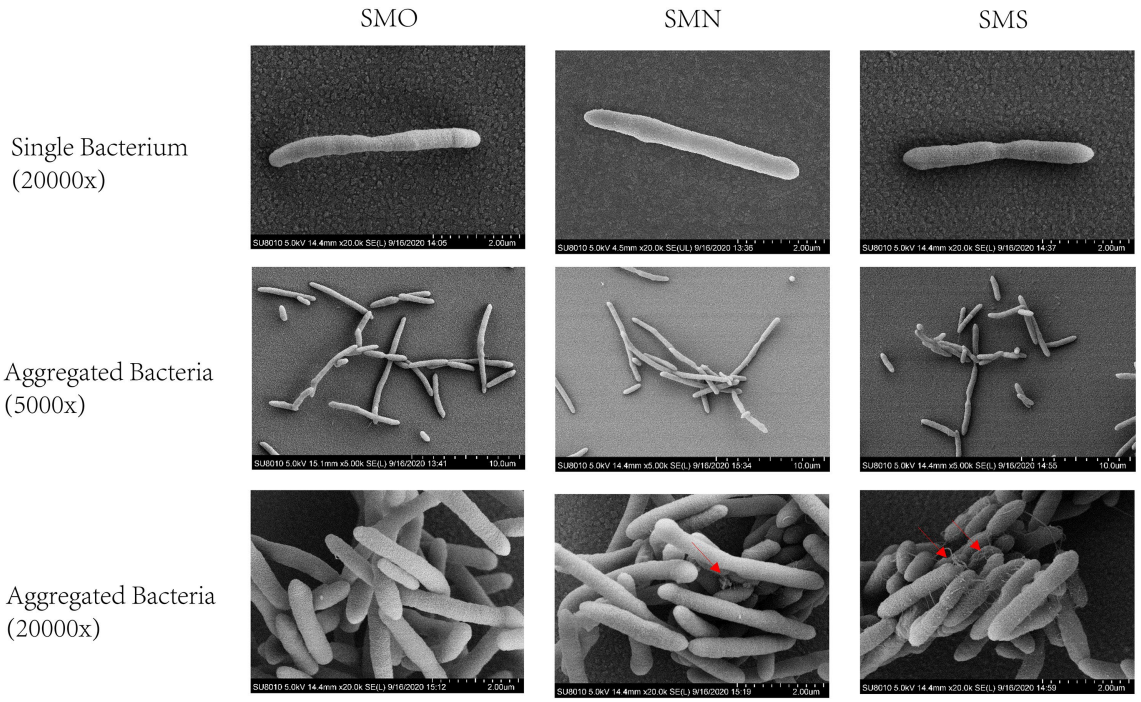


FIGURE 2 | Scanning electron micrographs (SEM) images of *Stenotrophomonas maltophilia*. The red arrow indicates the intercellular mucus-like substances. The **first row** showed the morphology of single bacterium under 20,000x SEM. The **middle row** showed the morphology of aggregated bacteria under 5,000x SEM. The **last row** showed the morphology of aggregated bacteria under 20,000x SEM.

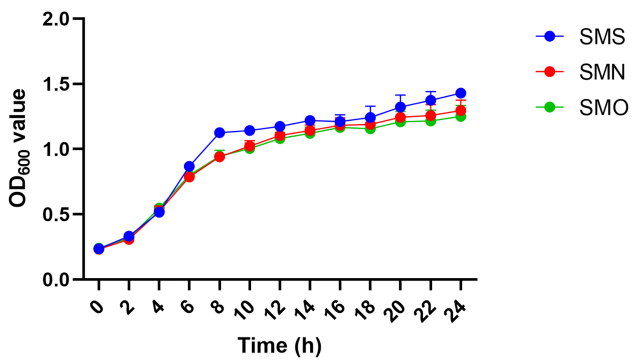


FIGURE 3 | Growth curves of three *S. maltophilia* samples. Growth curves of *Stenotrophomonas maltophilia* in simulated microgravity (SMS; blue), *S. maltophilia* in normal gravity (SMN; red), and original strain of *S. maltophilia* (SMO; green) were determined by measuring the OD_{600} value.

TABLE 2 | Biochemical characteristics of three *Stenotrophomonas maltophilia* samples.

Substrate	SMO	SMN	SMS
D-arabitol (D3)	+	+	w
myo-inositol (D4)	+	+	w
D-aspartic acid (D8)	+	+	w
L-pyroglutamic acid (E8)	+	+	w
quinic acid (F8)	+	+	w
D-lactic acid methyl ester (G3)	+	+	–
potassium tellurite (G12)	+	+	–

"+" represents a positive reaction; "–" represents a negative reaction; "w" represents a weakly positive reaction.

no statistically significant between SMS (97.6 ± 4.4) and SMN (96.3 ± 2.0) samples after 10^6 times dilution ($p = 0.5286$; **Figure 4G**). Furthermore, the biofilm formation ability results were confirmed *via* CLSM analysis. The image stacks obtained from each sample illustrated that the maximum biofilm thickness of SMS (**Figure 4C**) was significantly greater than that of both SMN ($p = 0.0002$; **Figure 4D**) and SMO ($p < 0.0001$; **Figure 4E**). There was no significant difference in biofilm thickness between SMN and SMO (**Figure 4F**).

Mobility

All three groups were able to move on the interface between the bottom of the plate and the medium in our study. On the swimming plates, SMS formed round and translucent colonies with dense cells at the inoculation site. The average colony diameter of SMS was 37.96 ± 1.34 mm ($n = 3$). However, the colony diameters of SMN and SMO were 22.56 ± 0.68 mm ($n = 3$), and 21.96 ± 1.13 mm ($n = 3$) respectively, both of which were smaller than that of SMS, and the difference was statistically significant ($p = 0.0066$, $p = 0.0057$; **Figure 5**).

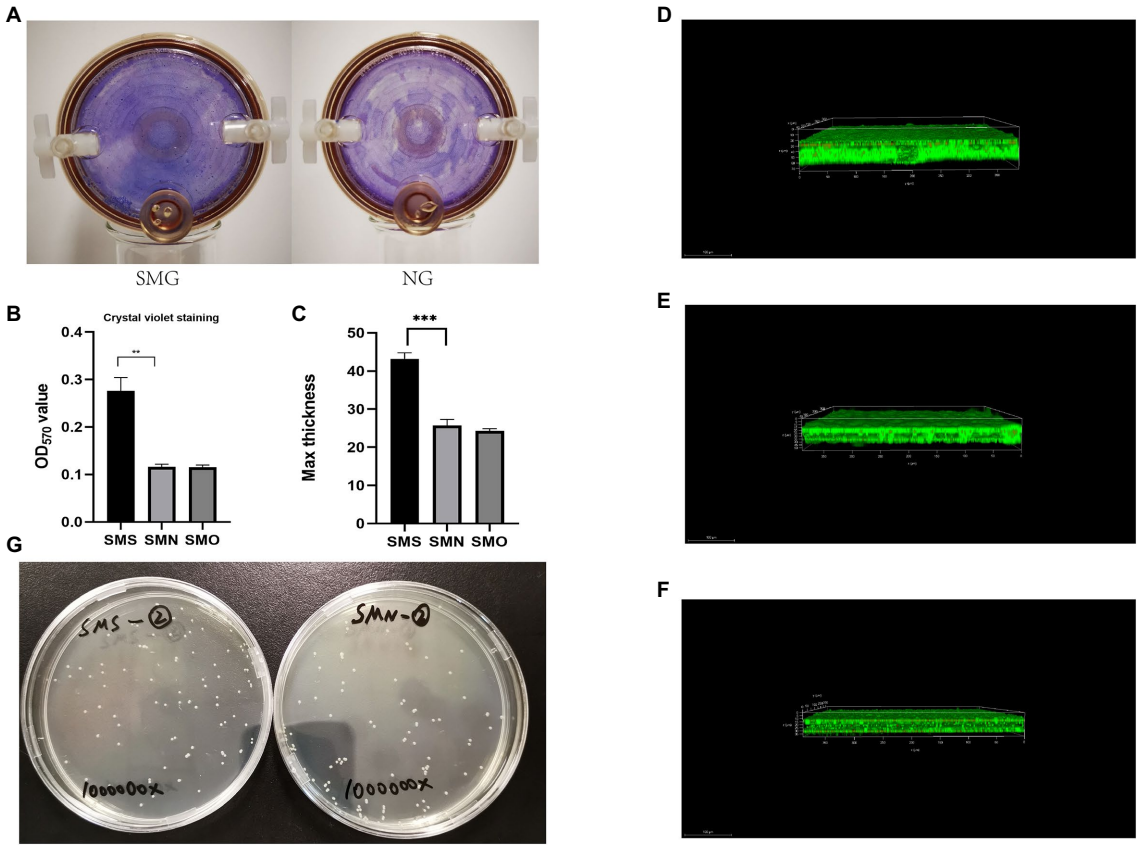


FIGURE 4 | Biofilm formation assay. **(A)** Quantitative analysis of biofilm formation ability *via* crystal violet. The three groups were cultured in 24-well plate at 37°C and 200 rpm for 24 h. Biofilm formation ability was measured by determining the OD_{570} of crystal violet. **(B)** Quantitative analysis of biofilm formation ability *via* crystal violet. The three groups were cultured in 24-well plate at 37°C and 200 rpm for 24 h. Biofilm formation ability was measured by determining the OD570 of crystal violet. **(C)** Biofilms of SMS, SMN, and SMO cultured in 35-mm confocal dished. Cells were stained with a Filmtracer™ (LIVE/DEAD Biofilm Viability Kit), and biofilm formation ability was detected using confocal laser-scanning microscopy (CLSM). Green fluorescence indicates live cells and red fluorescence indicates dead cells. The maximum thickness of biofilm of SMS, SMN, and SMO were (43.17 ± 2.73; 25.75 ± 1.59), and (25.36 ± 1.56) μm, respectively. $^{**}p < 0.01$ and $^{***}p < 0.001$. **(D)** Analysis of biofilm formation ability of SMS samples using CLSM. **(E)** Analysis of biofilm formation ability of SMN samples using CLSM. **(F)** Analysis of biofilm formation ability of SMO samples using CLSM. **(G)** Colony forming unit (CFU) assay of SMS and SMN samples.

Genome Sequencing, Assembly, and Annotation

The accession number of SMO is SRR14180817. The draft genome of SMO was estimated to be 4.78 Mbp with a GC content of 66.37%. In addition, SMO was predicted to have 4,437 coding sequences, 74 tRNAs, 13 sRNA, and seven gene islands. The average length of the coding genes was 941.55 bp, encompassing 87.40% of the genome. More information about the genome is shown in the circular map (**Figure 6**).

Comparative Genomic Analysis

The accession number for the resequencing data of the SMS is SRR14181238, and that for SMN is SRR14181566. SMO was used as a reference strain, and SMN and SMS were mapped to the genome of SMO to estimate genomic changes. One indel was found between SMN and SMO, including 10 related genes in coding regions and one related gene in an intergenic region. One indel was found between SMS and SMO, including 10 related genes in coding regions and one related gene in an intergenic region. In addition, one SNP was found between SMS and SMO, including eight related genes in coding regions and one related gene in an intergenic region. The reference base "G" was mutated to "C," which attributed to transversion mutation, the related upstream and downstream genes were mostly about ribosomal protein and ATP-binding protein. The details of the indels and SNP located in the CDS are presented in **Table 3**.

RNA-Seq Alignment and Comparative Transcriptomic Analysis

The accession number of the transcriptomic data for SMS is SRR14193993, and that for SMN is SRR14183401. The sequencing reads of SMS and SMN were mapped to the reference genome of SMO. The percentages of the total reads for SMS and SMN mapped to the reference strain were approximately 97.85 and

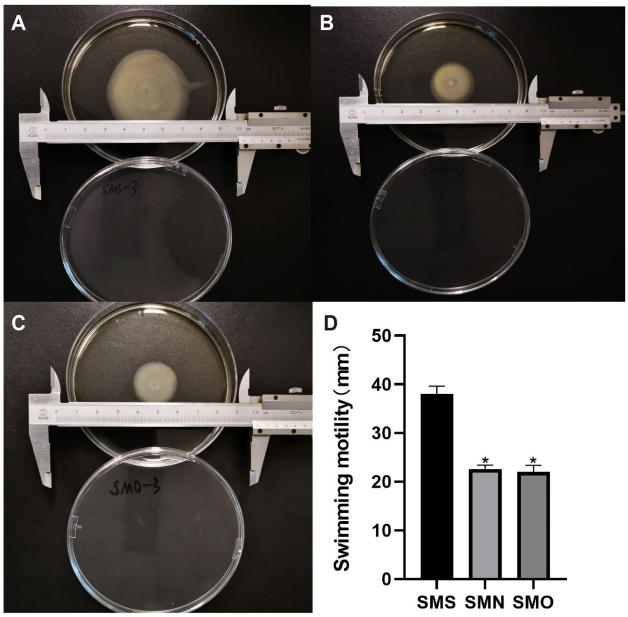


FIGURE 5 | Effect of SMG on mobility of *S. maltophilia*. For the swimming motility, three groups were inoculated on a surface of 0.1% (w/v) Bacto agar plates. Swimming haloes were gauged after 48 h of incubation at 37°C. Zone diameters (means ± SD; $n = 3$) are listed for each group. **(A)** SMS (37.96 ± 1.34 mm), **(B)** SMN (22.56 ± 0.68 mm), and **(C)** SMO (21.96 ± 1.13 mm). **(D)** Asterisks (*) indicate statistically significant change ($p < 0.05$) compared to that of SMS group. *$p < 0.05$.

98.47%, respectively. The uniquely mapped reads of SMS and SMN were 92.90 and 94.68%, respectively.

Overall, 188 DEGs were identified between SMS and SMN according to the TPM reads values. Compared with SMN, SMS contained 133 upregulated and 55 downregulated genes (**Figure 7**). In addition, 115 DEGs were mapped to COG categories, most of which were associated with cell motility, intracellular trafficking, secretion, vesicular transport, energy production and conversion, and inorganic ion transport and metabolism (**Figure 8A**). The pilP, pilM, flgE, flgG, and ronN genes, which are associated with cell motility, were upregulated. The pilA, pilO, pilQ, and ata genes, which are associated with fimbrial protein and outer membrane protein formation, were also upregulated.

Differentially expressed genes were also mapped according to the results of GO function analysis (**Figure 8B**). The DEGs were enriched in cellular components (CCs), molecular function (MF), and biological processes (BPs). For CCs, the DEGs between SMS and SMN were classified into two types. For MF, the DEGs were classified into five types, and six types were included in BPs. A total of 77 genes were annotated in the three categories, most of which were associated with cellular anatomical entities, binding, catalytic activity, and cellular processes. It is worth noting that the same DEGs probably existed in different categories.

In addition, the differentially expressed genes were annotated and enriched in the KEGG pathway database. In total, the 43 DEGs included in the first six categories and 19 secondary biochemical categories or branches of KEGG metabolic pathways, were statistically significant between the SMS and SMN groups, most of which are associated with signal transduction, cellular community, cell motility, and amino acid metabolism (**Figures 9A,B**).

Comparative Proteomic Analysis

The accession number of the proteomic data for SMS and SMN is PXD025290. The proteomic data showed 235 DEPs between SMS and SMN, among which 118 proteins were upregulated and 117 proteins were downregulated (**Figure 10A**). Subsequently, the DEPs were mapped to GO pathways and KEGG pathways.

According to the results of the GO functional analysis, it was clear that almost all the DEPs were enriched in biological processes and cellular components (**Figure 10B**). According to the results of the KEGG pathway analysis, 49 KEGG pathways of 109 DEPs were statistically significantly different between SMS and SMN. Further study indicated that the DEPs were involved in functions such as pilus ($p = 0.00001$), pilus organizations ($p = 0.00012$), cell adhesion ($p = 0.00006$), protein secretion by the type II secretion system ($p = 0.00029$), and protein transport across the cell outer membrane ($p = 0.00029$), all of which were significantly upregulated in SMS compared with SMN (**Figure 10C**).

Integration of Transcriptomic and Proteomic Analyses

To gain an understanding of the effects of SMG on *S. maltophilia* from a systematic biological perspective, the DEGs and DEPs were integrated to identify overlapping genes that were differentially expressed in both the transcriptome and proteome. In total, 30 genes were selected (**Figure 11A**), and genes with either upregulated or downregulated expression at both the mRNA level and protein levels were used for bioinformatic analysis. The GO enrichment analysis indicated that biological processes such as biological adhesion, cell adhesion, cell projection, and pilus organization might be affected in SMS. In addition, cellular components such as intrinsic components of the membrane, integral components of the membrane, membrane parts, and external encapsulating structure parts might also be affected in SMS (**Figure 11B**). The KEGG enrichment analysis indicated that biofilm formation and the two-component system between up-DEGs and up-DEPs might be affected in SMS (**Figure 11C**), which was in agreement with the DEGs enriched in the KEGG pathways at the transcriptomic level. Besides, four genes (chpC, pilJ, pilH, and pilG) involved in biofilm formation and seven genes (rpoN, chpB, chpC, pilA, pilG, pilH, and pilJ) involved in the two-component system were significantly upregulated.

DISCUSSION

The physiological changes and metabolic responses of microorganisms to the extreme environments of space have

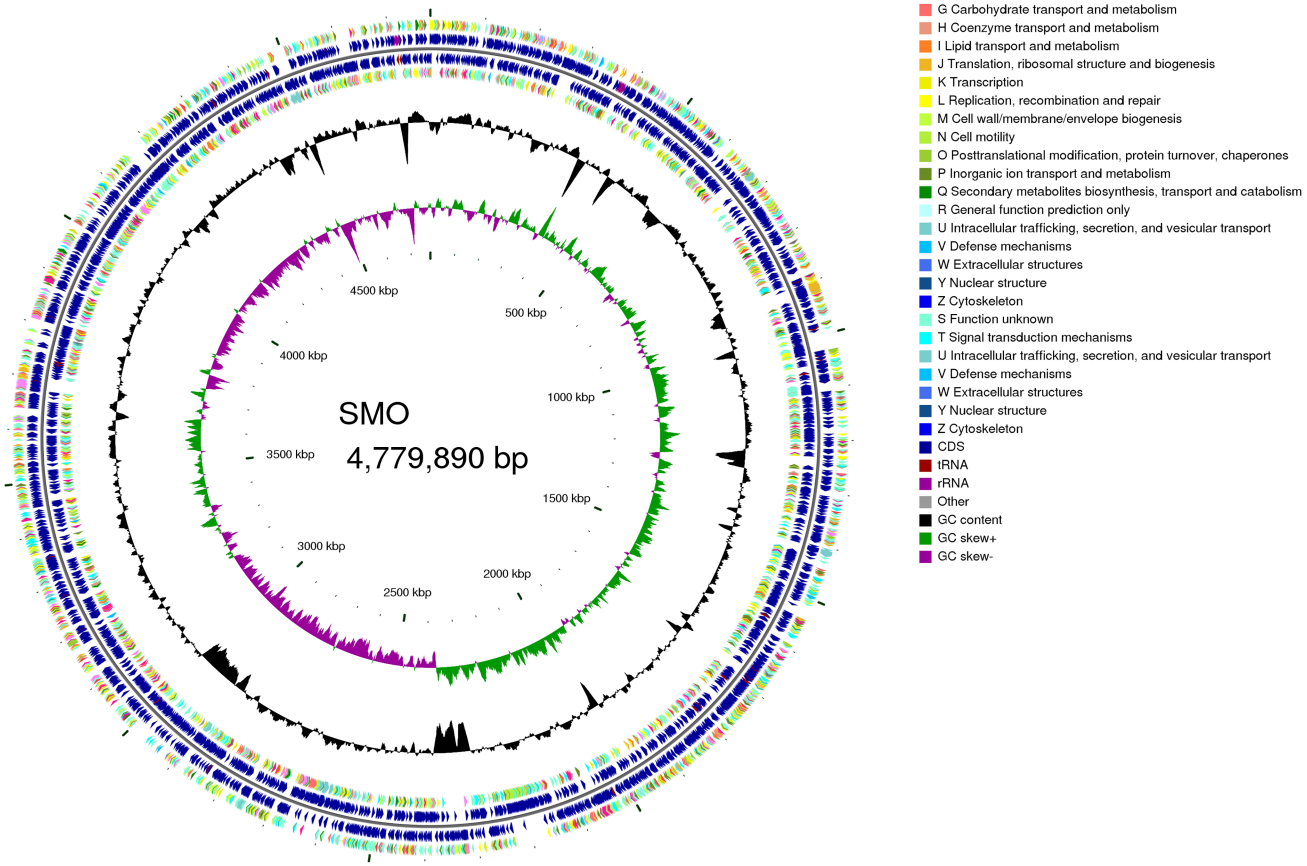


FIGURE 6 | Genome map of SMO. From outer to inner circle: CDS on the positive-strand, different colors indicate different Clusters of Orthologous Groups (COG) function classifications (ring 1), CDS, transfer RNA (tRNA), ribosome RNA (rRNA) on the positive-strand (ring 2), CDS, tRNA, and rRNA on the negative-strand (ring 3), CDS on the negative-strand, different colors indicate different COG function classifications (ring 4), GC content (ring 5), GC-Skew (ring 6).

attracted increasing attention. Monitoring microorganisms in space stations plays an important role in monitoring the safety of spacecraft and the health of astronauts. With the advancement of science and technology, more space and ground-assisted research technologies have helped us better understand the impact of the space environment on microorganisms. In recent years, a large number of studies have reported on the effects of space environment and simulated space environment on the physiological characteristics of microorganisms and the relationship between microorganisms and human beings. However, *S. maltophilia*, as a new opportunistic pathogen that is found in water, soil, humans, and animals, has rarely been studied for its effect on the space environment. *Stenotrophomonas maltophilia* is an important nosocomial pathogen and is closely related to clinical infections, especially in immunocompromised people. Hence, astronauts with low immunity are susceptible to *S. maltophilia* during spaceflight. Currently, there is no integrated genetic information available for *S. maltophilia* after cultivation in the SMG environment, which makes it difficult to explore the physiological changes and metabolic responses of the organism after exposure to this unique condition. Overall, our results confirm that in SMG environment, the physiological characteristics of *S. maltophilia* have undergo some changes, such as a faster growth rate, enhanced biofilm formation ability, and increased mobility. Furthermore, through multiomics analysis, 118 DEPs were found to be upregulated, most of which were related to biofilm formation, pilus organization, bacterial motility, locomotion, and the bacterial secretion system. At the genetic level, 113 DEGs were upregulated. For instance, DEGs such as pilP, pilM, flgE, flgG, and ronN, which are related to cell motility, were significantly upregulated. In addition, the chpB, chpC, rpoN, pilA, pilG, pilH, and pilJ genes, which are related to biofilm formation and cell outer membrane, were also significantly upregulated. These upregulated DEGs were consistent with the upregulated DEPs.

TABLE 3 | Summary of all indels and single-nucleotide polymorphisms (SNPs) located in the CDS of SMS and SMN.

Strain	Type	Gene ID	Annotation
SMS and SMN	Deletion	Gene3316	Peptidase M20
	Deletion	Gene3321	Hypothetical protein
	Deletion	Gene3322	Two-component system response regulator
	Deletion	Gene3313	Drug resistance transporter
	Deletion	Gene3314	NADPH-dependent ferric siderophore reductase
	Deletion	Gene3315	Hypothetical protein
	Deletion	Gene3317	Hypothetical protein
	Deletion	Gene3318	Phosphoethanolamine transferase
	Deletion	Gene3319	Hypothetical protein
	Deletion	Gene3320	Hypothetical protein
SMS	SNP	Gene1305	ABC transporter ATP-binding protein
	SNP	Gene1310	Lipid II flippase MurJ
	SNP	Gene1311	Bifunctional riboflavin kinase/FAD synthetase
	SNP	Gene1312	tRNA ligase
	SNP	Gene1306	50S Ribosomal protein L21
	SNP	Gene1307	50S Ribosomal protein L27
	SNP	Gene1308	Redox-regulated ATPase YchF
	SNP	Gene1309	30S Ribosomal protein S20

Considering the different culture conditions, cultivation times, and strains used, previous study showed that microorganisms exposed to a space environment or cultivated in the SMG environment on the ground exhibited different growth characteristics. Most researchers have shown that microgravity or weightlessness in space can significantly increase the growth rate of bacteria (Benoit and Klaus, 2007; Kim et al., 2013; Javadi et al., 2014; Yim et al., 2020). Compared with ground group, some studies have shown that under microgravity or weightlessness conditions, the growth rate of some bacteria does not change significantly (Coil et al., 2016; Fajardo-Cavazos et al., 2018). A few studies have even shown that the flight strain has a significantly decreased growth rate compared with that the normal gravity group (Fajardo-Cavazos and Nicholson, 2016). Our study illustrated (Fajardo-Cavazos and Nicholson, 2016) that SMS exhibited an increased growth rate compared with that of SMN and SMO. It is worth noting that under SMG or weightlessness, the underlying mechanism of the microbial growth rate is still unclear. Some experts believe that this may be related to the increase in the utilization of nutrients by bacteria in a specific environment, and the corresponding reduction in extracellular material transport (Brown et al., 2002). Through the GO enrichment analysis performed in this research, we found that the increased growth rate of SMS might be related to some upregulated DEPs involved in biological regulation, metabolic process, cellular anatomical entity, and catalytic activity. Moreover, according to the KEGG pathway analysis, some of the upregulated DEPs in amino acid metabolism, energy metabolism, and carbohydrate metabolism were also central to the increased growth rate. In addition, according to the COG annotations, 20 DEPs related to intracellular trafficking, secretion, and vesicular transport were overexpressed in SMS. In conclusion, we believe that the accumulation of these changes may increase the growth rate of SMS.

Astronauts during spaceflight are in a state of reduced immunity, which accompanied by changes in microbiota. Some studies have found that some characteristics of bacteria have a certain correlation with their motility (Portela et al., 2019). The differences in the structure and number of flagella, fimbriae, and pili of bacteria greatly affect their motility. As a key microbial behavior, motility plays a key role in nutrient absorption, tissue localization and invasion, biofilm formation, and pathogenicity. Changes in motility may cause changes in the bacterial distribution in the human body, and these motility changes may damage the health of astronauts. Therefore, the effect of microgravity on bacteria can be observed and explained by the change in bacterial motility to a certain extent. Although, domestic and foreign researchers have studied the influence of microgravity on microbial motility, the exact way in which microgravity causes these changes is still unclear. A reasonable explanation suggests that bacteria are affected by microgravity on account of the quiescent fluid environment enclosing bacteria in a liquid suspension culture, and another explanation is that the buoyant convection of low dense fluid in the vicinity of cell plays a key role (Tirumalai et al., 2017). Portela et al. (2019) found that after spaceflight, the morphology of *Escherichia coli* changed under Transmission Electron Microscope (TEM). Compared with the normal gravity group, the samples after spaceflight showed a decreased size and an increased number of outer membrane vesicles. Benoit and Klaus (2007) revealed that under the same spaceflight conditions, motile bacteria displayed greater cell numbers than nonmotile bacteria. Microbial motility is meditated by an electromotive gradient of ions across the cell membrane, so the disparity between each concentration and nutrient depletion produced by microgravity and SMG may have led to the difference in motility (DeRosier, 1998). Wilson et al. (2007) revealed that *Salmonella Typhimurium* cultivated in LB in HARVs showed a decrease in motility; downregulations of fliEST, flhD, and flgM; and upregulation

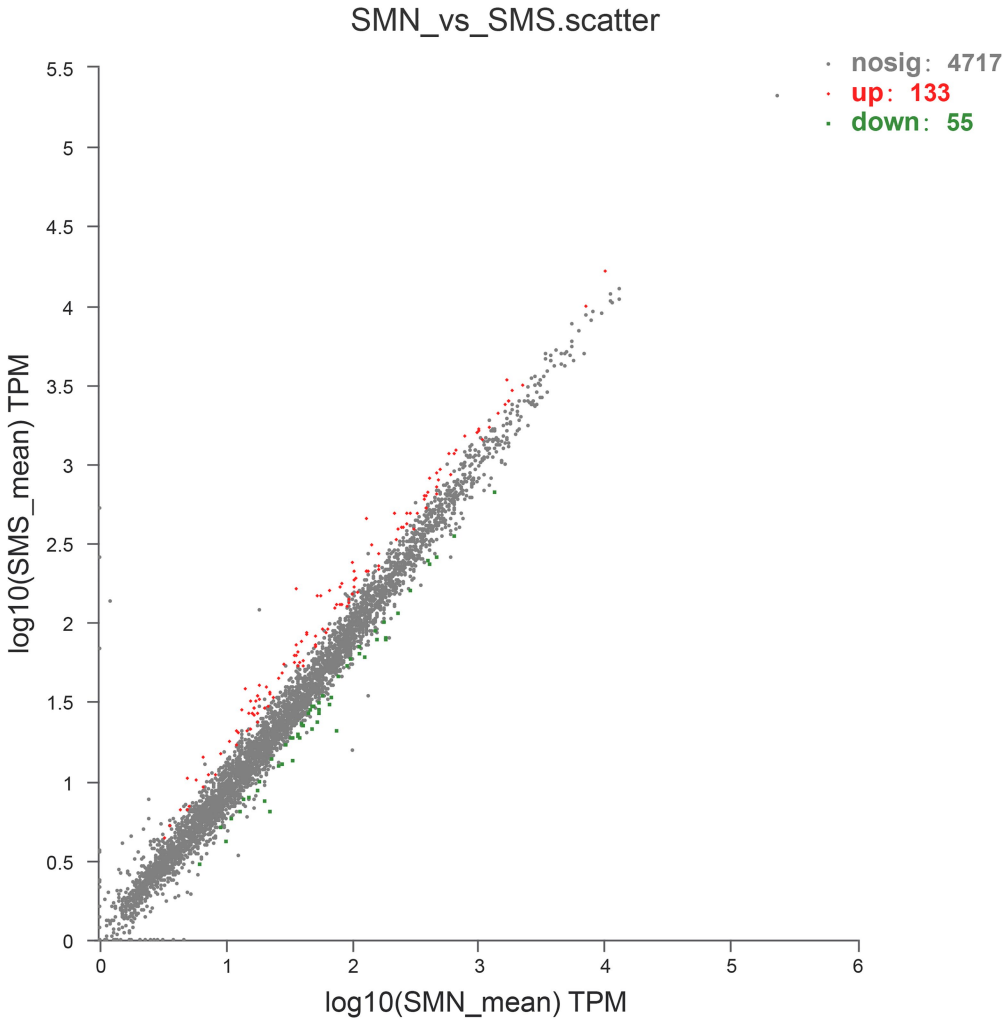


FIGURE 7 | Upregulated and downregulated genes expressed in SMS compared with that in SMN. The values of the abscissa and ordinate have been algorithmized, and each point represents a specific gene. The abscissa corresponding to a specific point is the expression level of the gene in SMN, and the ordinate is the expression level of the gene in SMS.

of filC. Crabbé et al. (2011) found upregulated gene expression related to motility in *Pseudomonas aeruginosa* under SMG, including fliACDGS, fleLNP, and flgM, but there were no obvious changes in the expression of genes related to motility after spaceflight. Duscher et al. (2018) revealed that *Vibrio fischeri* showed a trend of upregulation of flhA and flgM after 12 and 24 h of cultivating in SMG. Interestingly, Tucker et al. (2007) found that even in the same HARV environment, *E. coli* in different nutrients displayed different motility and had different expression of genes related to motility. In this study, compared with SMN and SMO, SMS showed more bacterial intercellular mucus and rougher membranes under 20,000x SEM. The swimming experiment revealed that the bacteria in the SMS group had increased motility compared with those in the SMN and SMO groups. In addition, considering that the swimming experiment was conducted under ground conditions and there would be some variables affecting the results, we will take videos of precultured bacterial cells directly from three samples to compare their motility in the subsequent experiment. Through the GO enrichment analysis performed in this research, we found that the increased motility of the SMS group may be related to some upregulated DEPs involved in locomotion, localization, biological adhesion, and binding. Moreover, according to the COG annotations, ten DEPs related to cell motility and seven DEPs related to cell wall, membrane, and envelope biogenesis were overexpressed in SMS. In addition, according to the GO enrichment analysis and KEGG enrichment analysis, DEGs, including pilP, pilM, flgE, flgG, and ronN, which are associated with cell motility, were overexpressed in SMS. These upregulated DEGs were consistent with the changes in the DEPs. In conclusion, we believe that the overlap of these changes may increase the motility of SMS in the SMG environment.

Most of the studies on *S. maltophilia* biofilms are related to the effects of antibiotics, hormones, or combinations of drugs. For instance, Wang et al. (2016a) discovered the synergistic

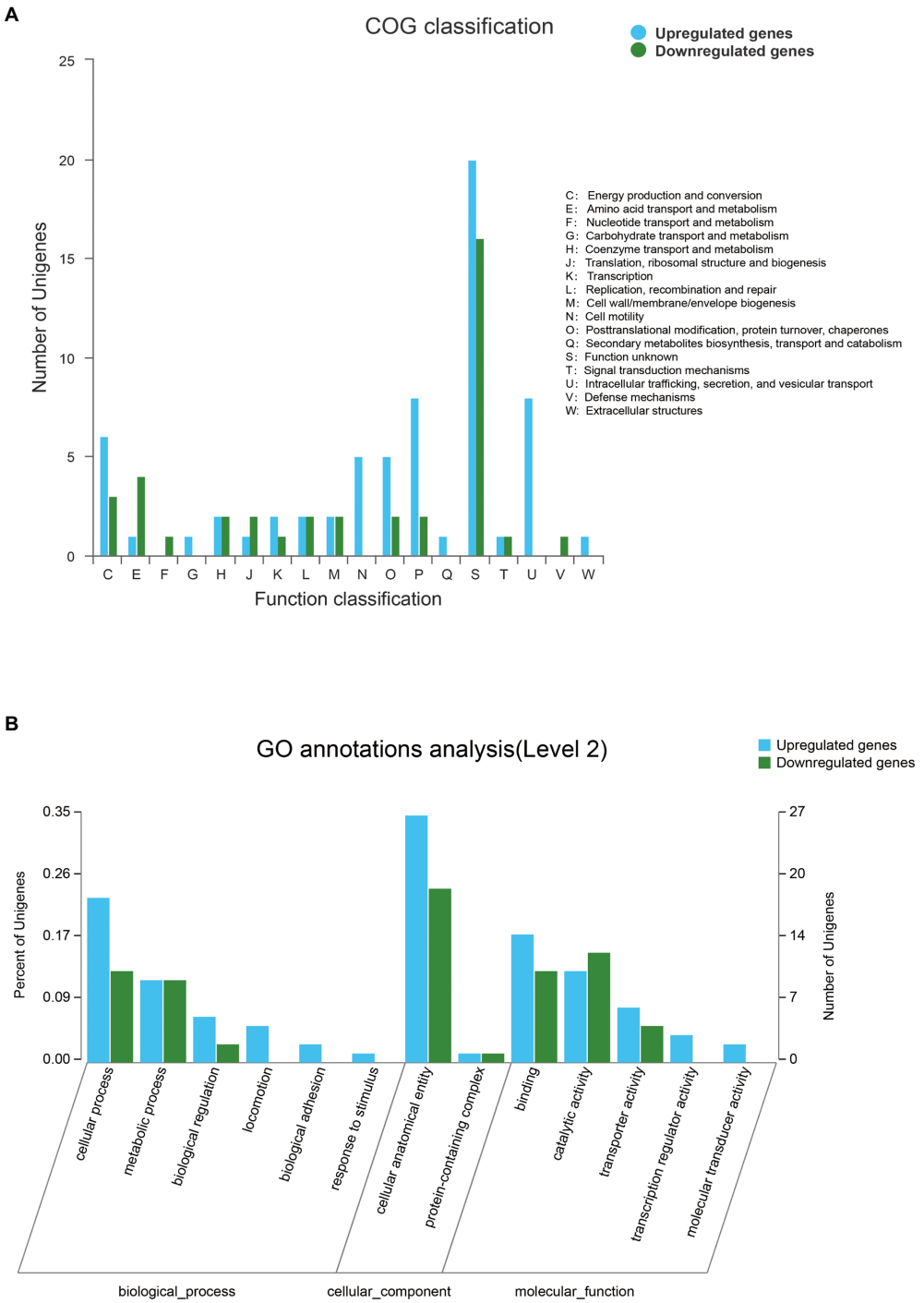


FIGURE 8 | (A) Distribution of differentially expressed genes (DEGs) between SMS and SMN in the cluster of COG. The X-axis represents the functional type of COG, and the *Y*-axis represents the number of genes. **(B)** Distribution of DEGs in Gene Ontology (GO) functional categories. The X-axis represents the secondary classification terms of GO, the *Y*-axis on the left represents the percentage of genes included in the secondary classification, the right represents the number of genes compared with the secondary classification, and the three colors represent the three major classifications.

inhibition of fluoroquinolones and azithromycin on the biofilm formation of *S. maltophilia*, and Ciacci et al. (2019) found an *in vitro* synergistic therapeutic effect of colistin and N-acetylcysteine on *S. maltophilia*. However, there are few studies on the characteristics of *S. maltophilia* in the space environment and SMG. Therefore, we focused on microgravity as a starting point to investigate the physiological characteristics of *S. maltophilia* in a microgravity environment. The formation

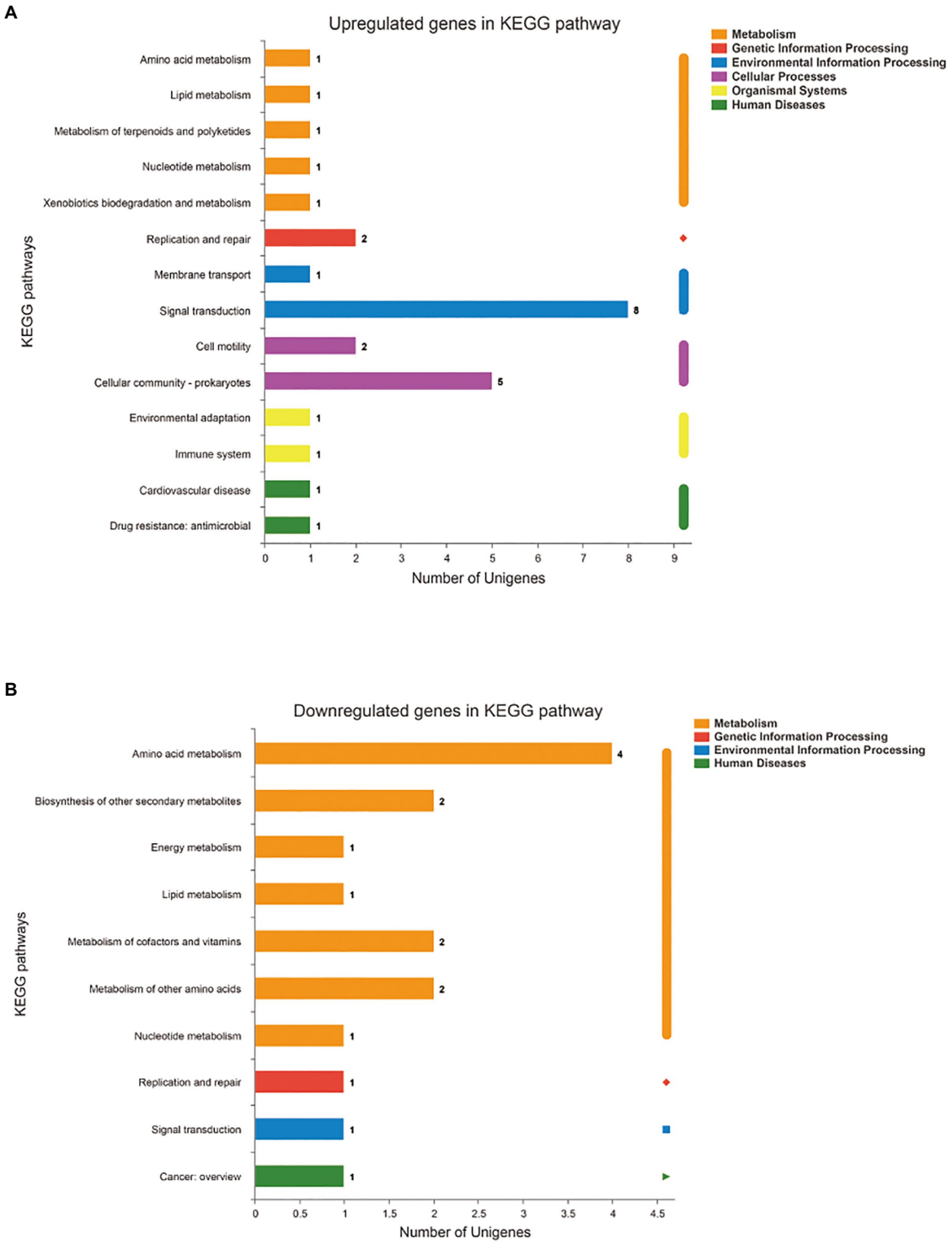


FIGURE 9 | Functional annotation and enrichment of genes of SMS. **(A)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of upregulated genes in SMS. **(B)** KEGG pathway of downregulated genes in SMS. The X-axis of **(A)** and **(B)** is the number of genes annotated to the KEGG pathway, the Y-axis of **(A)** and **(B)** is the name KEGG metabolic pathway.

of biofilms in bacteria was first carried out under microgravity in 2001 (McLean et al., 2001). After 24 h of exposure to microgravity, the formation of biofilms in *E. coli* increased (Lynch et al., 2006). *Escherichia coli* were cultured in a rotary cell culture system under microgravity and the biofilm formation ability increased in the SMG group compared with that in

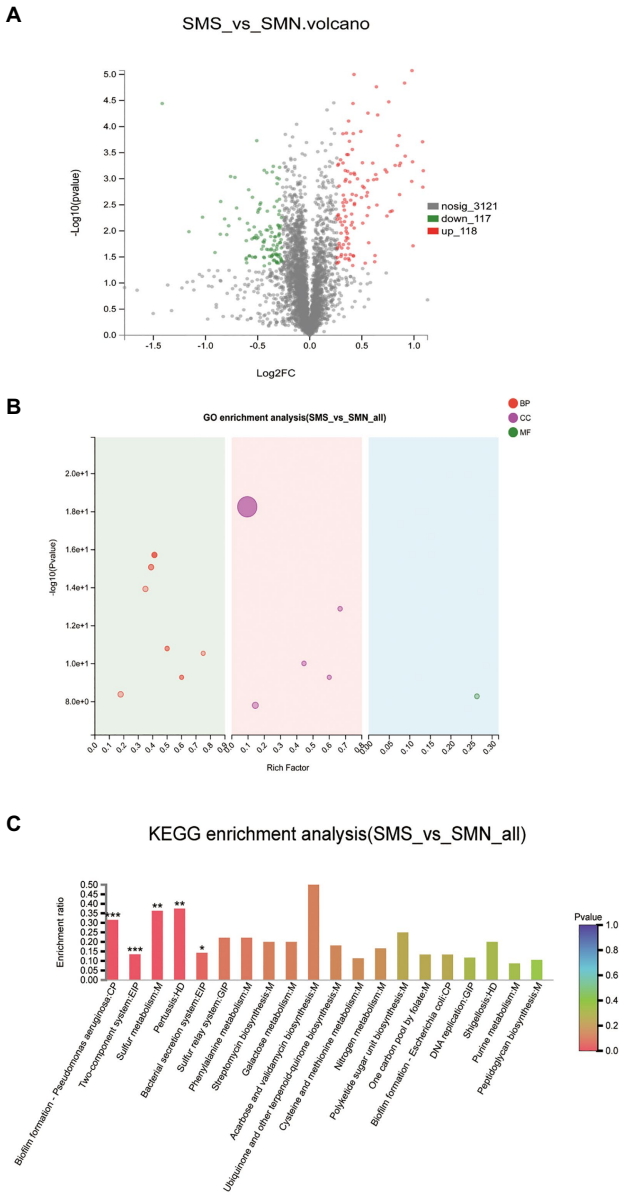


FIGURE 10 | Comparative proteomic analysis. **(A)**. Differentially expressed proteins (DEPs) between SMS and SMN. The X-axis is the fold change value of the difference in the protein expression between the two groups. The Y-axis is the statistical test value of the DEPs. The value of abscissa and ordinate have been logarithmized. Each point in the figure represents a specific protein. The green points are downregulated proteins and red points represent upregulated proteins. **(B)**. GO enrichment analysis of SMS and SMN. The X-axis is the rich factor. The Y-axis is -log10 (*value of p*). Each bubble in the figure represents a secondary GO category. The size of the bubble is directly proportional to the number of concentrated proteins enriched in this GO secondary classification (BP, biological process; CC, cellular component; and MF, molecular function). **(C)**. KEGG enrichment analysis of DEPs between SMS and SMN.

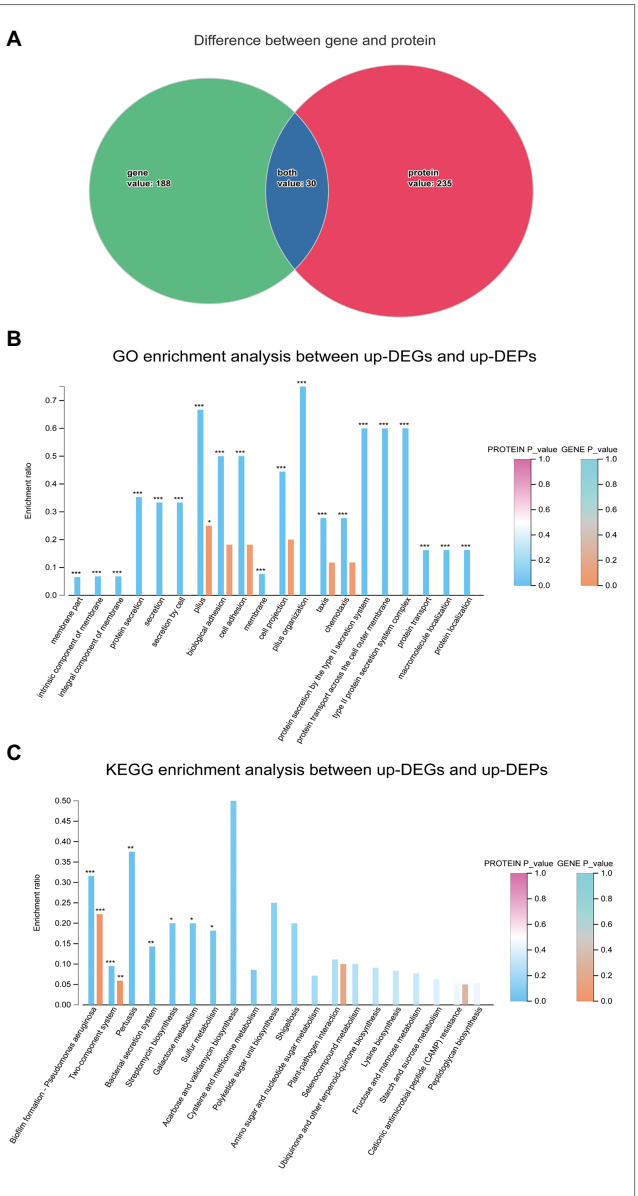


FIGURE 11 | Integration of transcriptomic and proteomic analysis. **(A)**. The overlapping genes expressed differently in both the transcriptome and proteome. **(B)**. GO enrichment analysis between up-DEGs and up-DEPs. **(C)**. KEGG enrichment analysis between up-DEGs and up-DEPs.

the NG group (Yim et al., 2020). The biofilm formation of the bacterial pathogen *S. typhimurium* was also increased after the space shuttle mission STS-115 (Wilson et al., 2007). These results were consistent with our results. However, Zhao et al. (2019) found that *Acinetobacter baumannii* showed decreased biofilm formation after a spaceflight on the Shenzhou 11 spacecraft of China. *Klebsiella pneumoniae* formed different subgroups, M1 and M2, on solid medium, and an increased biofilm forming ability was found in M1 compared with that in M2 in SMG (Wang et al., 2017). Therefore, there is no general conclusion on the influence of microgravity on microorganisms. The differences in strains, culture methods, and durations spent in microgravity may lead to different results. In our study, the biofilms that formed at the bottom of the HARV bioreactors in the SMS group were thicker than

those in the SMN group. Quantitative biofilm experiments showed that the biofilm formation ability of SMS was more than twice that of either SMN or SMO. CLSM analysis illustrated that the biofilm thickness in SMS was thicker than that in SMN or SMO. Through the GO and KEGG enrichment analyses performed in this research, we found that the biofilm formation ability of SMS might be related to some of the upregulated DEPs involved in intrinsic components of the membrane, integral components of the membrane, type II protein secretion systems, galactose metabolism, and sulfur metabolism. Moreover, according to the COG annotations, seven DEPs related to cell wall, membrane, and envelope biogenesis and 20 DEPs related to intracellular trafficking, secretion, and vesicular transport were overexpressed in SMS. In addition, according to the GO enrichment analysis and KEGG enrichment analysis, DEGs, including the chpB, chpC, rpoN, pilA, pilG, pilH, and pilJ genes, which are associated with biofilm formation and cell outer membrane, were overexpressed in SMS. These upregulated DEGs were consistent with the changes in the DEPs. In conclusion, we believe that these changes may increase the biofilm formation of SMS in the SMG environment.

To sum up, the changes in genome, transcriptome, and proteome can to some extent explain the phenotypic changes such as motility and biofilm formation ability. The present study might serve a basis for future studies of the complex mechanism by which bacteria adapt to the space. However, the overall changes in *S. maltophilia* at the level of DNA or RNA are still limited. Considering the complexity of space environment and the limitation of simulated experiment condition on the ground, more relevant experiments are needed for deep research.

DATA AVAILABILITY STATEMENT

All the data used to draw conclusions are presented in the article. The sequence reads used for WGS and RNA-seq have been submitted to NCBI with the project accession numbers of PRJNA720482 and PRJNA720572. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (Deutsch et al., 2020) partner repository with the dataset identifier PXD025290.

AUTHOR CONTRIBUTIONS

CL and JW designed and coordinated the project. XS performed the laboratory experiments and wrote the manuscript with assistance from all the authors. YG, TF, XJ, DW, DL, PB, and BZ performed the data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Key Program of Logistics Research (BWS17J030), and the Foundation Strengthening Program Technology Field Fund (2019-JCJQ-JJ-152).

ACKNOWLEDGMENTS

We would like to thank Professor Chunli Li of the Institute of Microbiology, Chinese Academy of Sciences, for his technical guidance on SEM. We would also like to thank the Professor Tingting Pan of the Translational Medicine Laboratory, Medical School of Chinese PLA, for her technical guidance of CLSM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.701265/full#supplementary-material

REFERENCES

Amin, R., and Waters, V. (2016). Antibiotic treatment for Stenotrophomonas maltophilia in people with cystic fibrosis. *Cochrane Database Syst. Rev.* 7:Cd009249. doi: 10.1002/14651858.CD009249.pub4

An, X., Tian, C., Xu, J., Dong, F., Liu, X., Wu, X., et al. (2020). Characterization of hexaconazole-degrading strain *Sphingobacterium multivorum* and analysis of transcriptome for biodegradation mechanism. *Sci. Total Environ.* 722:137171. doi: 10.1016/j.scitotenv.2020.137171

Benoit, M. R., and Klaus, D. M. (2007). Microgravity, bacteria, and the influence of motility. *Adv. Space Res.* 39, 1225–1232. doi: 10.1016/j.asr.2006.10.009

Brown, R. B., Klaus, D., and Todd, P. (2002). Effects of space flight, clinorotation, and centrifugation on the substrate utilization efficiency of *E. coli*. *Microgravity Sci. Technol.* 13, 24–29. doi: 10.1007/BF02881678

Cao, C., Wang, J., Liu, Y., Kwok, L. Y., Zhang, H., and Zhang, W. (2020). Adaptation of *Lactobacillus plantarum* to ampicillin involves mechanisms that maintain protein homeostasis. *mSystems* 5, e00853–e00819. doi: 10.1128/mSystems.00853-19

Castro, S. L., Nelman-Gonzalez, M., Nickerson, C. A., and Ott, C. M. (2011). Induction of attachment-independent biofilm formation and repression of Hfq expression by low-fluid-shear culture of *Staphylococcus aureus*. *Appl. Environ. Microbiol.* 77, 6368–6378. doi: 10.1128/AEM.00175-11

Checinska, A., Probst, A. J., Vaishampayan, P., White, J. R., Kumar, D., Stepanov, V. G., et al. (2015). Microbiomes of the dust particles collected from the international space station and spacecraft assembly facilities. *Microbiome* 3:50. doi: 10.1186/s40168-015-0116-3

Chen, Z., Zhong, W., Chen, S., Zhou, Y., Ji, P., Gong, Y., et al. (2021). TMT-based quantitative proteomics analyses of sterile/fertile anthers from a genic male-sterile line and its maintainer in cotton (*Gossypium hirsutum* L.). *J. Proteome* 232:104026. doi: 10.1016/j.jprot.2020.104026

Chi, B. B., Lu, Y. N., Yin, P. C., Liu, H. Y., Chen, H. Y., and Shan, Y. (2021). Sequencing and comparative genomic analysis of a highly metal-tolerant *Penicillium janthinellum* P1 provide insights Into its metal tolerance. *Front. Microbiol.* 12:663217. doi: 10.3389/fmicb.2021.663217

Chopra, V., Fadl, A. A., Sha, J., Chopra, S., Galindo, C. L., and Chopra, A. K. (2006). Alterations in the virulence potential of enteric pathogens and bacterial-host cell interactions under simulated microgravity conditions. *J. Toxicol. Environ. Health A* 69, 1345–1370. doi: 10.1080/15287390500361792

Ciacci, N., Boncompagni, S., Valzano, F., Cariani, L., Aliberti, S., Blasi, F., et al. (2019). In vitro synergism of Colistin and N-acetylcysteine against *Stenotrophomonas maltophilia*. *Antibiotics* 8:101. doi: 10.3390/antibiotics8030101

Clinical and Laboratory Standards Institute (CLSI). (2020). Performance Standards for Antimicrobial Susceptibility Testing, M100-30th Edition. Clinical and Laboratory Standards Institute, Wayne, PA, USA.

Coil, D. A., Neches, R. Y., Lang, J. M., Brown, W. E., Severance, M., Cavalier, D., et al. (2016). Growth of 48 built environment bacterial isolates on board the international Space Station (ISS). *PeerJ* 4:e1842. doi: 10.7717/peerj.1842

Crabbé, A., Schurr, M. J., Monsieurs, P., Morici, L., Schurr, J., Wilson, J. W., et al. (2011). Transcriptional and proteomic responses of Pseudomonas aeruginosa PAO1 to spaceflight conditions involve Hfq regulation and reveal a role for oxygen. *Appl. Environ. Microbiol.* 77, 1221–1230. doi: 10.1128/AEM.01582-10

Dai, J., Yu, X., Han, Y., Chai, L., Liao, Y., Zhong, P., et al. (2020). TMT-labeling proteomics of papillary thyroid carcinoma reveal invasive biomarkers. *J. Cancer* 11, 6122–6132. doi: 10.7150/jca.47290

DeRosier, D. J. (1998). The turn of the screw: the bacterial flagellar motor. *Cell* 93, 17–20. doi: 10.1016/S0092-8674(00)81141-1

Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., et al. (2020). The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* 48, D1145–d1152. doi: 10.1093/nar/gkz984

Duscher, A. A., Conesa, A., Bishop, M., Vroom, M. M., Zubizarreta, S. D., and Foster, J. S. (2018). Transcriptional profiling of the mutualistic bacterium Vibrio fischeri and an hfq mutant under modeled microgravity. *NPJ Microgravity* 4:25. doi: 10.1038/s41526-018-0060-1

Fajardo-Cavazos, P., Leehan, J. D., and Nicholson, W. L. (2018). Alterations in the Spectrum of spontaneous rifampicin-resistance mutations in the *Bacillus subtilis* rpoB gene after cultivation in the human spaceflight environment. *Front. Microbiol.* 9:192. doi: 10.3389/fmicb.2018.00192

Fajardo-Cavazos, P., and Nicholson, W. L. (2016). Cultivation of *Staphylococcus epidermidis* in the human spaceflight environment leads to alterations in the frequency and spectrum of spontaneous rifampicin-resistance mutations in the rpoB gene. *Front. Microbiol.* 7:999. doi: 10.3389/fmicb.2016.00999

Fujita, J., Yamadori, I., Xu, G., Hojo, S., Negayama, K., Miyawaki, H., et al. (1996). Clinical features of *Stenotrophomonas maltophilia* pneumonia in immunocompromised patients. *Respir. Med.* 90, 35–38. doi: 10.1016/S0954-6111(96)90242-5

Gao, H., Xie, F., Zhang, W., Tian, J., Zou, C., Jia, C., et al. (2020). Characterization and improvement of curdlan produced by a high-yield mutant of *Agrobacterium* sp. ATCC 31749 based on whole-genome analysis. *Carbohydr. Polym.* 245:116486. doi: 10.1016/j.carbpol.2020.116486

Haddad, G., Bellali, S., Takakura, T., Fontanini, A., Ominami, Y., Bou Khalil, J., et al. (2021). Scanning electron microscope: a new potential tool to replace gram staining for microbe identification in blood cultures. *Microorganisms* 9:1170. doi: 10.3390/microorganisms9061170

Hammond, T. G., and Hammond, J. M. (2001). Optimized suspension culture: the rotating-wall vessel. *Am. J. Physiol. Ren. Physiol.* 281, F12–F25. doi: 10.1152/ajprenal.2001.281.1.F12

Hao, M., Ma, W., Dong, X., Li, X., Cheng, F., and Wang, Y. (2021). Comparative genome analysis of multidrug-resistant Pseudomonas aeruginosa JNQH-PA57, a clinically isolated mucoid strain with comprehensive carbapenem resistance mechanisms. *BMC Microbiol.* 21:133. doi: 10.1186/s12866-021-02203-4

Horneck, G., Klaus, D. M., and Mancinelli, R. L. (2010). Space microbiology. *Microbiol. Mol. Biol. Rev.* 74, 121–156. doi: 10.1128/MMBR.00016-09

Hugh, R., and Leifson, E. J.Microbiology Society (1963). A description of the type strain of *Pseudomonas maltophilia*. *Int. J. Syst. Evol. Microbiol.* 13, 133–138.

Javadi, A., Shamaei, M., Mohammadi Ziazi, L., Pourabdollah, M., Dorudinia, A., Seyedmehdi, S. M., et al. (2014). Qualification study of two genomic DNA extraction methods in different clinical samples. *Tanaffos* 13, 41–47.

Jiang, X., Ren, S., Geng, Y., Jiang, C., Liu, G., Wang, H., et al. (2021). Role of the VirSR-VirAB system in biofilm formation of *Listeria monocytogenes* EGD-e. *Food Res. Int.* 145:110394. doi: 10.1016/j.foodres.2021.110394

Jiang, J., Yang, B., Ross, R. P., Stanton, C., Zhao, J., Zhang, H., et al. (2020). Comparative genomics of *Pediococcus pentosaceus* isolated from different niches reveals genetic diversity in carbohydrate metabolism and immune system. *Front. Microbiol.* 11:253. doi: 10.3389/fmicb.2020.00253

Kan, Y., Lyu, Q., Jiang, N., Han, S., Li, J., Burdman, S., et al. (2020). iTRAQ-based proteomic analyses of the plant-pathogenic bacterium Acidovorax citrulli during entrance into and resuscitation from the viable but nonculturable state. *J. Proteome* 211:103547. doi: 10.1016/j.jprot.2019.103547

Kim, W., Tengra, F. K., Shong, J., Marchand, N., Chan, H. K., Young, Z., et al. (2013). Effect of spaceflight on Pseudomonas aeruginosa final cell density is modulated by nutrient and oxygen availability. *BMC Microbiol.* 13:241. doi: 10.1186/1471-2180-13-241

Lai, C. H., Wong, W. W., Chin, C., Huang, C. K., Lin, H. H., Chen, W. F., et al. (2006). Central venous catheter-related *Stenotrophomonas maltophilia* bacteraemia and associated relapsing bacteraemia in haematology and oncology patients. *Clin. Microbiol. Infect.* 12, 986–991. doi: 10.1111/j.1469-0691.2006.01511.x

Landrum, M. L., Conger, N. G., and Forgione, M. A. (2005). Trimethoprim-sulfamethoxazole in the treatment of *Stenotrophomonas maltophilia* osteomyelitis. *Clin. Infect. Dis.* 40, 1551–1552. doi: 10.1086/429730

Lawal, A., Jejelowo, O. A., and Rosenzweig, J. A. (2010). The effects of low-shear mechanical stress on *Yersinia pestis* virulence. *Astrobiology* 10, 881–888. doi: 10.1089/ast.2010.0493

Lu, Z., Sun, W., Li, C., Ao, X., Yang, C., and Li, S. (2019). Bioremoval of non-steroidal anti-inflammatory drugs by *Pseudoxanthomonas* sp. DIN-3 isolated from biological activated carbon process. *Water Res.* 161, 459–472. doi: 10.1016/j.watres.2019.05.065

Lynch, S. V., Brodie, E. L., and Matin, A. (2004). Role and regulation of sigma S in general resistance conferred by low-shear simulated microgravity in *Escherichia coli*. *J. Bacteriol.* 186, 8207–8212. doi: 10.1128/JB.186.24.8207-8212.2004

Lynch, S. V., Mukundakrishnan, K., Benoit, M. R., Ayyaswamy, P. S., and Matin, A. (2006). Escherichia coli biofilms formed under low-shear modeled microgravity in a ground-based system. *Appl. Environ. Microbiol.* 72, 7701–7710. doi: 10.1128/AEM.01294-06

McLean, R. J., Cassanto, J. M., Barnes, M. B., and Koo, J. H. (2001). Bacterial biofilm formation under microgravity conditions. *FEMS Microbiol. Lett.* 195, 115–119. doi: 10.1111/j.1574-6968.2001.tb10507.x

Morrison, M. D., and Nicholson, W. L. (2020). Comparisons of transcriptome profiles from *Bacillus subtilis* cells grown in space versus high aspect ratio vessel (HARV) clinostats reveal a low degree of concordance. *Astrobiology* 20, 1498–1509. doi: 10.1089/ast.2020.2235

Murray, T. S., and Kazmierczak, B. I. (2006). FlhF is required for swimming and swarming in *Pseudomonas aeruginosa*. *J. Bacteriol.* 188, 6995–7004. doi: 10.1128/JB.00790-06

Nadeau, J., Lindensmith, C., Deming, J. W., Fernandez, V. I., and Stocker, R. (2016). Microbial morphology and motility as biosignatures for outer planet missions. *Astrobiology* 16, 755–774. doi: 10.1089/ast.2015.1376

Nickerson, C. A., Ott, C. M., Mister, S. J., Morrow, B. J., Burns-Keliher, L., and Pierson, D. L. (2000). Microgravity as a novel environmental signal affecting salmonella enterica serovar *Typhimurium virulence*. *Infect. Immun.* 68, 3147–3152. doi: 10.1128/IAI.68.6.3147-3152.2000

Papadakis, K. A., Vartivarian, S. E., Vassilaki, M. E., and Anaissie, E. J. (1995). *Stenotrophomonas maltophilia*: an unusual cause of biliary sepsis. *Clin. Infect. Dis.* 21, 1032–1034. doi: 10.1093/clinids/21.4.1032

Portela, R., Almeida, P. L., Sobral, R. G., and Leal, C. R. (2019). Motility and cell shape roles in the rheology of growing bacteria cultures. *Eur. Phys. J. E Soft Matter* 42:26. doi: 10.1140/epje/i2019-11787-9

Radtke, A. L., and Herbst-Kralovetz, M. M. (2012). Culturing and applications of rotating wall vessel bioreactor derived 3D epithelial cell models. *J. Vis. Exp.* 62:3868. doi: 10.3791/3868

Rasamiravaka, T., and El Jaziri, M. (2016). Quorum-sensing mechanisms and bacterial response to antibiotics in *P. aeruginosa*. *Curr. Microbiol.* 73, 747–753. doi: 10.1007/s00284-016-1101-1

Sanchez, M. B., Hernandez, A., and Martinez, J. L. (2009). Stenotrophomonas maltophilia drug resistance. *Future Microbiol.* 4, 655–660. doi: 10.2217/fmb.09.45

Sawyer, N. B., Worrall, L. K., Crowe, J. A., Waters, S. L., Shakesheff, K. M., Rose, F. R., et al. (2008). In situ monitoring of 3D in vitro cell aggregation using an optical imaging system. *Biotechnol. Bioeng.* 100, 159–167. doi: 10.1002/bit.21728

Shui, Y., Jiang, Q., Lyu, X., Wang, L., Lin, Y., Ma, Q., et al. (2021). Inhibitory effects of sodium new houttuyfonate on growth and biofilm formation of *Streptococcus mutans*. *Microb. Pathog.* 157:104957. doi: 10.1016/j.micpath.2021.104957

Takigawa, M., Noda, T., Kurita, T., Okamura, H., Suyama, K., Shimizu, W., et al. (2008). Extremely late pacemaker-infective endocarditis due to *Stenotrophomonas maltophilia*. *Cardiology* 110, 226–229. doi: 10.1159/000112404

Taylor, P. W. (2015). Impact of space flight on bacterial virulence and antibiotic susceptibility. *Infect. Drug Resist.* 8, 249–262. doi: 10.2147/IDR.S67275

Tirumalai, M. R., Karouia, F., Tran, Q., Stepanov, V. G., Bruce, R. J., Ott, C. M., et al. (2017). The adaptation of Escherichia coli cells grown in simulated microgravity for an extended period is both phenotypic and genomic. *NPJ Microgravity* 3:15. doi: 10.1038/s41526-017-0020-1

Tucker, D. L., Ott, C. M., Huff, S., Fofanov, Y., Pierson, D. L., Willson, R. C., et al. (2007). Characterization of Escherichia coli MG1655 grown in a low-shear modeled microgravity environment. *BMC Microbiol.* 7:15. doi: 10.1186/1471-2180-7-15

Vaishampayan, A., and Grohmann, E. (2019). Multi-resistant biofilm-forming pathogens on the international space station. *J. Biosci.* 44:125. doi: 10.1007/s12038-019-9929-8

Venkateswaran, K., Vaishampayan, P., Cisneros, J., Pierson, D. L., Rogers, S. O., and Perry, J. (2014). International Space Station environmental microbiome - microbial inventories of ISS filter debris. *Appl. Microbiol. Biotechnol.* 98, 6453–6466. doi: 10.1007/s00253-014-5650-6

Wang, Y., Jiang, N., Wang, B., Tao, H., Zhang, X., Guan, Q., et al. (2020). Integrated transcriptomic and proteomic analyses reveal the role of NprR in bacillus anthracis extracellular protease expression regulation and oxidative stress responses. *Front. Microbiol.* 11:590851. doi: 10.3389/fmicb.2020.590851

Wang, H., Li, W., Gu, L., Gao, X., Ni, B., Deng, H., et al. (2017). Emergence of two distinct subpopulations from *Klebsiella pneumoniae* grown in the stimulated microgravity environment. *Future Microbiol.* 12, 939–951. doi: 10.2217/fmb-2017-0032

Wang, A., Wang, Q., Kudinha, T., Xiao, S., and Zhuo, C. (2016a). Effects of fluoroquinolones and azithromycin on biofilm formation of *Stenotrophomonas maltophilia*. *Sci. Rep.* 6:29701. doi: 10.1038/srep29701

Wang, H., Yan, Y., Rong, D., Wang, J., Wang, H., Liu, Z., et al. (2016b). Increased biofilm formation ability in *Klebsiella pneumoniae* after short-term exposure to a simulated microgravity environment. *Microbiology* 5, 793–801. doi: 10.1002/mbo3.370

Wilson, J. W., Ott, C. M., Höner zu Bentrup, K., Ramamurthy, R., Quick, L., Porwollik, S., et al. (2007). Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. *Proc. Natl. Acad. Sci. U. S. A.* 104, 16299–16304. doi: 10.1073/pnas.0707155104

Wilson, J. W., Ott, C. M., Ramamurthy, R., Porwollik, S., McClelland, M., Pierson, D. L., et al. (2002a). Low-shear modeled microgravity alters the *Salmonella enterica serovar typhimurium* stress response in an RpoS-independent manner. *Appl. Environ. Microbiol.* 68, 5408–5416. doi: 10.1128/aem.68.11.5408-5416.2002

Wilson, J. W., Ramamurthy, R., Porwollik, S., McClelland, M., Hammond, T., Allen, P., et al. (2002b). Microarray analysis identifies *Salmonella* genes belonging to the low-shear modeled microgravity regulon. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13807–13812. doi: 10.1073/pnas.212387899

Wolf, D.A., Sams, C., and Schwarz, R.P. (1992). High-aspect-ratio rotating cell-culture vessel. NASA Tech Briefs. Document ID 19920000327, NASA Johnson Space Center, Houston, TX 16.

Ye, S. Q., Zou, Y., Zheng, Q. W., Liu, Y. L., Li, R. R., Lin, J. F., et al. (2021). TMT-MS/MS proteomic analysis of the carbohydrate-active enzymes in the fruiting body of *Pleurotus tuoliensis* during storage. *J. Sci. Food Agric.* 101, 1879–1891. doi: 10.1002/jsfa.10803

Yemisen, M., Mete, B., Tunali, Y., Yentur, E., and Ozturk, R. (2008). A meningitis case due to *Stenotrophomonas maltophilia* and review of the literature. *Int. J. Infect. Dis.* 12, e125–e127. doi: 10.1016/j.ijid.2008.03.028

Yim, J., Cho, S. W., Kim, B., Park, S., Han, Y. H., and Seo, S. W. (2020). Transcriptional profiling of the probiotic *Escherichia coli* Nissle 1917 strain under simulated microgravity. *Int. J. Mol. Sci.* 21:2666. doi: 10.3390/ijms21082666

Zhang, B., Bai, P., Zhao, X., Yu, Y., Zhang, X., Li, D., et al. (2019). Increased growth rate and amikacin resistance of *Salmonella enteritidis* after one-month spaceflight on China's Shenzhou-11 spacecraft. *Microbiology* 8:e00833. doi: 10.1002/mbo3.833

Zhang, L. S., Chu, M. Y., Zong, M. H., Yang, J. G., and Lou, W. Y. (2020). Carbon source modify lipids composition of *Rhodococcus opacus* intended for infant formula. *J. Biotechnol.* 319, 8–14. doi: 10.1016/j.jbiotec.2020.05.013

Zhang, S. Y., Zhang, S. W., Zhang, T., Fan, X. N., and Meng, J. (2021). Recent advances in functional annotation and prediction of the epitranscriptome. *Comput. Struct. Biotechnol. J.* 19, 3015–3026. doi: 10.1016/j.csbj.2021.05.030

Zhao, X., Yu, Y., Zhang, X., Huang, B., Bai, P., Xu, C., et al. (2019). Decreased biofilm formation ability of *Acinetobacter baumannii* after spaceflight on China's Shenzhou 11 spacecraft. *Microbiology* 8:e00763. doi: 10.1002/mbo3.763

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Su, Guo, Fang, Jiang, Wang, Li, Bai, Zhang, Wang and Liu. 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: 10 September 2021
doi: 10.3389/fmicb.2021.696102

Description, Taxonomy, and Comparative Genomics of a Novel species, *Thermoleptolyngbya sichuanensis* sp. nov., Isolated From Hot Springs of Ganzi, Sichuan, China

*Jie Tang[1], Liheng Li[2], Meijin Li[2], Lianming Du[1], Md Mahfuzur R. Shah[2], Michal M. Waleron[3], Malgorzata Waleron[4], Krzysztof F. Waleron[3] and Maurycy Daroch[2]**

[1] Antibiotics Research and Re-evaluation Key Laboratory of Sichuan Province, Sichuan Industrial Institute of Antibiotics, Chengdu University, Chengdu, China, [2] School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China, [3] Department of Pharmaceutical Microbiology, Faculty of Pharmacy, Medical University of Gdańsk, Gdańsk, Poland, [4] Department of Plant Protection and Biotechnology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Gdańsk, Poland

OPEN ACCESS

Edited by:
Brian P. Hedlund, University of Nevada, Las Vegas, United States

Reviewed by:
Xiuzhu Dong, Institute of Microbiology, Chinese Academy of Sciences (CAS), China
Wei Xie, Sun Yat-sen University, China

***Correspondence:**
Maurycy Daroch
m.daroch@pkusz.edu.cn

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

Received: *16 April 2021*
Accepted: *03 August 2021*
Published: *10 September 2021*

Citation:
Tang J, Li L, Li M, Du L, Shah MMR, Waleron MM, Waleron M, Waleron KF and Daroch M (2021) Description, Taxonomy, and Comparative Genomics of a Novel species, Thermoleptolyngbya sichuanensis sp. nov., Isolated From Hot Springs of Ganzi, Sichuan, China. Front. Microbiol. 12:696102. doi: 10.3389/fmicb.2021.696102

Thermoleptolyngbya is a newly proposed genus of thermophilic cyanobacteria that are often abundant in thermal environments. However, a vast majority of *Thermoleptolyngbya* strains were not systematically identified, and genomic features of this genus are also sparse. Here, polyphasic approaches were employed to identify a thermophilic strain, PKUAC-SCTA183 (A183 hereafter), isolated from hot spring Erdaoqiao, Ganzi prefecture, China. Whole-genome sequencing of the strain revealed its allocation to *Thermoleptolyngbya* sp. and genetic adaptations to the hot spring environment. While the results of 16S rRNA were deemed inconclusive, the more comprehensive polyphasic approach encompassing phenetic, chemotaxic, and genomic approaches strongly suggest that a new taxon, *Thermoleptolyngbya sichuanensis* sp. nov., should be delineated around the A183 strain. The genome-scale phylogeny and average nucleotide/amino-acid identity confirmed the genetic divergence of the A183 strain from other strains of *Thermoleptolyngbya* along with traditional methods such as 16S-23S ITS and its secondary structure analyses. Comparative genomic and phylogenomic analyses revealed inconsistent genome structures between *Thermoleptolyngbya* A183 and O-77 strains. Further gene ontology analysis showed that the unique genes of the two strains were distributed in a wide range of functional categories. In addition, analysis of genes related to thermotolerance, signal transduction, and carbon/nitrogen/sulfur assimilation revealed the ability of this strain to adapt to inhospitable niches in hot springs, and these findings were preliminarily confirmed using experimental, cultivation-based approaches.

Keywords: *Thermoleptolyngbya*, 16S rRNA, 16S-23S ITS, comparative genomics, thermophilic cyanobacterium, Oculatellaceae

INTRODUCTION

Cyanobacteria are widely distributed microorganisms in various ecological niches due to abundant features allowing extensive adaptations. In particular, cyanobacterial populations existing in thermal environments have attracted increased interests in light of a crucial role in energy metabolism and matter cycling in ecosystems (Amin et al., 2017). Among the cyanobacteria communities, strains morphologically assigned to genus *Leptolyngbya* are often reported to be prosperous in many thermal environments (Mackenzie et al., 2012; Amarouche-Yala et al., 2014; Tang et al., 2018b; Strunecký et al., 2019).

Identification of *Leptolyngbya*-like strains has been controversial because of their simple morphology, lacking significant discrimination. The heterogeneity of *Leptolyngbya* has been questioned since the establishment of this genus (Bruno et al., 2009). The genus *Leptolyngbya* has been recognized as polyphyletic (Johansen et al., 2011; Perkerson et al., 2011), and there are strong recommendations to conduct a taxonomic reevaluation of this genus. In light of limited information provided by cell morphology investigations of trichomes, genetic and molecular techniques have been applied to facilitate the establishment of correct taxonomy. The 16S rRNA gene has been proposed as a universal DNA barcoding marker for species-level identification of bacterial isolates, and as a complement to morphology-based taxa identification (Yarza et al., 2014). However, in some cases, the 16S rRNA gene cannot resolve cyanobacterial phylogeny at the species level (Niclas et al., 2010; Eckert et al., 2014). An additional genomic locus, 16S-23S intergenic spacer (ITS), has been used for cyanobacterial systematics. It has been applied for the construction of phylogenetic trees and through the analysis of secondary structures of 16S-23S ITS regions (Johansen et al., 2011). Numerous studies confirmed the integrated approach of analyzing 16S rRNA gene phylogeny and 16S-23S ITS secondary structure to be useful and robust for complex cyanobacterial taxonomy, as in the case of the species or genera within the family *Leptolyngbyaceae* (Komarek et al., 2014; Debnath et al., 2017; Shalygin et al., 2020). In light of these examples, the application of a polyphasic approach for cyanobacteria identification is crucial (Komárek, 2016, 2018).

Recently, *Thermoleptolyngbya*, a cryptogenus newly delineated using 16S rRNA gene and 16S-23S ITS, emerged from strains originally grouped into *Leptolyngbya* and was morphologically characterized by filaments, isopolar trichomes surrounded by a thin colorless sheath, and parietal thylakoids (Sciuto and Moro, 2016). Two species were ascribed to this new taxon: *T. albertanoae* and *T. oregonensis*. Still, the majority of strains ascribed to *Thermoleptolyngbya* were not systematically identified, and 16S rRNA gene sequences were exclusively used for their taxonomic recognition. As such, the genus requires expanded sequence information to complement the 16S rRNA gene taxonomy by multiple-locus sequence analysis (MLSA) or on the whole genome level. The widened sequence space, especially about the underrepresented members of the genus, is required to guide phylogenetic and taxonomic studies and eventual reclassification. In addition, the acquisition of a complete genome may provide new insights into the genomic features of genus *Thermoleptolyngbya*, particularly the survival mechanism in thermal conditions from a genomic perspective.

Strain A183, originally isolated from Erdaoqiao hot springs (Tang et al., 2018b) in Ganzi Prefecture, Sichuan Province, China, was previously reported as a potential new species (Tang et al., 2018a) and was used for taxonomic reevaluation in the current study. The morphological and molecular data were collected for this thermophilic *Thermoleptolyngbya*. This work aimed to provide deeper insights into the taxonomy and genomic features of *Thermoleptolyngbya* strains. Characterization in terms of morphology, physiology, and phylogeny was achieved by microscopic, experimental, and molecular analysis.

MATERIALS AND METHODS

Thermoleptolyngbya sichuanensis sp. nov. A183: Origins, Cultivation, and Basic Physiological Assessment

The stock culture of strain A183, cryopreserved for over 2 years as 10% DMSO in BG11 stock in −80°C, was used to establish the final precultures for experiments, essentially as described by Tang et al. (2018a). The shake-flask cultures were grown without carbon supplementation in a shaking incubator in BG11 medium at 45°C, 100 rpm, under a photoperiod of 16-h light (45 μmol m^{-2} s^{-1}) and 8-h darkness. Cells were subcultured every 2 weeks until the growth reached the logarithmic phase (OD_{685nm} of 1.0–1.5). Cells were subcultured by resuspending 2% volume of centrifuged cells into a 100-ml fresh medium in 250-ml conical flasks sealed with a perforated sealing film. The strain was initially denoted and deposited in Peking University Algae Collection as PKUAC-SCTA183 has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession number FACHB-2491.

The assessment of the capacity of the strain for nitrate, nitrite, and urea utilization was performed as described previously (Liang et al., 2019). Briefly, strains were cultivated in a nitrogen-free BG-11 medium supplemented with different nitrogen sources up to a concentration: 17 mM $NaNO_2$, 85 mM $NaNO_3$, 3 mM urea. The effect of sulfates and sulfites on the cultivation of the strain has been tested with the addition of 10 mM Na_2SO_4 and 10 mM $NaHSO_3$ to the BG-11 medium lacking sulfur, respectively. The growth parameters have been tested against the same strain grown in a standard BG-11 medium. The ability of the strain to fix molecular nitrogen has been performed in the course of 72 h using the acetylene reduction method according to the previously described methodology (Chen et al., 1998).

Microscopic Analysis

The isolated cyanobacterium was investigated using light microscopy LM, DP72; (Olympus, Tokyo, Japan). Approximately 5 μl of culture was dropped on the microscopy slide and observed under 400× magnification. The images were captured using a U-TV0.63XC camera (Olympus). Scanning electron microscopy (SEM) was performed as follows: cells were washed gently with

PBS (Servicebio, Boston, MA, United States, G0002), and fixed for 2 h in fixation solution (Servicebio, G1102). Subsequently, the cells were postfixed with 1% OsO_4 (Ted Pella Inc., Redding, CA, United States) in 0.1 M phosphate buffer (pH 7.4) for 1–2 h at room temperature. The fixed material was dehydrated in a graded ethanol series (30–100%) (Sinopharm, Shanghai, China) and isoamyl acetate (Sinopharm, Shanghai, China) for 15 min and dried with Critical Point Dryer (Quorum, Laughton, United Kingdom, K850). Specimens were then attached to the metallic stubs using carbon stickers and sputter-coated with gold for 30 s. Coated samples were examined directly under the scanning electron microscope (Hitachi, Tokyo, Japan, SU8100). For transmission electron microscopy (TEM), the fixation solution G1102 (Servicebio) was added to the isolated cells. The cells were subsequently pelleted and resuspended in the fresh fixation solution. Cooled 1% agarose solution was mixed with the cells, and the agarose blocks were post-fixed with 1% OsO_4 in 0.1 M phosphate buffer (pH 7.4) for 2 h. Then cells were dehydrated as described above for SEM and embedded in pure EMBed 812 resins 90529-77-4 (SPI, West Chester, PA, United States). Embedded cells were incubated in a 65°C oven for more than 48 h to complete polymerization. The sections were cut to 60–80 nm thin layers using the ultra-microtome Leica EM UC7 (Leica, Wetzlar, Germany), stained with 2% uranium acetate saturated alcohol solution and lead citrate for 8 min, and examined using TEM (Hitachi, HT7800).

Genome Sequencing and Assembly

Genomic DNA of strain A183 was extracted and purified using a bacterial genomic DNA isolation kit (Generay, Shanghai, China) according to the manufacturer's instructions. Purified genomic DNA was subjected to 1% agarose gel electrophoresis for the analysis of its integrity and assessed spectrophotometrically with Nanophotometer (Impeln) to determine the DNA concentration and optical purity. The whole-genome sequencing of A183 strain was performed using a hybrid sequencing strategy combining PacBio long reads and Illumina short reads. For Illumina sequencing, the libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, Ipswich, MA, United States) following manufacturer's recommendations, and index codes were added to attribute sequences to the sample, as the DNA sample was fragmented by sonication to a size of 350 bp. The short read of A183 was sequenced using Illumina NovaSeq PE150 at the Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). For PacBio sequencing the DNA library with an insert size of 10 kb was constructed and sequencing was performed with P6-C4 chemistry according to manufacturer's recommendations. Two SMRT cells were used for PacBio sequencing and yielded 57,735 adapter-trimmed reads (subreads) with an average read length of approximately 9 kb, corresponding to 94-fold coverage. Illumina NovaSeq sequencing of strain A183 generated a total of 6,810,074 filtered paired-end reads (clean data), providing approximately 185-fold coverage of the genome. These clean data were assembled into contigs using MaSuRCA v. 3.3.9 with default parameters (Zimin et al., 2013), finally generating a single contig. The genome obtained was mapped with Illumina reads using BWA v0.7.17 (Li and Durbin, 2009) and then Pilon v1.23 (Walker et al., 2014) to correct any assembly and sequence errors.

Phylogenetic Reconstruction

Sequences of the 16S rRNA gene and 16S-23S ITS were extracted from A183 strain genome for phylogenetic analysis. Reference sequences of cyanobacteria were also retrieved from GenBank through BLAST search for 16S rRNA gene and 16S-23S ITS dataset construction, respectively. Multiple alignments of sequences were generated with Muscle incorporated in Mega7 (Kumar et al., 2016). Alignments were subjected to manual editing where necessary. Sequences of each alignment were trimmed to the same length.

Phylogenetic trees of 16S rRNA and 16S-23S ITS sequence datasets were reconstructed using Maximum-Likelihood (ML), Maximum-Parsimony (MP), and Neighbor-Joining (NJ) methods, respectively. ML phylogenetic analyses were both carried out using PhyML version 3.0 (Guindon et al., 2010), and the substitution models were selected based on the Akaike information criterion (AIC) by Model Selection function implemented in PhyML (Vincent et al., 2017). The NJ trees were both constructed using the General Time Reversible (GTR) model implemented in Mega7. Non-parametric bootstrap tests (1000 replications) were applied to evaluate the robustness of tree topologies.

The phylogenomic relationship between A183 and focus taxa was inferred using the concatenated sequences from single-copy genes shared by all the genomes. Single-copy genes were identified by OrthoMCL (Li et al., 2003), concatenated using a customized Perl script and aligned by MAFFT v7.453 (Standley, 2013). The ML genomic tree was constructed by IQ-TREE v2.1.3 (Minh et al., 2020) using the substitution model selected by ModelFinder implemented in IQ-TREE. Tree topology was assessed by UltraFast Bootstrap (1000 replicates) (Hoang et al., 2018). The strains from the family *Leptolyngbyaceae* were used to root the tree.

Secondary Structure Prediction

The tRNAs presented in 16S-23S ITS sequences were predicted by tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997). The conserved domains (D1-D1', D2, D3, boxA, and D4) and the variable regions (V2, boxB, and V3) of 16S-23S ITS were detected as reported by Iteman et al. (2000). The secondary structures of these DNA fragments were individually folded by Mfold web server (Zuker, 2003). Except for the use of the structure draw mode untangle with loop fix, default conditions in Mfold were used in all cases.

Genome Annotation and Comparative Genome Analysis

The genome of A183 strain was annotated using a modified pipeline previously established by Tang et al. (2019). Briefly, gene prediction and annotation were automatically performed using the NCBI prokaryotic genome annotation pipeline (Pruitt et al., 2009), and further using the RAST annotation system to minimize poor calls. The insertion sequence (IS) was detected and annotated by ISsaga (Varani et al., 2011). Prophage regions

were predicted by PHASTER (Arndt et al., 2016). CRISPR loci were detected using CRISPRCasFinder server (Grissa et al., 2007). The protein sequences predicted by RAST were aligned against the NCBI non-redundant database using BLASTP with an *E*-value cut-off of 1e-5. The alignment results were imported into BLAST2GO V5.2.5 (Conesa et al., 2005) for GO term mapping. The results of BLAST2GO analysis were submitted to the WEGO (Ye et al., 2006) for GO classification under the biological process, molecular function, and cellular component ontologies. The circular plot of the A183 genome was generated by Circos v0.68 (Krzywinski et al., 2009).

The genome of the *Thermoleptolyngbya* sp. strain O-77 (AP017367) was used for comparative genomic analysis with strain A183 (CP053661). The genome sequence was also subjected to the annotation pipeline mentioned above to keep all the data analyzed under the same criteria. To compare the gene context, all-against-all BLASTP alignments were performed using the following thresholds: *E*-value cut-off of 1e-5 and 30% identity, and the best hit of alignments was selected. Orthologous genes were identified with the bidirectional best hit (BBH) criterion (Brilli et al., 2010).

The whole-genome average nucleotide identity (ANI) and average amino acid identity (AAI) between genomes were calculated using the ANI/AAI calculator with default settings[1]. Only genomes with near completeness (≥90%) and low contamination (<5%) were retrieved from NCBI for ANI/AAI analysis.

RESULTS AND DISCUSSION

Morphological Investigation

The cell morphology of the A183 strain indicated by light microscopy revealed straight, wavy, and occasionally bent trichomes (**Figure 1A**). The SEM and TEM (**Figures 1A–D**) showed that trichomes of strain A183 were unbranched and composed of 80–120 elongated barrel-shaped cells, 1.30–1.60 μm in length and 1.05–1.10 μm in width. Constrictions were observed at the cross-walls of cells (**Figures 1B–D**). Individual cells of the filaments were divided by centripetal invagination of the cell wall (**Figure 1E**). Intracellular connections between vegetative cells were not observed (**Figure 1D**). The TEM analysis also exhibited that the three to five thylakoid layers were located in parallel at the inner periphery of cells (**Figures 1C,D**) and can be described as parietal according to recent classification (Mareš et al., 2019). Sheath, septum, phycobilisome, carboxysomes, cyanophycin granule, and polyphosphate bodies were present in the cytoplasm (**Figures 1C–F**), and small lipid droplets were also observed (**Figure 1C**).

The morphological characteristics of A183 strain showed certain similarity to that of other *Thermoleptolyngbya* strains (Sciuto and Moro, 2016). The morphological description of all strains belonging to the genus is summarized in **Table 1**. Strain A183, together with strains ETS-08 and PCC 8501, were all blue-green filamentous cells and exhibited no vesicles. The sheath of A183 and PCC 8501 was unlayered, while ETS-08 showed a multilayered sheath. There were granules observed at the cross walls of A183 and ETS-08 but no of PCC 8501. This observation reinforces the claim that on the morphology level alone, it is impossible to make any final taxonomic conclusions.

Physiological Characteristics of Strain A183

Basic physiological characterization of the strain has been assessed by monitoring its growth with various modifications of the BG-11 medium (**Table 2**). The strain was capable to actively grow using externally added sodium bicarbonate up to a concentration of 0.5 M. This indicates that bicarbonate transporters, typical for many cyanobacteria are also active in A183. Physiological testing of sulfur compounds revealed that the strain responds positively to 10 mM sulfate concentration in the growth medium but negatively to a similar concentration of sulfite. This is similar to previous findings concerning *Thermosynechococcus* E542 (Liang et al., 2019). Analysis of the utilization of various nitrogen sources indicates that the strain is capable of using nitrate and urea as nitrogen sources. The latter promotes the growth of the strain at 3-mM concentrations and inhibits at 6 mM. The A183 strain is also capable of diazotrophy. Acetylene reduction proxy assay revealed that similar to other *Thermoleptolyngbya* strains (Yoon et al., 2017), the A183 possesses a functional nitrogenase capable of molecular nitrogen fixation. In a course of a 72-h-long assay conditions, the cells exhibited nitrogenase activity; the activity was highest in the first 24 h of the assay and then plateaued (**Supplementary Figure 1**).

General Features of Strain A183 Genome

The combined assembly of PacBio and Illumina sequencing data successfully generated the complete genome of strain A183. The genome (**Figure 2A**) comprises a single circular chromosome with a size of 5,525,100 bp (GC content, 56.38 mol%) and no plasmid. Gene prediction and annotation of strain A183 resulted in 5,166 protein-coding sequences (CDS), approximately half (50.3%) of which were identified as hypothetical proteins. Functional distribution on GO categories of all CDS identified was summarized in **Supplementary Figure 2**. Two ribosomal RNA (*rrn*) operons were detected and 45 tRNA genes were predicted in the A183 chromosome (**Table 3**).

In the A183 chromosome, 186 ISs representing 28 different IS families were identified. The most frequently observed IS type was the ISKra4 family (30.11%), followed by the IS630 family (26.88%) and IS4 family (19.89%). Numerous genes encoding transposase (**Supplementary Table 1**) were also observed, indicating that the genetic plasticity of the strain might be shaped by intragenomic rearrangements. It was proposed that transpositions play a crucial role in genomic rearrangements and are involved in gene regulation and adaptation processes that determine the directions of microevolutionary processes in cyanobacteria (Mikheeva et al., 2013).

The A183 chromosome harbored around 190 transporter-related genes (**Supplementary Table 1**). Among these

[1]http://enve-omics.ce.gatech.edu/ani/

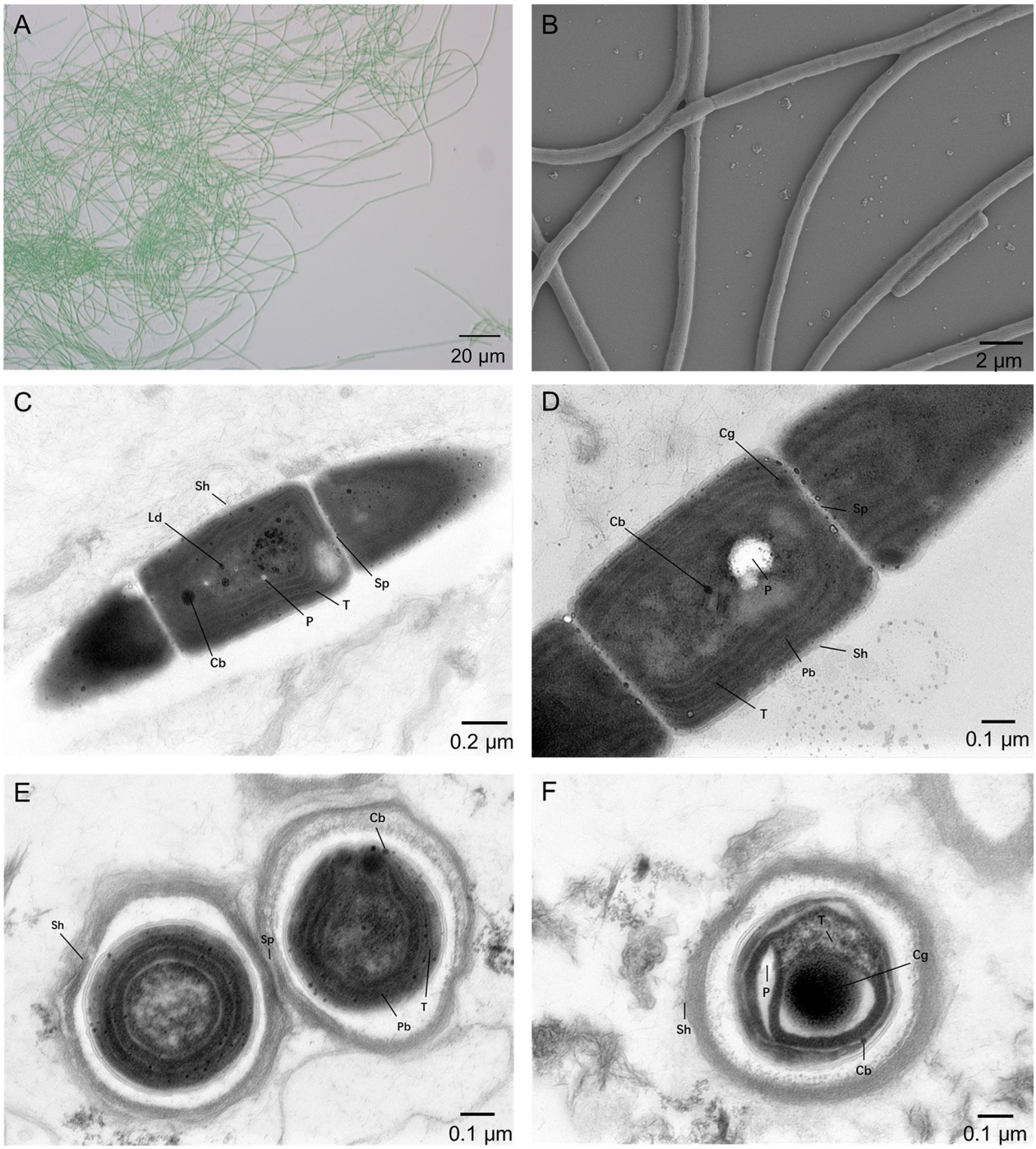


FIGURE 1 | Micrographs of A183. **(A)** Light microscopy image. **(B)** SEM image. **(C–F)** TEM images. Cb, carboxysome; Cg, cyanophycin granule; Cn, cyanophycin granule; Ld, lipid droplet; P, polyphosphate body; Pb, phycobilisome; Sh, sheath; Sp, septum; T, thylakoid membrane. Magnifications: 400× **(A)**, 5,000× **(B)**, 8,000× **(C)**, and 12,000× **(D–F)**.

transporters, ABC transporters accounted for the majority, and distinct bias was found in many genes, such as P-type ATPase transporter for copper, which only has two copies. Functionally, these transporters have been predicted as Na^+/H^+, iron, phosphate, amino acid, bicarbonate, CO_2 transporters, etc.

Phylogeny of 16S rRNA

To ascertain the taxonomic position of strain A183, a ML phylogenetic tree was reconstructed based on 16S rRNA gene sequences of the 58 cyanobacterial strains. The ML tree (**Figure 3**) resolved 13 well-defined clades of isolates corresponding to

TABLE 1 | Morphological features of known *Thermoleptolyngbya* strains.

Strain	Ecology	Cell width (μm)	Cell length (μm)	Sheath	Granule	Color	Thylakoids No.	References
A183	Hot spring	1.05–1.10	1.30–1.60	Unlayered	Present	Blue-green	3–5	This work
ETS-08	Thermal mud	<1	1.5–3	Multi-layered	Present	Blue-green	3–4	Sciuto and Moro, 2016
PCC 8501	Hot spring	0.8–1.8	2–6.5	Unlayered	Absent	Blue-green	4–6	Sciuto and Moro, 2016

TABLE 2 | The effect of various carbon and nitrogen sources on the growth of *Thermoleptolyngbya sichuanensis* A183.

Growth medium composition								
BG-11	100 mM $NaHCO_3$	300 mM $NaHCO_3$	500 mM $NaHCO_3$	85 mM $NaNO_3$	3 mM urea	6 mM urea	10 mM Na_2SO_4	10 mM $NaHSO_3$
+	+++	++	=	+++	+	−	+	−

"+" *Means that A183 can grow, and the more "+", the better the growth; "−" means that it cannot grow; " = " means no change in growth with respect to BG-11 control.*

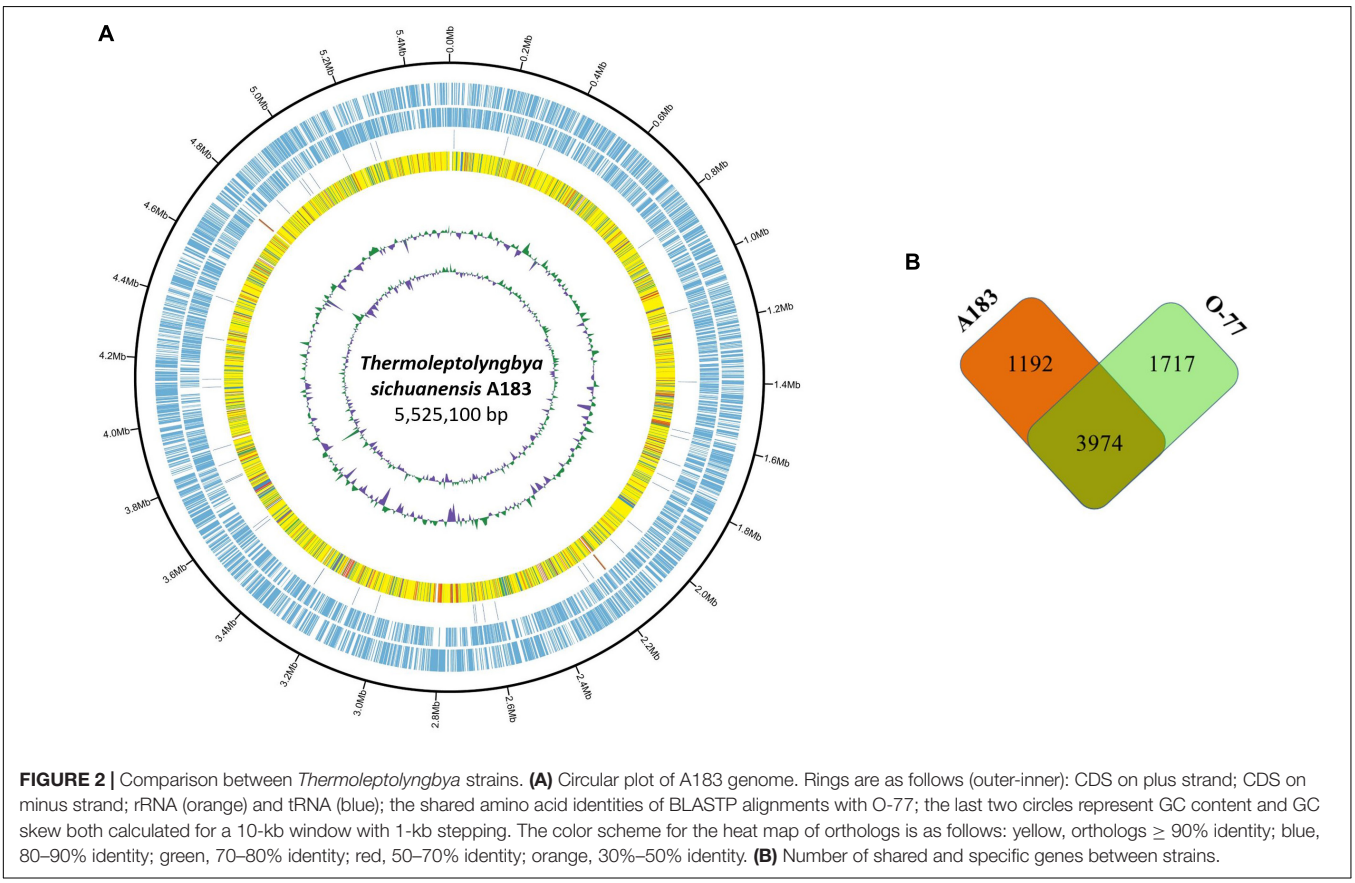


FIGURE 2 | Comparison between *Thermoleptolyngbya* strains. **(A)** Circular plot of A183 genome. Rings are as follows (outer-inner): CDS on plus strand; CDS on minus strand; rRNA (orange) and tRNA (blue); the shared amino acid identities of BLASTP alignments with O-77; the last two circles represent GC content and GC skew both calculated for a 10-kb window with 1-kb stepping. The color scheme for the heat map of orthologs is as follows: yellow, orthologs ≥ 90% identity; blue, 80–90% identity; green, 70–80% identity; red, 50–70% identity; orange, 30%–50% identity. **(B)** Number of shared and specific genes between strains.

previously described genera: *Alkalinema* (Vaz et al., 2015), *Halomicronema* (Abed et al., 2002), *Haloleptolyngbya* (Dadheech et al., 2012), *Kovacikia* (Miscoe et al., 2016), *Leptolyngbya sensu stricto* (Taton et al., 2010), *Nodosilinea* (Perkerson et al., 2011), *Oculatella* (Zammit et al., 2012), *Pantanalinema* (Vaz et al., 2015), *Phormidesmis* (Komárek et al., 2009), *Plectolyngbya* (Taton et al., 2011), *Stenomitos* (Miscoe et al., 2016), *Thermoleptolyngbya* (Sciuto and Moro, 2016), and *Gloeobacter* as an outgroup of the tree. Strain A183 investigated in this study closely clustered with 15 strains affiliated to genus *Thermoleptolyngbya*. Three clusters were not assigned to the previously described taxa and were marked as cluster A–C, respectively. In addition, the sequences of "*L. antarctica*" ANT.L67.1 and *L. indica* LKB did not collocate with any cluster and were placed in separate branches.

The sequence identities of 16S rRNA gene were calculated between A183 and other strains phylogenetically grouped into *Thermoleptolyngbya* clade (**Supplementary Table 2**). The A183 strain showed sequence identities ranging from 97.11 to 99.46%, compared to the other strains assigned to *Thermoleptolyngbya* clade. Notably, numerous strains within *Thermoleptolyngbya* clade were labeled by uncertain genus names (**Figure 3**), while the sequence identities of 16S rRNA gene (**Supplementary Table 2**) strongly indicated that these strains are members of the genus *Thermoleptolyngbya*, according to the recommended

TABLE 3 | Genome features of *Thermoleptolyngbya* strains A183 and O-77.

	A183	O-77
Accession number	CP053661	AP017367
Isolation source	Hot spring, Sichuan Province, China	Hot spring, Kumamoto, Japan
Temperature	40.8°C	Not specified (45–80°C range)
Size (bp)	5,525,100	5,480,261
Chromosome	1	1
GC content (mol%)	56.4	55.9
rRNA operons	2	2
tRNA	45	45
Number of subsystems	266	272
Coding sequences	5166	5691

threshold for bacterial species (98–99%) or genera (94.5–95%) demarcation (Rodriguez-R et al., 2018). Moreover, these *Thermoleptolyngbya* strains originated from thermal environments worldwide (**Supplementary Table 2**; Lacap et al., 2007; Oren et al., 2008; Peng et al., 2013; Nakamori et al., 2014; Gaisin et al., 2015; Bravakos et al., 2016; Sciuto and Moro, 2016; Heidari et al., 2018; Tang et al., 2018b), except for strain XAN 1 and CENA538. This is consistent with the general knowledge that organisms belonging to *Thermoleptolyngbya* appear to have thermal origins (Sciuto and Moro, 2016). Information on XAN 1 is sparse, and perhaps it was an inhabitant of hydrothermal spring based on the submission title on NCBI. CENA538 isolated from saline-alkaline Lake during the Brazilian dry season was subjected to desiccation periods, and the temperature and high salinity of the sampling site could be considered as a thermal environment (Andreote et al., 2014).

The results of phylogenetic reconstruction (**Figure 3**) and sequence identity (**Supplementary Table 2**) indicated that 16S rRNA gene might not effectively differentiate the species-level relationship of strains belonging to *Thermoleptolyngbya*. A verified example was the establishment of two different *Thermoleptolyngbya* species (*T. oregonensis* and *T. albertanoae*) (Sciuto and Moro, 2016), whereas the phylogenetic relationship (**Figure 3**) and sequence identity (99.17%) of the two strains were susceptible to reach an erroneous conclusion on species differentiation. Therefore, a polyphasic approach encompassing morphological, molecular, and phylogenetic analysis of more genomic loci is inordinately crucial for accurate species identification within *Thermoleptolyngbya*.

Phylogeny of 16S-23S ITS

The phylogenetic reconstruction based on full-length 16S-23S ITS sequences (**Figure 4**) was consistent with the phylogeny of the 16S rRNA gene, although fewer sequences were included in the analysis. The *Alkalinema* clade was rooted as an outgroup. Strains ascribable to *Thermoleptolyngbya* were placed into a well-supported clade and showed evident genetic divergence as revealed by branch length (**Figure 4**), indicating that there might be six species within genus *Thermoleptolyngbya*. The 16S-23S ITS tree also detected clades corresponding to previously described genera supported by robust bootstrap values, namely *Alkalinema*, *Kovacikia*, *Leptolyngbya sensu stricto*, *Oculatella*, *Phormidesmis*, *Plectolyngbya*, and *Stenomitos*.

There are two primary reasons attributed to fewer sequences included in the 16S-23S ITS phylogenetic analysis than that in 16S rRNA analysis. Firstly, the 16S-23S ITS sequences are unavailable for many strains. In this study, only half of *Thermoleptolyngbya* strains had both of their 16S rRNA and 16S-23S ITS sequences determined, hindering further comprehensive taxonomic recognition. Secondly, more importantly, the 16S-23S ITS marker is highly variable and is difficult to be aligned precisely if distantly related taxa were included in the database, affecting the outcome of phylogenetic reconstructions. It has been suggested that reliable phylogenies of 16S-23S ITS can be achieved by limiting the analysis to highly related taxa and with the support of secondary structure analysis (Johansen et al., 2011, 2014). However, to ascertain the species identity, a more accurate analysis at molecular level is required, utilizing MLSA and eventually whole genome sequence comparison.

The 16S-23S ITS marker contains highly variable regions and highly conserved domains. The sequence identities are distinct when using full-length ITS or regions/domains individually, probably bringing about misleading information without secondary structure domain analysis. This speculation has been manifested by previous reports (Johansen et al., 2011; Sciuto and Moro, 2016). Additionally, the hyper-variable region of 16S-23S ITS marker is nearly neglected in the phylogenetic analysis since gapped positions are excluded. Those regions, however, may be informative at the species level. Therefore, secondary structure domain analysis of 16S-23S ITS is essential as a complement to ultimate taxonomy determination and is beneficial for better resolving the taxonomic status of problematic strains with inconsistent phylogenetic inferences between 16S rRNA and 16S-23S ITS.

Secondary Structures of 16S-23S ITS

Hypothetical secondary structures of domains within ITS were estimated for strain A183 and representative strains from genus *Thermoleptolyngbya* in the 16S-23S ITS tree (**Figure 5**). Excluding two highly conserved tRNAs from full-length ITS sequences, the length of the remaining ITS sequences varied greatly from 297 to 535 bp (**Table 4**). The remaining ITS sequence of A183 strain was the longest among *Thermoleptolyngbya* strains, 535 bp in length. Identical sequences were observed in conserved domains D3 (GGTTC), boxA (GAACCTTGAAAA), and D4 (CTGCATA) among all *Thermoleptolyngbya* strains. Conserved domain D2 showed four sequence types with slight nucleotide difference, namely CTTCCAAACTAT in A183, O-77 and SHAFA S1B clone cl2, CTTCCAAGCTAG in ETS-08, CTTCCAAACTGT in CY11, and TTTCCAAACTAT in PCC 8501.

The inferred D1-D1' helices were drawn in **Figure 5A**. All helices exhibited the same length and nearly identical structures. However, nucleotide differences of D1-D1' primary structures were found among strains. From basal stem (GACCU-AGGUC) onward, all the helices were structured by seven residue asymmetrical loops followed by three residue stems, two residue symmetrical loops

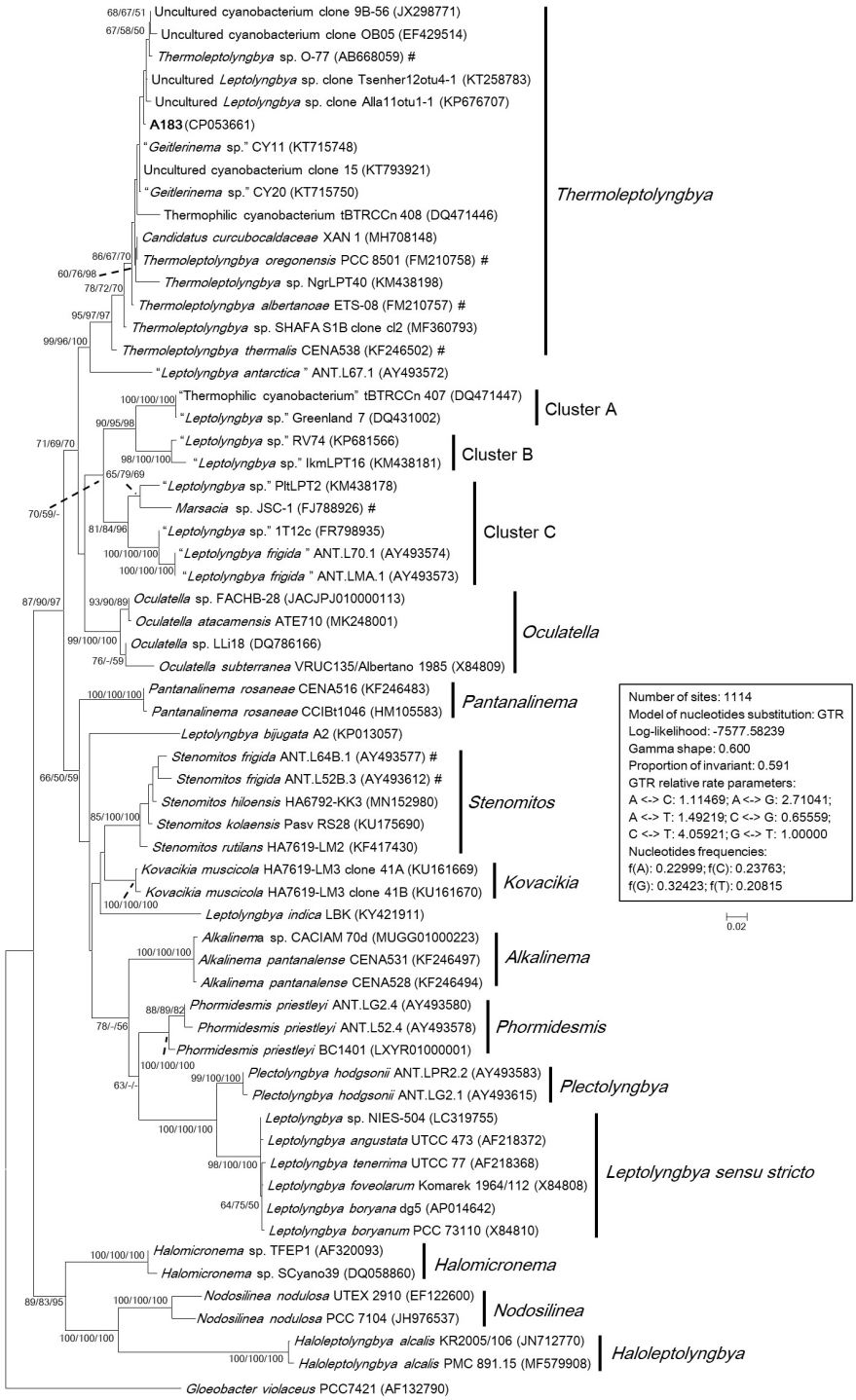


FIGURE 3 | Maximum-likelihood phylogenetic tree of 16S rRNA gene sequences. The numbers at nodes refers to the support values of ML, MP, and NJ analysis, respectively. Only bootstrap values > 50% (1000 non-parametric replications) are indicated at nodes. Genera or cluster are indicated by vertical bars. Scale bar = 2% substitutions per site. Strains in quotation marks have uncertain genus names. Strains marked with # have been recently reclassified or proposed to be reclassified.

followed by five residue stems, and one base left bulge followed by five residue stems, nine residue asymmetrical loops, two residue stems, and five residue asymmetrical loops. The only outlier of this structure is SHAFA S1B clone cl2 where seven residue asymmetrical loops and three residue stems are followed by four residue stems and four residue symmetrical loops; all the remaining parts of the structure were unchanged.

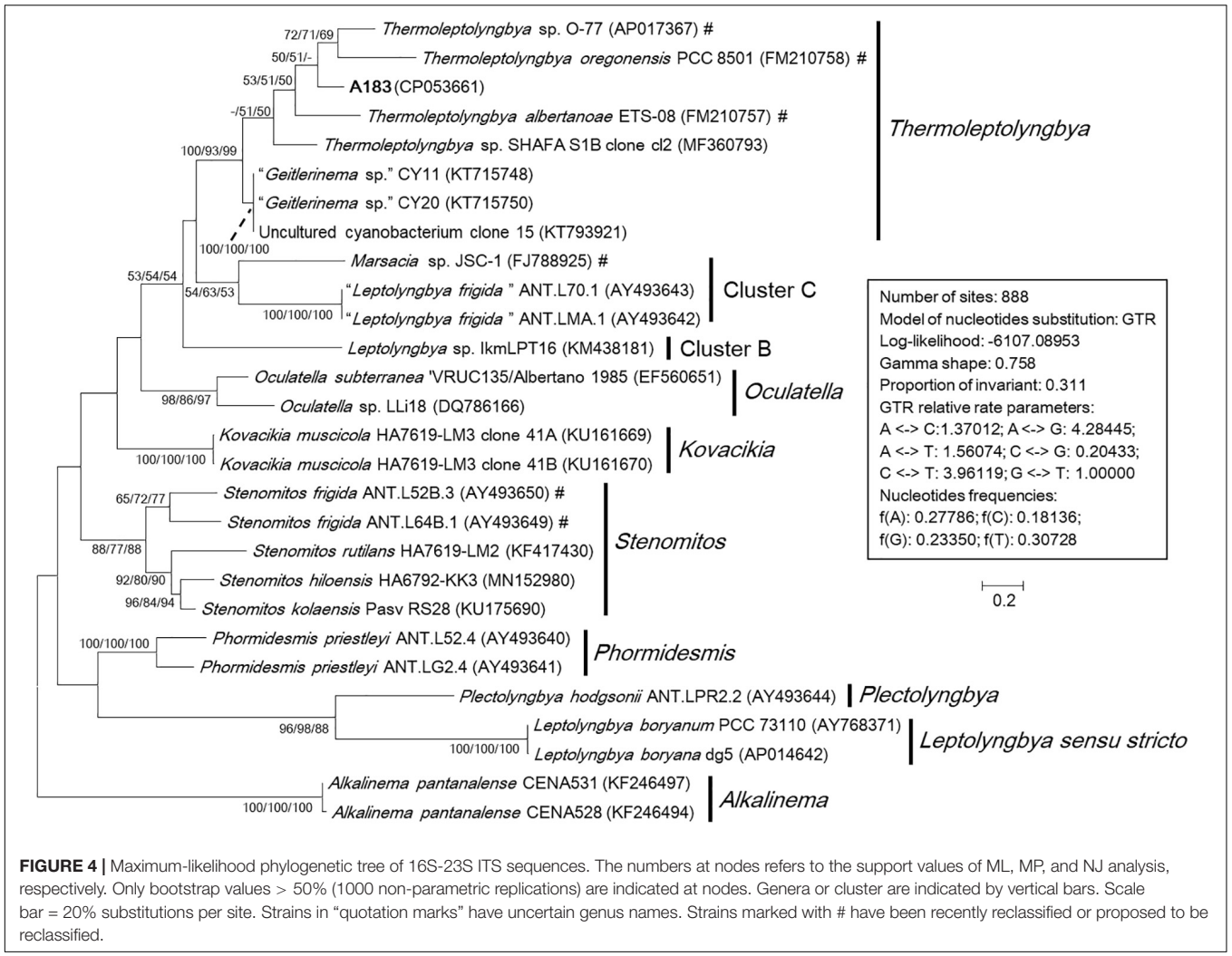


FIGURE 4 | Maximum-likelihood phylogenetic tree of 16S-23S ITS sequences. The numbers at nodes refers to the support values of ML, MP, and NJ analysis, respectively. Only bootstrap values > 50% (1000 non-parametric replications) are indicated at nodes. Genera or cluster are indicated by vertical bars. Scale bar = 20% substitutions per site. Strains in "quotation marks" have uncertain genus names. Strains marked with # have been recently reclassified or proposed to be reclassified.

Hypothetical V2 helices were tremendously distinct among *Thermoleptolyngbya* strains (**Supplementary Figure 3**), and no common basal structure was found. The highly variable helices were attributed to the divergent primary structures of V2 regions. The longest V2 helix was found to be 218 residues in strain A183, followed by 211 residues in strain O-77, 152 residues in ETS-08, and 75 residues in PCC 8501, respectively (**Table 4**). The V2 helices of the remaining two strains were only 11 and 39 residues in length (**Table 4**).

The depicted boxB helices (**Figure 5B**) indicated that a basal stem structure (AGCA-UGCU) was shared by all *Thermoleptolyngbya* strains. Although A183 had similar residue length to ETS-08 and CY11 (**Table 4**), the boxB helix structure of A183 was clearly distinct from that of the two strains. The boxB helix of A183 was mainly composed of a stem orderly fragmented by three residue asymmetrical loops, single base left bulge, two residue symmetrical loops, three residue asymmetrical loops, and terminated with seven residue hairpin loops. The boxB helices of the other *Thermoleptolyngbya* strains (**Figure 5B**) all terminated with hairpin loops variable in residues sequence and length, while the main stem structure of boxB helices considerably varied among strains in single base right bulge (in EST-08, PCC 8501, O-77, and SHAFA S1B clone cl2), single base left bulge (in EST-08, PCC 8501, O-77, and CY11), two base left bulge (in PCC 8501), asymmetrical loop (in EST-08, PCC 8501, SHAFA S1B clone cl2, and CY11), and symmetrical loop (in O-77).

The V3 helices shared a basal stem structure (GUC-GAC) among all *Thermoleptolyngbya* strains (**Figure 5C**). The V3 helix of A183 comprised two asymmetrical loops, two symmetrical loops, single base left bulge, fragmented stems, and terminated with four residue hairpin loops. Although all the V3 helices showed the same helix length (**Table 4**), the structures (**Figure 5C**) differed from each other in terms of bulge, loop, and stem. Unfortunately, V3 helix of SHAFA S1B clone cl2 was not inferred due to incomplete sequences in this region.

In summary, the secondary structures of V2, boxB, and V3 undoubtedly differentiate A183 from the other *Thermoleptolyngbya* strains, whilst the structure of the D1-D1' domain remains conserved. The result of the 16S-23S ITS secondary structure analysis is in agreement with the phylogenetic reconstructions of 16S-23S ITS, confirming the verification of the A183 strain as a new species of

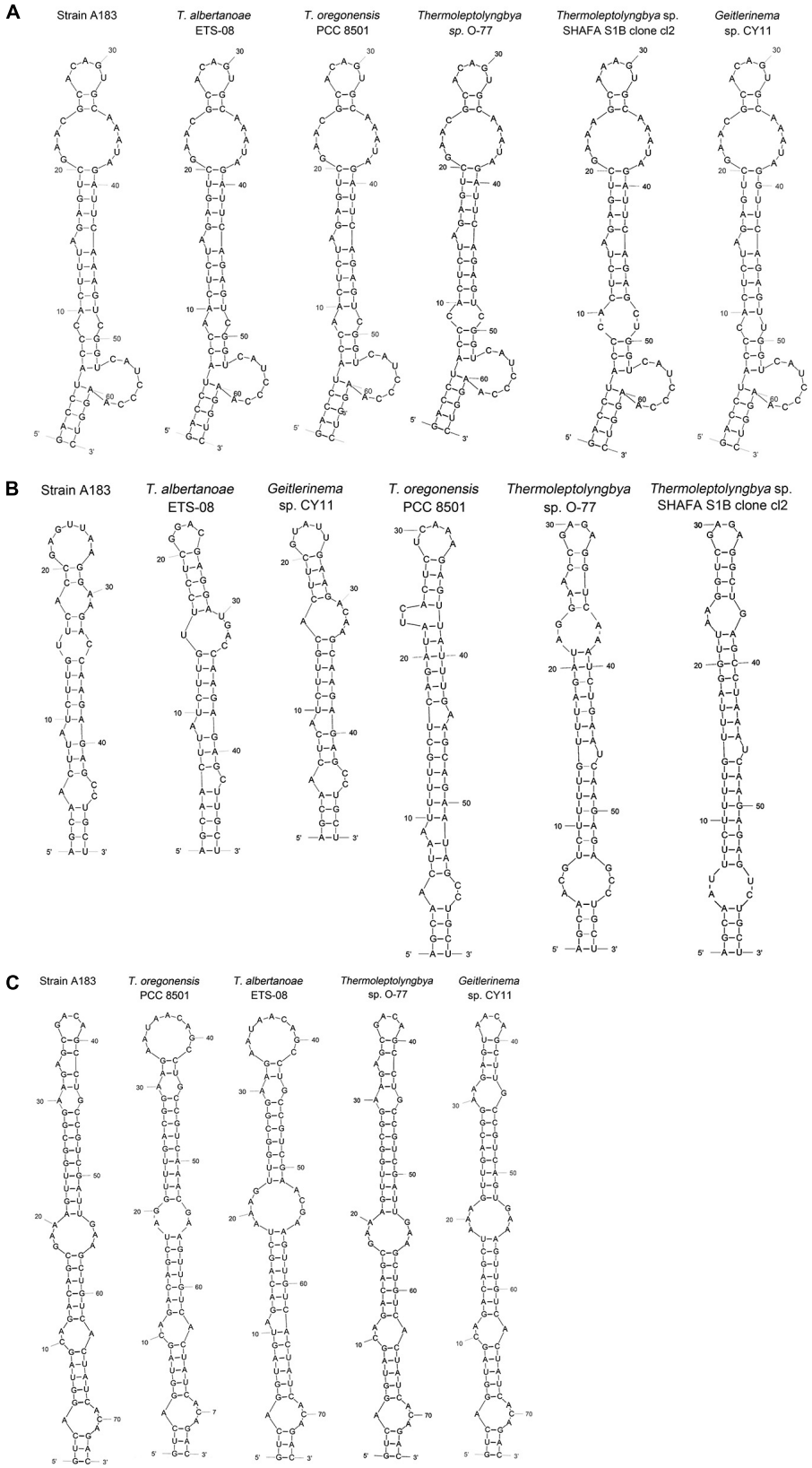


FIGURE 5 | Predicted secondary structures of D1-D1' helix **(A)**, boxB **(B)**, and V3 helix **(C)** of 16S-23S ITS of *Thermoleptolyngbya* strains.

TABLE 4 | The length (bp) summary of regions within 16S-23S ITS of strains studied.

Strain	ITS length (tRNA removed)	D1-D1' helix	D2	D3	boxA	D4	V2 helix	boxB helix	V3 helix
A183	535	64	12	5	12	7	218	48	74
Thermoleptolyngbya albertanoae ETS-08	441	64	12	5	12	7	152	47	74
Thermoleptolyngbya oregonensis PCC 8501	344	64	12	5	12	7	75	60	74
Thermoleptolyngbya sp. O-77	486	64	12	5	12	7	211	60	74
Thermoleptolyngbya sp. SHAFA S1B clone cl2	358	64	12	5	12	7	39	60	NA
Geitlerinema sp. CY11	297	64	12	5	12	7	11	48	74

NA, not available.

Thermoleptolyngbya. The secondary structure analysis of V2, boxB, and V3 helices appeared to be effective for species-level identification. Although the V2 helix was the most variable, it was the least taxonomic-informative in light of its high variability and absence in some cyanobacterial strains (Iteman et al., 2000; Sciuto and Moro, 2016). Although it was reported that D1-D1' helix, compared to boxB and V3 helix, is more important for interspecies discrimination within a given genus (Perkerson et al., 2011; Vieira Vaz et al., 2015), it turned out to be not that effective in the case of *Thermoleptolyngbya*. The utilization of boxB and V3 for species distinction within *Thermoleptolyngbya* has been verified by the successful differentiation of *T. albertanoae* and *T. oregonensis* (Sciuto and Moro, 2016).

Interestingly, phylogenetic analysis and secondary structure analysis of 16S-23S ITS both indicated that *Thermoleptolyngbya* strains listed in **Table 4** are probably different species to each other within the genus *Thermoleptolyngbya*, even though the phylogeny (**Figure 2**) and sequence identity (**Supplementary Table 2**) of 16S rRNA did not initially reveal such differentiation. Undoubtedly, detailed information regarding morphology and DNA sequence of more loci or complete genome would be essential for the taxonomic revision of SHAFA S1B clone cl2 and CY11within genus *Thermoleptolyngbya*.

Comparative Genome Analyses

As shown in **Table 5**, results of genome-wide ANI and AAI conformed to the suggested values for species (ANI > 96%, AAI ≥ 95%) and genus (ANI < 83%, AAI ≤ 70%) delimitation (Walter et al., 2017; Jain and Rodriguez, 2018), again confirming the taxonomy delineation of strain A183 as a novel species of *Thermoleptolyngbya*. Moreover, the concatenated alignment of 751 single-copy genes from the complete genomes of related strains produced an ML genomic tree with 100% bootstrap support (**Figure 6**), the topology of which was consistent with that of 16S rRNA and ITS. All the results verified the conclusion that strain A183 was a novel species of *Thermoleptolyngbya*. Unfortunately, the unavailability of genomes led to the failure of a comprehensive snapshot of divergences in genomes among all strains related to A183. To compensate for this shortcoming, partial data extracted from the metagenomic bins that show resemblance to *Thermoleptolyngbya* have been also analyzed with the combination of MLSA (**Supplementary Figure 4**), patristic distance (**Supplementary Table 3**), and *in silico* DNA-DNA hybridization (**Supplementary Table 4**) approaches. Provided analyses further reinforce the claim that the newly isolated strain is a novel species within the *Thermoleptolyngbya* genus. The A183 strain clearly separates from previously described O-77 and PCC 8501 strains. Interestingly, two recently deposited metagenomic bins (C42_A2020 and M55_K2018) obtained from the hot springs of Cahuelmo, Chile (temperature 42°C, pH 9.2) and Chhattisgarh, India (temperature 55°C, pH 7.9) show sequences of closer resemblance to A183 than either O-77 or PCC 8501. The lack of fully sequenced genomes and some key phylogenetic markers such as 16S or ITS as well as morphological data makes this analysis somehow incomplete and prevents them from being unequivocally described as members of *Thermoleptolyngbya*. These results, however, support delineation of a new species around A183, and an implied global distribution of *Thermoleptolyngbya*, indicating also potentially novel, closely related filamentous cyanobacteria.

To have an in-depth understanding of the A183 strain functioning in its ecosystem and to clarify the genetic features of the strain, a comparative genome analysis was performed between the A183 strain and its relative *Thermoleptolyngbya* (O-77). Generally, the two strains shared a similar genome size, GC content, and the number of rRNA operons and tRNAs, but a significant difference in CDS number (**Table 3**). The subsystem category distribution was also almost identical to both strains (**Supplementary Figure 5**). Analogously, gene ontology (GO) analysis of all CDS showed a similar distribution of functional categories in both strains (**Supplementary Figure 2**).

An ortholog table (**Supplementary Table 5**) was constructed based on all-against-all BLASTP alignment. As indicated in **Figure 2B**, strain A183 shared 3,974 gene-encoding homologous sequences with strain O-77, accounting for 76.9% annotated genes of the A183 genome. Unique genes were also found in each strain, namely 1,192 genes in A183 and 1,717 genes in O-77 (**Figure 2B**). Further, GO analysis of the unique genes showed that these genes were distributed in a wide range of functional category (**Supplementary Figure 6**).

No intact prophage region was detected in either strain. Three and 11 CRISPR arrays were identified in the genomes of A183 and O-77, respectively (**Supplementary Table 6**). A higher number of spacers (597) was observed in the O-77 genome, approximately 3.5 times higher than that of A183. All direct repeats of the CRISPR arrays were 35 bp long in the A183 genome, while there were two kinds of length (34 and 35 bp long) in the O-77 genome. The A183 had only one CRISPR array with adjacent Cas genes. Although O-77 had two CRISPR arrays with Cas genes, one of them did not have a full set of genes encoding the Cas system

TABLE 5 | Summary of ANI (Average Nucleotide Identity) and AAI (Average Amino acid Identity) between cyanobacterial genomes studied.

Strain	A183	O-77	70d	dg5	JSC-1	FACHB-28	BC1401
A183	100	93.59	59.18	60.24	64.54	64.59	61.36
Thermoleptolyngbya sp. O-77	89.97	100	58.64	59.75	64.12	63.99	60.95
Alkalinema sp. CACIAM 70d	74.88	75.95	100	61.17	59.35	58.18	62.17
Leptolyngbya boryana dg5	75.80	76.35	76.91	100	60.51	60.58	66.71
Marsacia sp. JSC-1	75.55	74.88	79.21	75.05	100	64.96	61.34
Oculatella sp. FACHB-28	74.12	74.00	74.83	76.25	75.06	100	62.73
Phormidesmis priestleyi BC1401	75.25	74.57	74.39	75.05	73.85	74.52	100

The numbers above and below the diagonal indicate the AAI and ANI values (%), respectively.

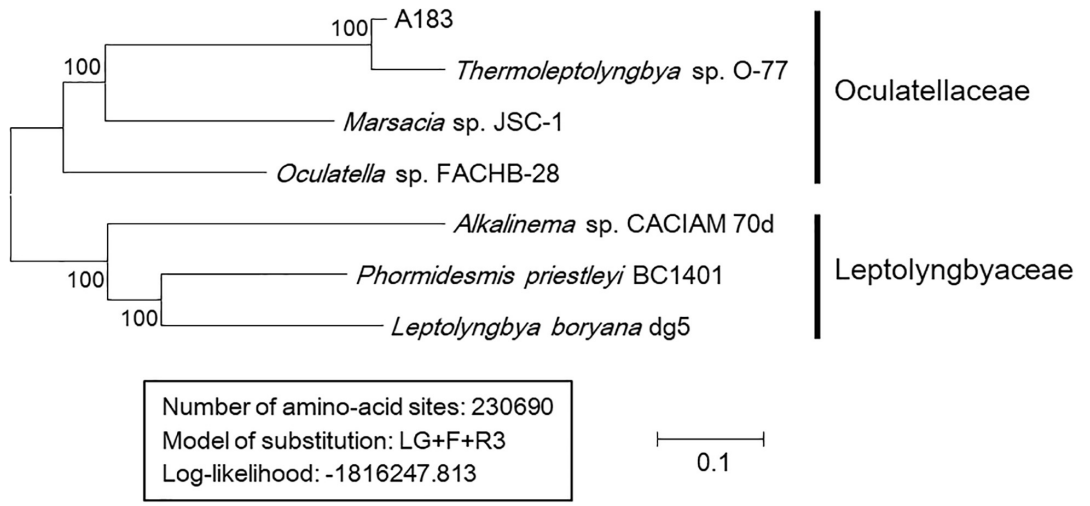


FIGURE 6 | Maximum-likelihood phylogenomic tree of concatenated protein alignment of 751 single-copy genes shared by all genomes. Bootstrap values (1000 replications) are indicated at nodes. Scale bar = 10% substitutions per site.

and was not assigned to a specific CRISPR-Cas system. The type III-D CRISPR-Cas system in both strains might confer resistance to foreign mobile genetic elements from bacteriophage or viruses (Makarova et al., 2011).

Thermotolerance

A183 was originally isolated from the hot spring with a temperature of 40.8°C (Tang et al., 2018b) and is capable of growing at a maximum of 50°C (Tang et al., 2018a). A survival strategy must be well-prepared for A183 to survive in these thermal environments. It is known that heat shock proteins (Hsps) are essential for maintaining and restoring protein homeostasis. In the A183 genome, dozens of genes were identified to encode heat shock proteins (Hsps), including the Hsp100, Hsp90, Hsp70, and Hsp60 family as well as small Hsps.

Hsps belonging to Hsp100 family may favor the refolding, disaggregation, and removal of heat-damaged proteins. For instance, ATP-dependent protease *clpX*, associated with *clpP*, promoted disassembly and degradation of heat-aggregated substrates (e.g., *ftsZ*) (Labreck et al., 2017). Homologs of the *clp* family were found in the A183 genome: *clpB*, *-C*, *-P*, *-S*, *-X* (TS0414, TS0934, TS1039, TS1204-1206, TS1331, TS1444, TS1861-1862, TS2363, TS3986, and TS4547). An identical composition and distribution of *clp* genes were found in the O-77 genome (**Supplementary Table 1**).

There was only one copy of the *htpG* gene (TS0820) in A183, and so was O-77 (**Supplementary Table 1**). The *htpG* protein of the Hsp90 family was suggested to be more of a general stress protein in that it played a role in several abiotic stresses (Hossain and Nakamoto, 2002, 2003). Particularly, the *htpG* primarily protected the photosynthetic machinery from heat stress in cyanobacteria (Sato et al., 2010).

Proteins of Hsp70 family appeared to be prosperous in the A183 genome, mainly including *dnaK* (TS1830, TS2881, TS4129, and TS4674) and *dnaJ* (TS1379, TS2355, TS2878, TS4052, TS4637, TS4910, and TS5059). Multiple copies of *dnaK* and *dnaJ* genes have also been reported in other cyanobacteria (Rajaram et al., 2014). Nevertheless, it was suggested that *dnaK* and *dnaJ* proteins might function differently, and only some of them contributed to thermotolerance (Schneider, 2010; Duppre et al., 2011). Moreover, the gene *grpE* (TS2882) was found, the protein encoded by which might act as a cofactor of Hsp70 family and participate actively in response to heat shock by preventing the aggregation of stress-denatured proteins (Schneider, 2011). The homologs of Hsp70 family in O-77 showed a high similarity to that in A183 (**Supplementary Table 1**).

Interestingly, two distinct *groEL* genes (TS0663, TS2946) of the Hsp60 family, also referred to as the *groE* chaperone machinery, were identified in the A183 genome. One of them (TS2946, *groEL*-1) formed *groESL* operon together with the small Hsp *groES* (TS2947). The *Gloeobacter* PCC 7421 genome contains two *groEL* genes, both of which have *groES* immediately upstream of each *groEL*. Therefore, it was speculated that one of the two *groESL* operons had lost its *groES* during the evolutionary cyanobacterial diversification. The *HrcA* repressor system (TS4519) was also found in the A183 genome, which may negatively regulate the expression of *groE* genes (Nakamoto et al., 2003). A similar composition of *groE* genes was observed in the O-77 genome.

In addition to the genes mentioned above, genes encoding small Hsps and proteases were also identified in the A183 genome (**Supplementary Table 1**) and might be involved in thermotolerance. For example, small *Hsp16* (TS0692), also referred to as *hspA*, may act as a chaperone and interact with dozens of proteins at high temperature, playing multiple roles ranging from protein folding to stabilization of thylakoid and periplasmic membranes (Basha et al., 2004); the *ftsH* protease (TS0606, TS1376, TS1741, TS1916, and TS2838) may be responsible for heat-induced degradation of photodamaged D1 protein by up-regulated expression of *ftsH* (Kamata et al., 2005). However, further, detailed investigations on the actual functions of these genes are necessary to elucidate the mechanisms of adaptation of A183 to high temperatures. Based on previous results concerning the thermophilic cyanobacteria from this region's thermal springs it could be concluded that high temperatures are physiological to these strains and the putative thermotolerance genes are expressed constitutively (results not published). In order to fully elucidate the impact of these genes on strain thermostability, a loss of function study, such as targeted gene knock-out or silencing, will be needed. This in turn, requires development of the entire repertoire of genetic tools that is currently unavailable for *Thermoleptolyngbya* sp. and related strains.

Signal Transduction

The two-component regulatory systems have been commonly observed in cyanobacteria and elucidated for the perception of environmental stress and the subsequent transduction of stress signals (Los et al., 2010). In the A183 genome, 36 and 31 genes were identified to encode histidine kinases and response regulators, respectively (**Supplementary Table 1**). The system composed of these genes is likely to play major roles in the core part of acclimation to changing environments. However, the scattered distribution of these genes in the A183 genome hindered the association of genes for histidine kinases with their respective cognate response regulators. This was in sharp contrast to cases in *E. coli* (Bourret and Silversmith, 2010) or *Bacillus subtilis* (Aguilar et al., 2001) that the genes for a single two-component system were organized into operons or located close one to another. Thus, investigation on sensors and their cognate regulators requires individual mutation on these genes. Similarly, the scattered genes encoding histidine kinases and response regulators were also found in the O-77 genome (**Supplementary Table 1**).

The serine/threonine protein kinases (Spks) had similar purposes as two-component systems for signal transduction. A total of 17 genes were identified in A183 genome to encode Spks (**Supplementary Table 1**). Unfortunately, the functions of only several Spks in *Synechocystis* have been characterized to date, such as *spkA* and *spkB* involved in the control of cell motility, and *spkE* involved in the regulation of nitrogen metabolism (Zhang et al., 2006). Although the Spks proteins were conserved among *Thermoleptolyngbya* A183 and O-77 strains as revealed by their high similarity of amino acid sequences (87.61–96.71%, **Supplementary Table 5**), the Spks proteins of *Thermoleptolyngbya* strains were quite divergent from those of other cyanobacterial strains. These putative Spks proteins in *Thermoleptolyngbya* are required to be more carefully investigated in the future.

Different environmental conditions or developmental signals often cause major changes in transcription pattern by inducing a swap of sigma factors in the RNA polymerase holoenzyme (Los et al., 2010). The A183 genome inhabited seven genes encoding Group 2 sigma factor *sigD* (TS0188, TS0493, TS1286, TS1338, TS2057, TS3542, and TS4081) and two genes encoding Group 3 sigma factor *sigF* (TS1335 and TS3418). The *sigD* is the only *sig* gene that produced moderate amounts of transcripts in the dark and was not affected by any of the stress treatments (Tuominen et al., 2003), suggesting its crucial role in transcription regulation particularly in adverse conditions. The exact function of *sigD* in cyanobacteria remains to be clarified.

In addition, seven genes encoding GGDEF/EAL domain proteins were observed in the A183 genome (**Supplementary Table 1**). It was reported that GGDEF/EAL domain proteins function as diguanylate cyclase/phosphodiesterase that synthesizes/degrades cyclic di-GMP and participate in a cyclic-di-GMP signaling pathway that may regulate biofilm formation, motility, virulence, and cell cycle (Agostoni et al., 2013). A recent study showed that GGDEF/EAL domains were also involved in blue-light-induced cell aggregation in *Thermosynechococcus* BP-1 and NIES-2134 (Enomoto et al., 2015). Interestingly, *Thermosynechococcus* genomes have 9–13 genes encoding GGDEF/EAL domains, while hot-spring *Synechococcus* genomes (strain JA-3-3-Ab and JA-2-3Ba) only have four genes (Cheng et al., 2020). These data implied that the complexity of cyclic-di-GMP signaling pathways appeared to be distinct among thermotolerant strains.

Carbon Assimilation

Cyanobacteria in hot springs have to deal with many environmental stresses, one of which is low CO_2 solubility at high temperatures. The CO_2-concentrating-mechanism cyanobacteria have evolved into can partially alleviate this problem by actively transporting and accumulating inorganic carbon (Ci: CO_2 and HCO_3^-) for the sake of a satisfactory rate of CO_2 fixation under carbon-limiting concentrations (Price et al., 2008). The uptake of gaseous CO_2 systems in cyanobacteria relied on NADPH dehydrogenase (NDH-1) complexes (Price et al., 2008). In the A183 genome, there were

19 NDH-1 genes encoding *ndhA*, *ndhB*, *ndhC*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*, *ndhL*, and *ndhM*, respectively (**Supplementary Table 1**). Except for *ndhE* and *ndhF*, the other NDH-1 genes were present in one copy. Meanwhile, several genes clustered together (*ndhE-ndhG-ndhI-ndhA*, TS1642-1645; *ndhJ-ndhK-ndhC*, TS2756-2758; *ndhF-ndhE*, TS4963-4964), while *ndhE* and *ndhF* were composed of clusters coupled with low-affinity CO_2 hydration proteins, namely *ndhF-ndhE-cphY* (TS4168-4170) and *ndhF-ndhE-cphX* (TS4275-4277). These gene clusters probably contribute to constitutively expressed NDH-1 complex involved in low-affinity CO_2 uptake (Shibata et al., 2001). Similar clusters were also found in the O-77 genome.

Our previous study showed that *Thermoleptolyngbya* A183 survived at the concentration exceeding 0.5 M $NaHCO_3$ (Tang et al., 2018a), suggesting that the A183 strain can assimilate bicarbonate and convert it to CO_2 for photosynthesis. This was further evidenced by the HCO_3^- uptake systems, as suggested by the genome analysis. First, a homolog (TS0333) of a low-affinity, high flux, Na^+-dependent bicarbonate transporter (*bicA*) was found in the A183 genome. Second, the A183 genome also harbored a homolog (TS2219) of *sbtA*, an inducible, high-affinity Na^+-dependent bicarbonate transporter (Shibata et al., 2002). The transporter shows 69.3% sequence identity to *sbtA* of other thermophilic cyanobacterium *Thermosynechococcus lividus* PCC6715 (Liang et al., 2019). The different bicarbonate uptake systems might be flexibly utilized by A183 to meet the conditional demand for carbon gain. Interestingly, the *BCT1* bicarbonate transporter typical for many other cyanobacteria was not detected in the A183 genome, suggesting a difference regarding a bicarbonate uptake mechanism from *Thermosynechococcus* BP-1 that lacks *sbtA* entirely but is equipped with *BCT1* (Price et al., 2008). In addition, three ABC-type bicarbonate transporters were observed in the A183 genome. The O-77 genome also possessed genes encoding *BicA* and *sbtA* and showed high protein similarities with A183 (94.99 and 95.56%, respectively), but the number of genes encoding an ABC-type bicarbonate transporter was higher than that of A183 (9 vs. 3). These results are consistent with the physiological characterization of the strain (**Table 2**).

Nitrogen Assimilation

In the A183 genome, eight genes encoding nitrogenases were identified, including *nifB*, *nifR*, *nifH*, *nifO*, *nifW*, and *nifX* (**Supplementary Table 1**). These genes encode proteins required for catalytic activity, Fe–Mo cofactor biosynthesis, and maturation and stability of the nitrogenase protein complex (Steunou et al., 2008). O-77 exhibited similar gene components regarding nitrogenases. These results suggested that both strains are nitrogen-fixing non-heterocystous cyanobacteria. N_2 fixation is an energetically expensive metabolic reaction catalyzed by nitrogenase, which is inhibited by O_2 generated during photosynthesis (Steunou et al., 2008). Therefore, *Thermoleptolyngbya* A183 also exhibits alternative strategies in light of an economy of nitrogen assimilation.

Specific transporters are essential for organisms to concentrate ambient nitrogen sources within cells to survive in oligotrophic aquatic environments (Esteves-Ferreira et al., 2018). In the A183 genome, the ABC-type nitrate transport system (*nrtABCD*, TS4258-4261) was detected and formed an operon with two essential nitrogen-related genes located at both sides of *nrtABCD*, encoding ferredoxin-nitrite reductase (*nirA*, TS4257) and ferredoxin-nitrate reductase (*narB*, TS4262), respectively. This result was consistent with many freshwater cyanobacterial strains (Maeda et al., 2015). Furthermore, a complete gene set (*urtABCDE*) of ABC-type urea transport system and seven genes encoding urease (*ureA* – *ureG*) was observed (**Supplementary Table 1**), suggesting the ability of A183 to import and utilize urea as a nitrogen source. In addition, two genes were identified as ammonium transporter (TS132, TS5152). The O-77 genome showed similar components of nitrogen-related transporters. The above results reinforced the importance of these transporters for cyanobacterial growth in oligotrophic environments and also implied that A183 can depend on multiple forms of nitrogen sources. The ability of both strains to fix molecular nitrogen and utilize urea has been confirmed experimentally (**Table 2**).

Sulfur Assimilation

The A183 strain was isolated from a hot spring abundant in sulfur (Tang et al., 2018b). The analysis of the A183 genome reveals three transporter systems involved in sulfate uptake. First, the sulfate-thiosulfate permease (*sulT*), belonging to the ABC-type transporter, comprised four subunits encoded by the *cysPTW* operon (TS1272-1274) and *cysA* (TS4980) that was located far away from the operon. The O-77 genome showed a different operon in terms of *cysPTW* but a similar distribution of *cysA* gene. The organization of the *cysPTW* operon in the *Thermoleptolyngbya* strains was different from that of *Thermosynechococcus* strains, such as BP-1 and PKUAC-SCTE542, resulting from the replacement of *cysP* by *sbpA* (Liang et al., 2019). The second sulfate permease in the A183 genome was *sulP* (TS1400), encoded by a single polypeptide and functioned as inorganic anion uptake carriers or anion:anion exchange transporters (Aguilar-Barajas et al., 2011). No homolog of *sulP* was identified in the O-77 genome. Besides, sulfate can also be transported by the high-affinity ModABC molybdate transport system (Aguilar-Barajas et al., 2011), and the ModABC transporter in the A183 genome was encoded by the *modABC* operon (TS2382-2384). An analogous ModABC transporter was also found in the O-77 genome. The predominant pathway and substrate specificity need to be experimentally clarified in future research. Physiological studies confirmed the strain's ability of growth in the presence of sulfates (**Table 2**).

CONCLUSION

The polyphasic approach used in this study, including phylogenetic, phylogenomic, ultrastructural, physiological, and morphological surveys, came up with a proposal of a new species, *Thermoleptolyngbya sichuanensis*, and the delineation of this new taxon around the representative strain A183. Although basic phylogenetic analysis and sequence identities of 16S rRNA showed high similarity among *Thermoleptolyngbya* strains, more advanced genome-based approaches strongly confirmed the

delineation of a new species of *Thermoleptolyngbya*. Analysis of metagenome-assembled genomes revealed that the proposed species is not endemic but present in hot springs globally. Meanwhile, comparative genome analysis revealed distinct genome structures of *Thermoleptolyngbya* strains. Moreover, genes related to thermotolerance, signal transduction, and carbon/nitrogen/sulfur assimilation were thoroughly analyzed and partially verified experimentally for illustrating the ability of this strain to adapt to inhospitable niches in hot springs.

Taxonomic Treatment and Description of *Thermoleptolyngbya sichuanensis* Daroch, Tang, and Shah et al. sp. nov.

The classification system that was applied was based on Komarek et al. (2014).

Taxon description in accordance with the prescriptions of the International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen code) (Turland et al., 2018).

Phylum: Cyanobacteria

Order: Synechococcales

Family: Leptolyngbyaceae

Description: The cell morphology of the A183 strain indicated by light microscopy revealed straight, wavy, and occasionally bent trichomes (**Figure 1A**). The SEM and TEM (**Figures 1B–D**) showed that trichomes of strain A183 were unbranched and composed of 80–120 elongated barrel-shaped cells, 1.30–1.60 μm in length and 1.05–1.10 μm in width. Constrictions were observed at the cross-walls of cells (**Figures 1B–D**). Individual cells of the filaments were divided by centripetal invagination of the cell wall (**Figure 1E**). Intracellular connections between vegetative cells were not observed (**Figure 1D**). The TEM analysis also exhibited that the three to five thylakoid layers were located in parallel at the inner periphery of cells (**Figures 1C,D**) and can be described as parietal according to recent classification (Mareš et al., 2019). Sheath, septum, phycobilisome, carboxysomes, cyanophycin granule, and polyphosphate bodies were present in the cytoplasm (**Figures 1C–F**), and small lipid droplets were also observed (**Figure 1C**).

Type strain: is A183 (= FACHB-2491).

Etymology: Species epithet derives from the name of collection site.

Type locality: Thermal springs in Ganzi Prefecture of Sichuan Province, China.

Ecology of type locality: the sample occurred as macroscopic green mat attached to the sinter around the pond with a small amount of mucilage around the entire mat. Sample collection was done in 12.05.2016 with the humidity being close to 71%. Temperature at the time of collection was 15°C and the light intensity was around 1000 lux. The pH of the spring was 6.32 and concentration of total dissolved solids was 447 mmol L^{-1}

Habitat: thermal springs in Ganzi Prefecture of Sichuan Province, China (30°05'14" N, 101°56'55" E) *Thermoleptolyngbya* species exhibiting peak $NaHCO_3$ tolerance as high as 1 M, and 0.5 M during prolonged cultivation. The strain can utilize urea up to a concentration of 3 mM and withstand 10 mM Na_2SO_4, but not an equivalent concentration of Na_2HSO_3. The strain is diazotrophic and exhibits experimentally verified nitrogenase activity with acetylene reduction assay (**Table 2** and **Supplementary Figure 1**).

Holotype here designated: the culture of *Thermoleptolyngbya sichuanensis* was initially denoted and deposited in Peking University Algae Collection as PKUAC-SCTA183 has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession number FACHB-2491 as *Thermoleptolyngbya* species after identification and authentication on the basis of the full-length sequencing of the 16S rRNA gene along with folding of the secondary structures of the 16S–23S ITS region. After proper identification and authentication, the culture is being maintained in the FACHB under the accession number FACHB-2491.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, CP053661.

AUTHOR CONTRIBUTIONS

JT: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing-original draft, writing-review and editing, visualization, supervision, project administration, and funding acquisition. LL and ML: formal analysis, investigation, data curation, and writing-original draft. LD and MMW: methodology, software, and data curation. MS: methodology, validation, and writing-original draft. MW: methodology and writing-review and editing. KW: conceptualization, methodology, and writing-review and editing. MD: conceptualization, methodology, resources, data curation, writing-original draft, writing-review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (31970092 and 32071480), Shenzhen Knowledge and Innovation Basic Research grant number (JCYJ20180302153648993), Key Laboratory of Coarse Cereal Processing (Ministry of Agriculture and Rural Affairs, China) (2019CC12), and Tenure-Track Fund to MD. Funding bodies had no influence over design and execution of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.696102/full#supplementary-material

REFERENCES

Abed, R. M., Garcia-Pichel, F., and Hernández-Mariné, M. (2002). Polyphasic characterization of benthic, moderately halophilic, moderately thermophilic cyanobacteria with very thin trichomes and the proposal of halomicronema excentricum gen. nov., sp. nov. *Arch. Microbiol.* 177, 361–370. doi: 10.1007/s00203-001-0390-2

Agostoni, M., Koestler, B. J., Waters, C. M., Williams, B. L., and Montgomery, B. L. (2013). Occurrence of cyclic di-GMP-modulating output domains in cyanobacteria: an illuminating perspective. *mBio* 4:e00451.

Aguilar, P. S., Hernandez-Arriaga, A. M., Cybulski, L. E., Erazo, A. C., and De, M. D. (2001). Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*. *Embo J.* 20, 1681–1691. doi: 10.1093/emboj/20.7.1681

Aguilar-Barajas, E., Díaz-Pérez, C., Ramírez-Díaz, M. I., Riveros-Rosas, H., and Cervantes, C. (2011). Bacterial transport of sulfate, molybdate, and related oxyanions. *Biometals* 24, 687–707. doi: 10.1007/s10534-011-9421-x

Amarouche-Yala, S., Benouadah, A., Abd, E. O. B., and López-García, P. (2014). Morphological and phylogenetic diversity of thermophilic cyanobacteria in Algerian hot springs. *Extremophiles* 18, 1035–1047. doi: 10.1007/s00792-014-0680-7

Amin, A., Ahmed, I., Salam, N., Kim, B.-Y., Singh, D., Zhi, X.-Y., et al. (2017). Diversity and distribution of thermophilic bacteria in hot springs of Pakistan. *Microb. Ecol.* 74, 116–127. doi: 10.1007/s00248-017-0930-1

Andreote, A. P. D., Vaz, M., Genuario, D. B., Barbiéro, L., and Fiore, M. F. (2014). Nonheterocytous cyanobacteria from Brazilian saline-alkaline lakes. *J. Phycol.* 50, 675–684. doi: 10.1111/jpy.12192

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.

Basha, E., Lee, G. J., Breci, L. A., Hausrath, A. C., Buan, N. R., Giese, K. C., et al. (2004). The identity of proteins associated with a small heat shock protein during heat stress in Vivo indicates that these chaperones protect a wide range of cellular functions. *J. Biol. Chem.* 279, 7566–7575. doi: 10.1074/jbc.m310684200

Bourret, R. B., and Silversmith, R. E. (2010). Two-component signal transduction. *Curr. Opin. Microbiol.* 13, 113–115.

Bravakos, P., Kotoulas, G., Skaraki, K., Pantazidou, A., and Economou-Amilli, A. (2016). A polyphasic taxonomic approach in isolated strains of cyanobacteria from thermal springs of Greece. *Mol. Phylogenet. Evol.* 98, 147–160. doi: 10.1016/j.ympev.2016.02.009

Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-*proteobacteria*: a comparative genomic analysis. *BMC Syst. Biol.* 4:52. doi: 10.1186/1752-0509-4-52

Bruno, L., Dan, I., Bellezza, S., and Albertano, P. (2009). Cytomorphological and genetic characterization of troglobitic Leptolyngbya strains isolated from roman hypogea. *Appl. Environ. Microbiol.* 75, 608–617. doi: 10.1128/aem.01183-08

Chen, Y. B., Dominic, B., Mellon, M. T., and Zehr, J. P. (1998). Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. strain IMS 101. *J. Bacteriol.* 180, 3598–3605. doi: 10.1128/jb.180.14.3598-3605.1998

Cheng, Y.-I., Chou, L., Chiu, Y.-F., Hsueh, H.-T., Kuo, C.-H., and Chu, H.-A. (2020). Comparative genomic analysis of a novel strain of Taiwan hot-spring cyanobacterium thermosynechococcus sp. CL-1. *Front. Microbiol.* 11:82. doi: 10.3389/fmicb.2020.00082

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, 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

Dadheech, P. K., Mahmoud, H., Kotut, K., and Krienitz, L. (2012). Haloleptolyngbya alcalis gen. et sp. nov., a new filamentous cyanobacterium from the soda lake Nakuru, Kenya. *Hydrobiologia* 691, 269–283. doi: 10.1007/s10750-012-1080-6

Debnath, M., Singh, T., and Bhadury, P. (2017). New records of cyanobacterial morphotypes with *Leptolyngbya indica* sp. nov. from terrestrial biofilms of the lower Gangetic Plain, India. *Phytotaxa* 316, 101–120.

Duppre, E., Rupprecht, E., and Schneider, D. (2011). Specific and promiscuous functions of multiple DnaJ proteins in *Synechocystis* sp. PCC 6803. *Microbiology* 157, 1269–1278. doi: 10.1099/mic.0.045542-0

Eckert, E. M., Fontaneto, D., Coci, M., and Callieri, C. (2014). Does a barcoding gap exist in prokaryotes? Evidences from species delimitation in cyanobacteria. *Life* 5, 50–64. doi: 10.3390/life5010050

Enomoto, G., Ni-Ni-Win, W., Narikawa, R., and Ikeuchi, M. (2015). Three cyanobacteriochromes work together to form a light color-sensitive input system for c-di-GMP signaling of cell aggregation. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8082–8087. doi: 10.1073/pnas.1504228112

Esteves-Ferreira, A. A., Inaba, M., Fort, A., Araújo, W. L., and Sulpice, R. (2018). Nitrogen metabolism in cyanobacteria: metabolic and molecular control, growth consequences and biotechnological applications. *Crit. Rev. Microbiol.* 44, 541–560. doi: 10.1080/1040841x.2018.1446902

Gaisin, V. A., Kalashnikov, A. M., Sukhacheva, M. V., Namsaraev, Z. B., Barhutova, D. D., Gorlenko, V. M., et al. (2015). Filamentous anoxygenic phototrophic bacteria from cyanobacterial mats of Alla hot springs (Barguzin Valley, Russia). *Extremophiles* 19, 1067–1076. doi: 10.1007/s00792-015-0777-7

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, 52–57.

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

Heidari, F., Zima, J., Riahi, H., and Hauer, T. (2018). New simple trichal cyanobacterial taxa isolated from radioactive thermal springs. *Fottea* 18, 137–149. doi: 10.5507/fot.2017.024

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hossain, M. M., and Nakamoto, H. (2002). HtpG plays a role in cold acclimation in cyanobacteria. *Curr. Microbiol.* 44, 291–296. doi: 10.1007/s00284-001-0005-9

Hossain, M. M., and Nakamoto, H. (2003). Role for the cyanobacterial htpG in protection from oxidative stress. *Curr. Microbiol.* 46, 70–76. doi: 10.1007/s00284-002-3831-5

Iteman, I., Rippka, R., Tandeau, D. M. N., and Herdman, M. (2000). Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. *Microbiology* 146, 1275–1286. doi: 10.1099/00221287-146-6-1275

Jain, C., and Rodriguez, R. L. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Johansen, J. R., Bohunická, M., Lukešová, A., Hrèková, K., Vaccarino, M. A., Chesarino, N. M., et al. (2014). Morphological and molecular characterization within 26 strains of the genus cylindrospermum (Nostocaceae, Cyanobacteria), with descriptions of three new species. *J. Phycol.* 50, 187–202. doi: 10.1111/jpy.12150

Johansen, J. R., Kovacik, L., Casamatta, D. A., Iková, K. F., and Kaštovskı, J. (2011). Utility of 16S-23S ITS sequence and secondary structure for recognition of intrageneric and intergeneric limits within cyanobacterial taxa: leptolyngbya corticola sp. nov. (Pseudanabaenaceae, Cyanobacteria). *Nova Hedwigia* 92, 283–302. doi: 10.1127/0029-5035/2011/0092-0283

Kamata, T., Hiramoto, H., Morita, N., Shen, J.-R., Mann, N. H., and Yamamoto, Y. (2005). Quality control of photosystem II: an FtsH protease plays an essential role in the turnover of the reaction center D1 protein in synechocystis PCC 6803 under heat stress as well as light stress conditions. *Photochem. Photobiol. Sci.* 4, 983–990. doi: 10.1039/b506068k

Komárek, J. (2016). A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *Eur. J. Phycol.* 51, 346–353. doi: 10.1080/09670262.2016.1163738

Komárek, J. (2018). Several problems of the polyphasic approach in the modern cyanobacterial system. *Hydrobiologia* 811, 7–17. doi: 10.1007/s10750-017-3379-9

Komarek, J., Kaštovskı, J., Mares, J., and Johansen, J. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera), using a polyphasic approach. *Preslia* 86, 295–335.

Komárek, J., Kaštovskı, J., Ventura, S., Turicchia, S., and Šmarda, J. (2009). The cyanobacterial genus phormidesmis. *Algol. Stud.* 129, 41–59. doi: 10.1127/1864-1318/2009/0129-0041

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

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

Labreck, C. J., Shannon, M., Viola, M. G., Joseph, C., and Camberg, J. L. (2017). The protein chaperone ClpX targets native and non-native aggregated substrates for remodeling, disassembly, and degradation with ClpP. *Front. Mol. Biosci.* 4:26. doi: 10.3389/fmolb.2017.00026

Lacap, D. C., Barraquio, W., and Pointing, S. B. (2007). Thermophilic microbial mats in a tropical geothermal location display pronounced seasonal changes but appear resilient to stochastic disturbance. *Environ. Microbiol.* 9, 3065–3076. doi: 10.1111/j.1462-2920.2007.01417.x

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, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liang, Y., Tang, J., Luo, Y., Kaczmarek, M. B., Li, X., and Daroch, M. (2019). Thermosynechococcus as a thermophilic photosynthetic microbial cell factory for CO2 utilisation. *Bioresour. Technol.* 278, 255–265. doi: 10.1016/j.biortech.2019.01.089

Los, D. A., Zorina, A., Sinetova, M., Kryazhov, S., Mironov, K., and Zinchenko, V. V. (2010). Stress sensors and signal transducers in cyanobacteria. *Sensors* 10, 2386–2415. doi: 10.3390/s100302386

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Mackenzie, R., Pedrós-Alió, C., and Díez, B. (2012). Bacterial composition of microbial mats in hot springs in Northern Patagonia: variations with seasons and temperature. *Extremophiles* 17, 123–136. doi: 10.1007/s00792-012-0499-z

Maeda, S.-I., Murakami, A., Ito, H., Tanaka, A., and Omata, T. (2015). Functional characterization of the FNT family nitrite transporter of marine picocyanobacteria. *Life* 5, 432–446. doi: 10.3390/life5010432

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* 9, 467–477.

Mareš, J., Strunecký, O., Bučinská, L., and Wiedermannová, J. (2019). Evolutionary patterns of thylakoid architecture in cyanobacteria. *Front. Microbiol.* 10:277. doi: 10.3389/fmicb.2019.00277

Mikheeva, L. E., Karbysheva, E. A., and Shestakov, S. V. (2013). The role of mobile genetic elements in the evolution of cyanobacteria. *Russ. J. Genet.* 3, 91–101.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Miscoe, L. H., Johansen, J. R., Kociolek, J. P., Lowe, R. L., Vaccarino, M. A., Pietrasiak, N., et al. (2016). "The diatom flora and cyanobacteria from caves on Kauai, Hawaii. II," in *Novel Cyanobacteria from Caves on Kauai, Hawaii*, eds L. H. Miscoe, J. R. Johansen, J. P. Kociolek, R. L. Lowe, M. A. Vaccarino, N. Pietrasiak, et al. (Stuttgart: Bibliotheca Phycologica), 75–152.

Nakamori, H., Yatabe, T., Yoon, K. S., and Ogo, S. (2014). Purification and characterization of an oxygen-evolving photosystem II from *Leptolyngbya* sp. strain O-77(enzymology, protein engineering, and enzyme technology). *J. Biosci. Bioeng.* 118, 119–124. doi: 10.1016/j.jbiosc.2014.01.009

Nakamoto, H., Suzuki, M., and Kojima, K. (2003). Targeted inactivation of the hrcA repressor gene in cyanobacteria. *FEBS Lett.* 549, 57–62. doi: 10.1016/s0014-5793(03)00768-3

Niclas, E., Cameron, C. R., and Williamh, G. (2010). 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus lyngbya. *J. Phycol.* 46, 591–601. doi: 10.1111/j.1529-8817.2010.00840.x

Oren, A., Ionescu, D., Hindiyeh, M., and Malkawi, H. (2008). Morphological, phylogenetic and physiological diversity of cyanobacteria in the hot springs of zerka ma. *BioRisk* 3, 20–23.

Peng, X., Xu, H., Jones, B., Chen, S., and Zhou, H. (2013). Silicified virus-like nanoparticles in an extreme thermal environment: implications for the preservation of viruses in the geological record. *Geobiology* 11, 511–526.

Perkerson, R. B., Johansen, J. R., Kovácik, L., Brand, J., Kaštovský, J., and Casamatta, D. A. (2011). A unique pseudanabaenalean (cyanobacteria) genus *Nodosilinea* gen. nov. based on morphological and molecular data. *J. Phycol.* 47, 1397–1412. doi: 10.1111/j.1529-8817.2011.01077.x

Price, G. D., Badger, M. R., Woodger, F. J., and Long, B. M. (2008). Advances in understanding the cyanobacterial CO2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* 59, 1441–1461. doi: 10.1093/jxb/erm112

Pruitt, K. D., Tatusova, T., Klimke, W., and Maglott, D. R. (2009). NCBI reference sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 37, D32–D36.

Rajaram, H., Chaurasia, A. K., and Apte, S. K. (2014). Cyanobacterial heat-shock response: role and regulation of molecular chaperones. *Microbiology* 160, 647–658. doi: 10.1099/mic.0.073478-0

Rodriguez-R, L. M., Santosh, G., Harvey, W. T., Ramon, R.-M., Tiedje, J. M., Cole, J. R., et al. (2018). The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of archaea and bacteria at the whole genome level. *Nucleic Acids Res.* 46, W282–W288.

Sato, T., Minagawa, S., Kojima, E., Okamoto, N., and Nakamoto, H. (2010). HtpG, the prokaryotic homologue of Hsp90, stabilizes a phycobilisome protein in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Mol. Microbiol.* 76, 576–589. doi: 10.1111/j.1365-2958.2010.07139.x

Schneider, D. (2010). Similarities and singularities of three DnaK proteins from the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* 51, 1210–1218. doi: 10.1093/pcp/pcq074

Schneider, D. (2011). Thermostability of two cyanobacterial GrpE thermosensors. *Plant Cell Physiol.* 52, 1776–1785. doi: 10.1093/pcp/pcr116

Sciuto, K., and Moro, I. (2016). Detection of the new cosmopolitan genus *Thermoleptolyngbya* (Cyanobacteria, Leptolyngbyaceae) using the 16S rRNA gene and 16S–23S ITS region. *Mol. Phylogenet. Evol.* 105, 15–35. doi: 10.1016/j.ympev.2016.08.010

Shalygin, S., Shalygina, R., Redkina, V., Gargas, C., and Johansen, J. (2020). Description of *Stenomitos kolaenensis* and S. hiloensis sp. nov. (Leptolyngbyaceae, Cyanobacteria) with an emendation of the genus. *Phytotaxa* 440, 108–128. doi: 10.11646/phytotaxa.440.2.3

Shibata, M., Katoh, H., Sonoda, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., et al. (2002). Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: function and phylogenetic analysis. *J. Biol. Chem.* 277, 18658–18664. doi: 10.1074/jbc.m112468200

Shibata, M., Ohkawa, H., Kaneko, T., Fukuzawa, H., Tabata, S., Kaplan, A., et al. (2001). Distinct constitutive and low-CO2 induced CO2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11789–11794. doi: 10.1073/pnas.191258298

Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772. doi: 10.1093/molbev/mst010

Steunou, A.-S., Jensen, S. I., Brecht, E., Becraft, E. D., Bateson, M. M., Kilian, O., et al. (2008). Regulation of nif gene expression and the energetics of N2 fixation over the diel cycle in a hot spring microbial mat. *ISME J.* 2, 364–378. doi: 10.1038/ismej.2007.117

Strunecký, O., Kopejtka, K., Goecke, F., Tomasch, J., Lukavský, J., Neori, A., et al. (2019). High diversity of thermophilic cyanobacteria in Rupite hot spring identified by microscopy, cultivation, single-cell PCR and amplicon sequencing. *Extremophiles* 23, 35–48. doi: 10.1007/s00792-018-1058-z

Tang, J., Du, L.-M., Liang, Y.-M., and Daroch, M. (2019). Complete genome sequence and comparative analysis of Synechococcus sp. CS-601 (SynAce01), a cold-adapted cyanobacterium from an oligotrophic Antarctic habitat. *Int. J. Mol. Sci.* 20:152. doi: 10.3390/ijms20010152

Tang, J., Jiang, D., Luo, Y., Liang, Y., Li, L., Shah, M. M. R., et al. (2018a). Potential new genera of cyanobacterial strains isolated from thermal springs of western Sichuan, China. *Algal Res.* 31, 14–20. doi: 10.1016/j.algal.2018.01.008

Tang, J., Liang, Y., Jiang, D., Li, L., Luo, Y., Shah, M. M. R., et al. (2018b). Temperature-controlled thermophilic bacterial communities in hot springs of western Sichuan, China. *BMC Microbiol.* 18:134. doi: 10.1186/s12866-018-1271-z

Taton, A., Grubisic, S., Ertz, D., Hodgson, D. A., Piccardi, R., Biondi, N., et al. (2010). Polyphasic study of Antarctic cyanobacterial strains. *J. Phycol.* 42, 1257–1270. doi: 10.1111/j.1529-8817.2006.00278.x

Taton, A., Wilmotte, A., Smarda, J., Elster, J., and Komarek, J. (2011). Plectolyngbya hodgsonii: a novel filamentous cyanobacterium from Antarctic lakes. *Polar Biol.* 34, 181–191. doi: 10.1007/s00300-010-0868-y

Tuominen, I., Tyystjärvi, E., and Tyystjärvi, T. (2003). Expression of primary sigma factor (PSF) and PSF-Like sigma factors in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 185, 1116–1119. doi: 10.1128/jb.185.3.1116-1119.2003

Turland, N., Wiersema, J., Barrie, F. R., Greuter, W., and Smith, G. F. (2018). *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the 19th International Botanical Congress Shenzhen, China, July 2017*. China: Koeltz Botanical Books.

Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V., and Chandler, M. (2011). ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* 12:R30.

Vaz, M. G. M. V., Genuário, D. B., Andreote, A. P. D., Malone, C. F. S., Sant'Anna, C. L., Barbiero, L., et al. (2015). *Pantanalinema* gen. nov. and *Alkalinema* gen. nov.: novel pseudanabaenacean genera (Cyanobacteria) isolated from saline–alkaline lakes. *Int. J. Syst. Evol. Microbiol.* 65, 298–308. doi: 10.1099/ijs.0.070110-0

Vieira Vaz, M. G., Genuário, D. B., Andreote, A. P., Malone, C. F., Sant'Anna, C. L., Barbiero, L., et al. (2015). Pantanalinema gen. nov. and *Alkalinema* gen. nov.: novel pseudanabaenacean genera (Cyanobacteria) isolated from saline-alkaline lakes. *Int. J. Syst. Evol. Microbiol.* 65(Pt 1):298.

Vincent, L., Jean-Emmanuel, L., and Olivier, G. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

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

Walter, J. M., Coutinho, F. H., Dutilh, B. E., Swings, J., Thompson, F. L., and Thompson, C. C. (2017). Ecogenomics and Taxonomy of cyanobacteria phylum. *Front. Microbiol.* 8:2132. doi: 10.3389/fmicb.2017.02132

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

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, 293–297.

Yoon, K. S., Nguyen, N. T., Tran, K. T., Tsuji, K., and Ogo, S. (2017). Nitrogen fxation genes and nitrogenase activity of the non-heterocystous cyanobacterium *Thermoleptolyngbya* sp. O-77. *Microbes Environ.* 32, 324–329. doi: 10.1264/jsme2.me17015

Zammit, G., Billi, D., and Albertano, P. (2012). The subaerophytic cyanobacterium *Oculatella subterranea* (Oscillatoriales, Cyanophyceae) gen. et sp. nov.: a cytomorphological and molecular description. *Eur. J. Phycol.* 47, 341–354. doi: 10.1080/09670262.2012.717106

Zhang, C. C., Jang, J., Sakr, S., and Wang, L. (2006). Protein phosphorylation on Ser, Thr and Tyr residues in cyanobacteria. *J. Mol. Microbiol. Biotechnol.* 9, 154–166. doi: 10.1159/000089644

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677. doi: 10.1093/bioinformatics/btt476

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Tang, Li, Li, Du, Shah, Waleron, Waleron, Waleron and Daroch. 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: 28 September 2021
doi: 10.3389/fmicb.2021.685254

Performance of Multiple Metagenomics Pipelines in Understanding Microbial Diversity of a Low-Biomass Spacecraft Assembly Facility

*Jason M. Wood, Nitin K. Singh, Lisa Guan, Arman Seuylemezian, James Nick Benardini and Kasthuri Venkateswaran**

Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States

OPEN ACCESS

Edited by:
Rafael R. de la Haba, University of Seville, Spain

Reviewed by:
Marc Warwick Van Goethem, Lawrence Berkeley National Laboratory, United States
Digvijay Verma, Babasaheb Bhimrao Ambedkar University, India

***Correspondence:**
Kasthuri Venkateswaran
kjvenkat@jpl.nasa.gov

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

Received: *24 March 2021*
Accepted: *27 August 2021*
Published: *28 September 2021*

Citation:
Wood JM, Singh NK, Guan L, Seuylemezian A, Benardini JN and Venkateswaran K (2021) Performance of Multiple Metagenomics Pipelines in Understanding Microbial Diversity of a Low-Biomass Spacecraft Assembly Facility. Front. Microbiol. 12:685254. doi: 10.3389/fmicb.2021.685254

NASA planetary protection (PP) requires an assessment of the biological contamination of the potential microbial burden on spacecraft destined to explore planetary bodies that may harbor signs of life, like Mars and Europa. To help meet these goals, the performance of multiple metagenomic pipelines were compared and assessed for their ability to detect microbial diversity of a low-biomass clean room environment used to build spacecraft destined to these planetary bodies. Four vendors were chosen to implement their own metagenomic analysis pipeline on the shotgun sequences retrieved from environmental surfaces in the relevant environments at NASA's Jet Propulsion Laboratory. None of the vendors showed the same microbial profile patterns when analyzing same raw dataset since each vendor used different pipelines, which begs the question of the validity of a single pipeline to be recommended for future NASA missions. All four vendors detected species of interest, including spore-forming and extremotolerant bacteria, that have the potential to hitch-hike on spacecraft and contaminate the planetary bodies explored. Some vendors demonstrated through functional analysis of the metagenomes that the molecular mechanisms for spore-formation and extremotolerance were represented in the data. However, relative abundances of these microorganisms varied drastically between vendor analyses, questioning the ability of these pipelines to quantify the number of PP-relevant microorganisms on a spacecraft surface. Metagenomics offers tantalizing access to the genetic and functional potential of a microbial community that may offer NASA a viable method for microbial burden assays for planetary protection purposes. However, future development of technologies such as streamlining the processing of shotgun metagenome sequence data, long read sequencing, and all-inclusive larger curated and annotated microbial genome databases will be required to validate and translate relative abundances into an actionable assessment of PP-related microbes of interest. Additionally, the future development of machine learning and artificial

intelligence techniques could help enhance the quality of these metagenomic analyses by providing more accurate identification of the genetic and functional potential of a microbial community.

Keywords: planetary protection, metagenomics, Spacecraft Assembly Facility, microbial diversity, low-biomass

INTRODUCTION

Planetary protection (PP) requires a periodic assessment of the potential biological contamination for microbial burden of space flight hardware destined to or nearby planetary bodies that may or may not have harbored signs of life such as Mars or the Icy Worlds of the outer solar system (COSPAR, 2011). Space hardware exploring the possibility of life on a planetary body of interest, such as recent missions to Mars, are required to be cleaned to have less than an average of 300 spores/m^2 on spacecraft surfaces and less than 5×10^5 spores at launch (Benardini et al., 2014). For future missions, such as Mars Sample Return, in addition to limiting outbound contamination, the science community has deemed it important to collect an Earth-based assessment of the potential biological and organic contamination that the hardware could have experienced as part of the hardware integration and test operations (Beaty et al., 2018). Upcoming missions to the outer planets and the Icy Worlds, such as Europa Clipper, require less than 1×10^{-4} probability of contamination of the subsurface ocean by Earth-based microorganisms (National Research Council, 2000). Currently, the NASA standard assay (NSA) is used to test for spores on spacecraft, but this assay will not detect most microorganisms (NASA., 2010). Heat tolerant microorganisms (80°C; 15 min) that can grow aerobically in a nutrient rich Tryptic Soy Agar (TSA) medium, incubated at 32°C, for 72 h are captured by this NSA measurement, but slow growing microorganisms that (i) prefer cooler or hotter temperatures, (ii) are obligate anaerobes, (iii) or have dietary requirements not met by TSA may not be detected by this NSA method.

Metagenomics has been identified as an alternative biological verification technique to replace or supplement a culture-based assay that NASA could potentially use to detect the widest possible spectrum of microorganisms, including those of PP interest. NASA is seeking to develop this technology to demonstrate bioburden control assessments for enumeration of relevant viable organisms or contamination risk assessment. In doing so, this approach would likely include an enumeration, phylogenetic identification, and high-resolution characterization of microbial traits and biochemical capabilities from spacecraft surfaces. Microorganisms such as those which are radiation resistant, psychrophilic or anaerobic, etc., are likely to withstand space flight conditions (launch to landing) and may be able to survive on an extraterrestrial planet (forward contamination) (Thompson et al., 2017). High throughput metagenomics sequencing and bioinformatics pipelines have the potential to rapidly identify most of the microbial diversity (Horneck et al., 2012; Tamames and Puente-Sánchez, 2019) and to classify those microorganisms into PP relevant categories, but are not yet available.

The goal of this work is to test the ability of modern metagenomics pipelines to enable NASA in defining new PP requirements for detecting microorganisms on space hardware for future life detection missions. Proven modern molecular technologies are essential for NASA's mission because none of the techniques used to clean and sterilize spacecraft components and subsystems (e.g., dry-heat sterilization, vapor hydrogen peroxide, oxygen plasma, etc.) are compatible with fully assembled, modern day spacecraft (Chung et al., 2008; Shirey et al., 2017). Hence, hitch-hiking microorganisms associated with spacecraft must be measured and characterized using modern molecular techniques throughout assembly of the spacecraft. Techniques like 16S rRNA gene cloning (1980s) (Pace, 1997), microarray (early 2000s) (Vaishampayan et al., 2010), and targeted amplicon sequencing (2010s) (Minich et al., 2018) were tried and utilized to measure the widest possible spectrum of microorganisms associated with spacecraft components, however, these molecular methods are unable to characterize functional pathways which are essential for identifying PP related microorganisms. Shotgun metagenomics sequencing and analyses pipelines can detect, quantify, and assess potential metabolisms of both cultivable and uncultivable microorganisms (Singh et al., 2018).

To test the ability of various metagenomics pipelines to detect the widest possible spectrum of microorganisms and of relevance to the PP mission, four vendors (labeled C, J, L, and N) were chosen to perform analyses of metagenomes generated from surface samples collected at the Jet Propulsion Laboratory (JPL) Spacecraft Assembly Facility (SAF) between March and August, 2016, just prior to the beginning of assembly of the Mars 2020 rover, a bioburden controlled and sample return biological contamination knowledge capture mission. These vendors are affiliated with a private industry (Vendor L), national laboratory (Vendor J), and academia (Vendor C, N), in the United States. The SAF is a good test environment for this project because (i) it is where Mars bound spacecraft have been assembled, (ii) it is a critical cleanroom for projects where component to assembly level integration and testing occurs, (iii) a rigorous cleaning regime and bio-control ensures that surface samples are low in biomass like samples from spacecraft surfaces, and (iv) it is a good testbed where microorganisms of PP concern could be detected and have been previously documented (Probst et al., 2010; Vaishampayan et al., 2012). Each vendor implemented their own metagenomics analysis pipeline and choose their own sequence databases for comparison without our interference, but all vendors were provided the same metagenomics sequence and associated metadata.

To measure the effectiveness of each pipeline, various requirements were considered. Quality control measures such as adapter removal, filtering of human reads, and sequence length and quality cutoffs were expected to be performed by each

vendor. Control samples that displayed high or low similarity with other samples were included in the dataset sent to each vendor to test how controls were processed. Databases used by each vendor were considered for their ability to detect microorganisms of PP concern. The proportion of high-quality sequences assigned an annotation is also important, as a more comprehensive description of the sampled diversity clarifies the PP risk presented by detected microorganisms.

MATERIALS AND METHODS

Description of Samples and Sample Selection Criteria

Among more than 200 samples collected from the JPL SAF clean room floors and control samples that were shotgun metagenome sequenced (Hendrickson et al., 2017), 20 were selected for this exercise (**Table 1**). These consist of handling and instrument controls, which include two dirty samples (FC2; 872,202 reads and MC2; 573,199 reads), and a clean sample (FC9; 5,471 reads). Additionally, SAF floor samples containing $<1 \times 10^6$ reads (n = 2; Group L) and $>1 \times 10^6$ reads (n = 15; Group H) were included. Among samples containing $>1 \times 10^6$ reads, three consist of microorganisms relevant for PP (Actinobacteria, Firmicutes, etc., Group H-F).

Sample collection and processing methods were previously published (Hendrickson et al., 2017). Briefly, 1 m^2 of floor was wiped with sterile, pre-moistened 23 × 23 cm polyester wipes (Texwipe; TX1009, Kernersville, NC, United States), and placed in a sterile bottle containing 200 mL phosphate buffer saline (PBS). The sample was mixed thoroughly, and the resulting particulates and microorganisms suspended in the PBS were concentrated with a CP-150 concentrating pipette (InnovaPrep, Drexel, MO, United States). DNA was extracted from concentrated samples using a Maxwell 16 (Promega, Madison, WI, United States) and sequenced using Illumina MiSeq (San Diego, CA, United States). Handling control samples (pre-moistened polyester wipes waved in the air of SAF) were also examined exactly like the SAF environmental samples. Reagent controls were also included during DNA extraction with the Maxwell 16 (well filled with PBS rather than concentrated sample).

Metagenome Sequencing

Shotgun metagenome sequencing was carried out as described previously (Singh et al., 2018; Avila-Herrera et al., 2020). Briefly, DNA libraries were prepared for sequencing using the NextEra DNA Library Preparation Kit (Illumina, Inc., San Diego, CA, United States). Quality and fragment size were assessed on the Agilent Tapestation 4200 (Agilent Technologies, Santa Clara, CA, United States). Libraries were quantitated using the Qubit fluorimeter (Thermo Fisher Scientific, Waltham, MA, United States) and normalized to equivalent DNA quantities, pooled, and diluted according to the manufacturer's standard recommendations. Shotgun metagenomic sequencing was performed using an Illumina NextSeq 500 with the NextSeq Series High Output Kit v2 (Illumina Inc., San Diego, CA, United States), using 150 base pair, paired end reads. Fastq files are generated from the sequencing results using the bcl2fastq software (Illumina) and given to vendors.

Taxonomic and Functional Assignment

Each vendor was asked to perform their own taxonomic and functional assignment for all metagenomic reads provided to them using their preferred pipeline. Each vendor used different methods for quality control (QC) and classification of sequence data, used different databases for taxonomic and functional annotation. Detailed descriptions of the methodologies of each vendor are provided below.

Vendor L Methods

Quality Control

To support multiple platforms, metadata included in the read files are analyzed to determine the appropriate QC pipeline and threshold settings. Settings are persisted to the database (for default values) and GUI support is implemented to allow users to enter custom threshold values for subsequent steps. An executable binary file is constructed and launched as either a parallel worker thread or parallel executor (Spark support) to construct an interactive HTML5 compliant page. This page is served to the user through workflow support in the GUI. Low quality reads are removed and cataloged (metadata to RDS) according to user-selected values.

Read Pairing

For sequencing technologies and protocols that support paired reads, similar sample paired read files are merged leveraging proprietary merge packages implemented in Java. Output is directed to a file system and file paths are persisted in the database allowing for simple GUI navigation.

Adapter Prediction

For paired reads that do not include known inserts and adapter sequences, the paired reads are run through a predictive algorithm. Metadata is pulled from the proprietary algorithm and persisted in the database.

Trimming

Merged paired reads are trimmed by either the known sequence input, or the predicted adapters that can be optionally generated. Trimmed, quality control screened files are then used in subsequent analysis using TrimGalore Version 0.6.5[1].

Phylogenetic Analysis

High-throughput BLAST servers were set up on GPU-based AWS Elastic Cloud Compute (EC2) instances to support a high-throughput parallel read blast. Each individual read was mapped against libraries of known microbial genome assemblies. Reads that meet either default or user set thresholds can be aligned with all known microbial genomes and depending on similarity score can be assigned a taxonomic identification.

[1]https://github.com/FelixKrueger/TrimGalore

TABLE 1 | Characteristics of samples analyzed.

						Vendor L			Vendor J			Vendor C			Vendor N	
Sample	Group	Location	Sample Type	Date of Sampling	Total number of raw reads	QC Reads[1]	Taxonomically Classified Reads[2]	>1%[3]	QC Reads[1]	Taxonomically Classified Reads[4]	>1%[3]	QC Reads[1]	Taxonomically Classified Reads[5]	>1%[3]	QC Reads[1]	Taxonomically Classified Reads[6]
FC2	Control	Field Control	Air of SAF	15-March-2016	872,202	74,023	338	332	320,075	9,263	9,163	41,880	9,911	9,705	15,726	N/A
FC9	Control	Field Control	Air of SAF	12-July-2016	5,471	232	–	–	704	44	9	401	44	44	83	N/A
MC2	Control	Maxwell Control	Extraction Control	15-March-2016	573,199	129,426	923	893	290,218	81,808	30,287	118,891	58,717	57,179	100,730	N/A
S14	L	L6	Floor of SAF	15-March-2016	529,438	108,702	225	222	259,750	31,984	2,669	101,610	4,406	3,813	77,363	N/A
S99	L	L11	Floor of SAF	26-July-2016	427,481	2,671	12,223 7	10,932	2,653	1,894	432	2,564	194	194	2,544	N/A
S12	H	L2	Floor of SAF	15-March-2016	1,097,438	787,877	1,108	1,061	869,207	267,667	48,750	564,322	59,647	45,850	676,490	N/A
S16	H	L5	Floor of SAF	15-March-2016	1,205,398	862,588	4,030	3,896	954,606	604,825	100,850	823,471	270,698	260,492	754,851	N/A
S17	H	L9	Floor of SAF	15-Mar-2016	10,647,036	9,726,824	26,421	25,298	9,782,262	1,176,911	108,555	9,733,552	502,672	458,315	8,680,384	N/A
S18	H	L13	Floor of SAF	15-March-2016	1,254,780	858,915	2,028	1,938	974,280	389,330	85,782	829,518	126,622	112,610	743,226	N/A
S20	H	L10	Floor of SAF	15-Mar-2016	2,001,909	1,584,542	4,593	4,371	1,644,129	1,001,993	240,780	1,535,396	283,615	205,900	1,305,244	N/A
S41	H	L1	Floor of SAF	17-May-2016	3,329,331	2,857,079	4,886	4,641	2,811,738	519,088	63,155	1,346,708	93,215	67,717	2,639,304	N/A
S42	H	L10	Floor of SAF	17-May-2016	2,726,783	2,212,223	3,793	3,497	2,208,614	803,127	68,856	1,261,498	204,248	153,587	2,056,000	N/A
S43	H	L8	Floor of SAF	17-May-2016	17,258,784	15,574,766	27,828	25,427	15,455,888	4,682,775	464,348	14,069,392	1,328,792	1,111,901	14,659,904	N/A
S44	H	L12	Floor of SAF	17-May-2016	1,335,957	1,104,044	1,310	1,291	1,099,278	312,955	46,684	729,636	56,630	43,348	1,025,038	N/A
S45	H	L10	Floor of SAF	17-May-2016	2,369,108	1,948,423	2,274	2,174	1,940,897	549,588	59,777	1,549,155	146,026	103,581	1,809,578	N/A
S57	H	L5	Floor of SAF	1-June-2016	2,661,331	2,319,103	8,192	7,743	2,317,306	1,205,865	238,697	2,166,598	218,332	147,781	2,189,283	N/A
S108	H	L9	Floor of SAF	16-August-2016	2,150,195	2,090,570	5,759	5,632	2,068,840	1,316,931	151,302	2,063,778	463,731	354,047	2,005,116	N/A
S29	H-F	L10	Floor of SAF	30-March-2016	4,117,535	3,477,442	21,199	19,808	3,568,720	2,324,729	316,881	3,229,576	1,996,816	1,926,851	3,050,205	N/A
S30	H-F	L1	Floor of SAF	30-March-2016	3,592,661	3,220,908	11,982	11,388	3,243,095	2,487,003	405,394	3,209,450	2,106,955	2,028,611	2,846,917	N/A
S76	H-F	L5	Floor of SAF	28-June-2016	14,027,039	12,697,322	5,797	5,418	12,636,003	1,332,390	136,865	12,709,423	251,337	211,846	12,079,811	N/A

Groups includes all control samples (n = 3), group [L]ow includes samples <1 × 10^6 reads (n = 2), and group [H]igh includes samples >1 × 10^6 reads (n = 15). A subset of group H, group H-F, includes samples known to be abundant in Actinobacteria and Firmicutes. [1]Quality Control pipelines are detailed in section "Materials and Methods". [2]Proprietary pipeline of Vendor L. [3]Low abundance taxa (<1% of a library) were removed from each vendor provided table to simplify comparison. [4]Diamond BlastX/MEGAN6 analysis by Vendor J. [5]KrakenUniq analysis by Vendor C. [6]MetaPhlAn2 analysis by Vendor N provides relative abundance in terms of %, not absolute read counts. [7]Anomaly.

Vendor J Methods

Metagenome Sequence Data Processing

Paired-end 150 bp reads were processed with Trimmomatic (Bolger et al., 2014) to trim adapter sequences and low-quality ends, with a minimum Phred score of 20 across the entire length of the read used as a quality cutoff. Reads shorter than 80 bp after trimming were discarded. All reads were normalized across samples as previously recommended (Nayfach and Pollard, 2016). High-quality filtered reads were clustered to respective taxonomic levels (domains through species) using the lowest common ancestor (LCA) algorithm provided by MEGAN6 (Huson et al., 2007).

Taxonomic and Functional Assignment

For lower downstream processing and visualization, the MEGAN6 metagenomics toolkit was used (Huson et al., 2017). The NCBI taxonomy database (Sayers et al., 2008), containing over 6.6×10^5 reference sequences, and NCBI-NR protein sequence database, consisting of entries from GenPept, Swiss-Prot, PIR, PDB, and RefSeq, were used to assign taxonomic features to reads by using DIAMOND (Buchfink et al., 2015) and the weighted LCA algorithm of MEGAN6 (Huson et al., 2016). The identification of the reads to a taxon is not based on the genes only, but it is based on the comparison of the reads with the reference sequences deduced from the genomes of the curated NCBI taxonomy database (Sayers et al., 2008). Briefly, taxonomic and functional binning of the metagenomic reads is carried out using MEGAN6 (Huson et al., 2017), with the following settings: minScore = 50, maxExpected = 0.01, topPercent = 10, and minSupportPercent = 0.01. The resulting taxon assignments are presented in this manuscript. Functional analysis was carried out by mapping filtered DNA sequences against a reference database of all proteins within eggNOG (Powell et al., 2011), SEED (Overbeek et al., 2005), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) databases. The search for translated DNA sequences was executed using DIAMOND, and hits that spanned $\geq$20 amino acids with $\geq$90% similarity were retained. In cases where one read matched these criteria against multiple proteins, only the protein or proteins (in the event of a tie) with the maximum bit score were considered. Pathways were analyzed by summing counts of each KEGG orthology in a pathway. Using different databases allowed a detailed view of reads defined by gene function consisting of a collection of biologically defined (i) subsystems, (ii) clusters of orthologous groups, and (iii) collection of metabolic pathways.

Vendor C Methods

Quality Control

Adapters and low-quality bases were removed using AdapterRemoval v2 (Schubert et al., 2016). Bases with a quality of 1 were removed as were ambiguous bases. Reads shorter than 50 bp after trimming were discarded. The remaining reads were aligned against the human genome with alternate contigs using Bowtie2 (Langmead and Salzberg, 2012) and "–sensitive" settings. Read pairs where only one read aligned were discarded.

Jellyfish (Marçais and Kingsford, 2011) was used to count k-mers in the processed reads. All k-mers, including singletons were counted. Various statistics were calculated on k-mers using a python script[2]. The two statistics presented here are (1) fraction of k-mers which are singletons, the number of k-mers which only occurred once vs. the total number of unique k-mers, and (2) k-mer entropy, Shannon entropy calculated over the probability of drawing each k-mer at random, as shown in previous work (Cleary et al., 2015; Costea et al., 2017).

Identification of likely negative controls from QC data was performed by manual inspection. Decision factors included the k-mer complexity (high singleton fraction, low entropy), the number of reads filtered, and taxonomic similarity to other samples (Levy and Borenstein, 2012; Thompson et al., 2017).

Taxonomic Profiling

Initial taxonomic profiling was performed by mapping clean-reads to all of RefSeq Microbial (1,977,559 unique taxa) using KrakenUniq (Breitwieser et al., 2018). KrakenUniq can produce false positive species calls, so we aggressively filtered results (McIntyre et al., 2017). We removed all species which were identified using fewer than 1,024 unique marker k-mers across the entire dataset. Additionally, species were removed if they were identified with fewer than 10,000 unique marker k-mers unless at least 10% of all known marker k-mers for that species were found. To provide an additional comparison, taxonomic profiles generated by MetaPhlAn2 (Segata et al., 2012) were also analyzed.

De novo Genome Assembly

Bacterial genomic sequences were assembled using metaSPAdes (Nurk et al., 2017), the best in-class metagenomic assembler. Resulting sequences were filtered for quality and duplicates, and annotated by aligning them to known sequences in the NCBI NT database. Annotations were processed using a series of GitHub scripts[3] to identify a single likely annotation per outcome. Possible genes were annotated on the assembled *Acinetobacter* and *Bacillus* genomes (planetary protection relevant microbes), using PROKKA[4]. Many contigs, particularly larger contigs, did not precisely match any known taxa and may be from novel microbial strains.

Estimating Growth Rate

The estimated rate of growth for the two major taxa identified was evaluated using Growth Rate Index (GRiD) (Emiola and Oh, 2018). GRiD uses the peak to trough ratio of coverage on a microbial genome and a sophisticated series of filters to estimate that genomes rate of replication. GRiD is designed to work well even with low coverage samples and low-quality genome assemblies.

[2]https://github.com/dcdanko/gimmebio/blob/master/gimmebio/kmers/gimmebio/kmers/cli.py

[3]https://github.com/dcdanko/gimmebio/blob/develop/gimmebio/assembly/gimmebio/assembly/cli.py

[4]https://github.com/tseemann/prokka

Functional Analysis

Microbial function was evaluated using the Human Microbiome Project (HMP) Unified Metabolic Analysis Network (HUMAnN2) (Franzosa et al., 2018). HUMAnN2 maps reads to UniRef90, a database of functional genes and combines genes into known metabolic pathways. Pathways are summarized by the total abundance of genes in the pathway and by the fraction of genes in the pathway which are identified. Pathways with less than 50% coverage were filtered from further analysis.

Comparison to MetaSUB Data

Vendor C used MetaSUB[5], a large database of metagenomic samples from urban environments, to contextualize PP samples. This includes 2,126 MetaSUB samples with surface annotations and 371 MetaSUB samples collected from the air of six cities. These MetaSUB samples were processed for taxonomy and metabolic function analogously to how Vendor C processed the 20 samples in this study (n = 2,517 total).

After processing all 2,517 samples, taxonomic and functional profiles were reduced to a binary representation, indicating if a given taxa or pathway was detected in a given sample or not. Without performing this reduction of profiles to a binary representation, clean-room samples would cluster separate from MetaSUB samples. UMAP (McInnes et al., 2020) was used to perform dimensionality reductions for both taxonomy and function and observation of the results (Hart et al., 2015).

Vendor N Methods

Sequencing Metrics and Quality Control

The sequencing quality of all 20 datasets was assessed using the publicly available FASTX-toolkit package[6], as well as FastQC package[7]. Prior to analysis with MetaPhlAn2 and Kraken2 (see below), the paired reads were trimmed to removed adapters and low-quality bases using Trimmomatic (Bolger et al., 2014) with the following parameters: ILLUMINACLIP:/path/to/adapter.fasta:2:25:10 SLIDINGWINDOW:5:20 MINLEN:60.

Metagenomic Analysis of Sample Composition

All sequence reads were segmented into non-overlapping 50-mer fragments, which were then taxonomically classified down to the species level using the MTSv pipeline. The MTSv pipeline[8] uses a local copy of NCBI's GenBank database (accessed as of 06/15/2018) as its reference database. The pipeline performed a true alignment with up to three mismatches tolerated against all taxonomically classified sequences. Only the sequences that align to one-and-only-one taxonomic unit within the entirety of NCBI's GenBank are used for taxonomic classification. In this fashion, only the sequence fragments that unambiguously support the presence of an organism are used, significantly reducing the false positive rate associated with metagenomic characterization of complex samples. Only organisms achieving at least 300 unique signature hits are reported.

The 20 metagenome datasets were also analyzed with MetaPhlAn2 (Segata et al., 2012), Kraken2 (Wood et al., 2019) and the downstream statistical module, Bracken (Lu et al., 2017). Bracken takes Kraken2 output and computes species abundance using Bayesian reestimation. In contrast to MetaPhlAn2, which relies upon a subset of markers for a given genome for classification, thereby limiting its ability to resolve low abundance organisms, Kraken2 is a k-mer based alignment against a reduced, but still large, version of RefSeq. Thus, Kraken2/Bracken is generally able to identify organisms in low abundance, but conversely is susceptible to false positives.

Functional Analysis of Samples

HUMAnN2 (Franzosa et al., 2018) is a tool for gene annotation metabolic pathway discovery from both metagenomic and metatranscriptomic data, and was developed by the same group that published MetaPhlAn2. HUMAnN2 generates three types of output: gene family abundance; pathway abundance; and pathway coverage. Each sample was run through HUMAnN2 where the analysis focused on gene family abundance for each genome that was discovered by MetaPhlAn2.

Comparative Analyses of Vendor Results

All four vendors provided a table containing abundances for microbial species detected in each metagenomic libraries. Since some vendors provided species in addition to strain names while others provided only species names, all strain names were excised and abundances were summed as necessary to better facilitate a comparison between vendors. All read counts in the subsequent species tables were normalized by the number of annotated reads in each metagenomics library if not already performed by the vendor. Output tables from each vendor containing taxon abundance were merged together and utilized for comparative analyses.

Venn diagrams demonstrating overlap among species by the four vendors were generated using InteractiVenn (Heberle et al., 2015) for surface samples and controls separately. Analysis of similarities (ANOSIM) and non-metric multidimensional scaling (NMDS) comparing taxon tables from the four vendors were performed using the vegan R package (Oksanen et al., 2013) and custom R scripts[9], [10]. Network diagrams for surface samples or controls were generated by averaging relative abundance for each taxon for each vendor and converting the resulting species matrix into a network using Cytoscape (v.3.8.0) (Shannon et al., 2003) and a custom Perl script[11].

RESULTS

A key need for PP efforts is a rigorous genetic catalog of the presence of microorganisms on and around spacecraft associated environments, as well as an assessment of their

[5]http://www.metasub.org

[6]http://hannonlab.cshl.edu/fastx_toolkit/

[7]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[8]https://github.com/FofanovLab/MTSv

[9]https://github.com/sandain/R/blob/main/anosim.R

[10]https://github.com/sandain/R/blob/main/mds.R

[11]https://github.com/sandain/pigeon/blob/main/scripts/species-matrix2graphml.pl

likelihood of survival or persistence as well as proliferation in extraterrestrial conditions. To aid in this effort, we analyzed 20 PP-related samples for microbial taxonomy, metabolic function, growth rate, and assembled large genomic contigs. Shotgun metagenomics raw reads (BioProject PRJNA668809) were given to four different vendors to analyze bioinformatically using their favorite publicly available or in-house pipelines to hypothesize on the composition of the microbial community sampled. One of the main features of the analysis pipeline that deviated from other vendors was that Vendor J performed a protein-based taxonomic assignment, whereas other vendors carried out nucleotide-based taxonomic assignment. All vendors removed adapter sequences and human reads, retaining <1% (of control sample FC9) to >90% (of sample S108) of the total raw reads (7.22×10^7; **Table 1**) which allowed for further taxonomic classification. To simplify comparative analyses, low abundance taxa identified by each vendor that comprised <1% relative abundance of a library were excised, reducing the effective number of reads analyzed by 0 to 92% (**Table 1**).

Vendor L Results

After QC steps followed by the Vendor L pipeline to remove adapter sequences and human reads generated 61,637,680 reads from 20 samples tested, and hence this vendor retained 85% of the raw reads for further analyses (**Table 1**, **Figure 1**, and **Supplementary Figure 1A**). The number of QC reads associated with 17 SAF floor samples ranged from 2,671 (S99) to 15,574,766 (S43), whereas the number of reads from field controls ($n = 2$) and Maxwell reagent control ($n = 1$) samples were less than 1,000. Only 0.22% of these QC reads were taxonomically classified by Vendor L. In general, Vendor L detected 82 species with high confidence (>1% relative abundance in at least one library) in the 17 SAF surface samples analyzed (**Figure 2A**), including 4 species of potential PP concern, *Bacillus flexus*, *Brevundimonas diminuta, Acinetobacter lwoffii*, and *A. johnsonii*. Additionally, Vendor L detected multiple species that are normally associated with the human skin and gut microbiome, including *Ruminococcus gnavus*, *Staphylococcus aureus*, *S. epidermidis*, and *Propionibacterium acnes*.

One of the field control samples (FC9) was found to possess only 232 reads after QC, hence no useful chromatogram could be generated to determine the distribution of species. In the other two control samples, Vendor L detected 11 species in Maxwell control (MC2) and 10 species in field control (FC2) samples (**Supplementary Figure 1B**). There were 2 bacterial species of potential PP concern [*Bacillus flexus* (83 reads) and *Anoxybacillus kestanbolensis* (61 reads)] found in Maxwell control (MC2) with very low number of reads. Similarly, Vendor L detected bacterial species associated with the human microbiome in controls samples (30–162 reads in FC2 and 28–489 reads in MC2), including *Staphylococcus aureus*, *S. epidermidis*, and *Propionibacterium acnes*.

During this study, Vendor L communicated that they have developed a proprietary pipeline to assemble genomes for unknown taxonomic units. Vendor L's approach includes a method to determine contiguous regions of unmapped reads

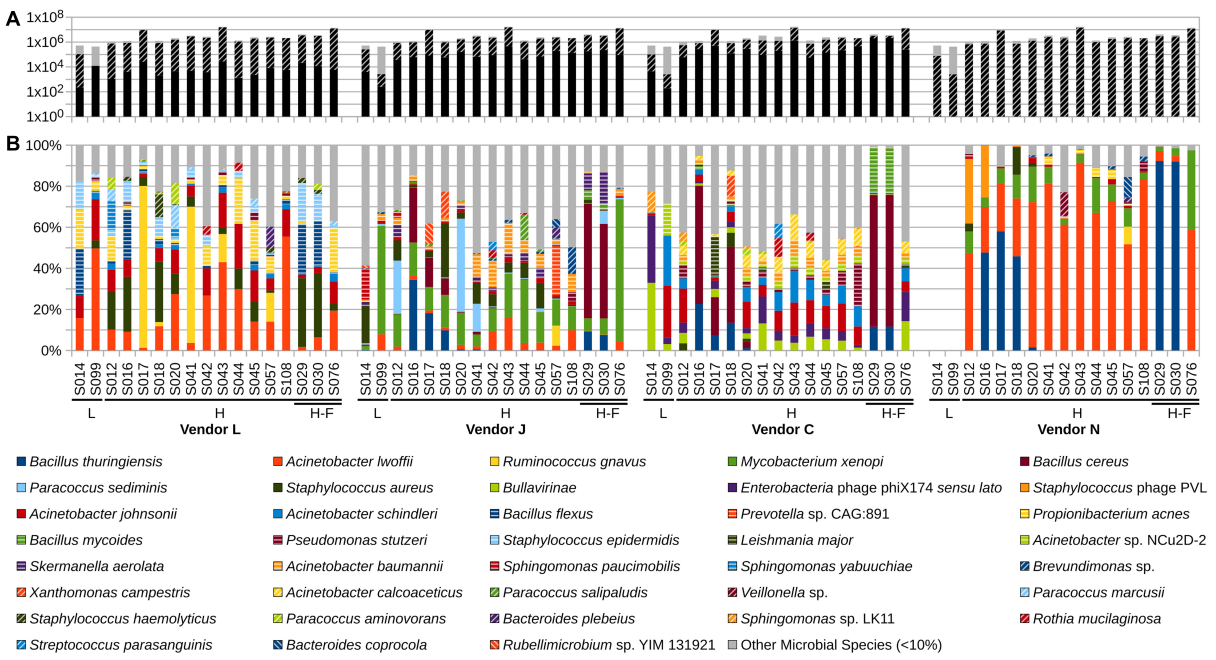


FIGURE 1 | Abundance of reads in SAF floor samples that passed quality control (QC) measures of each vendor (black and black with cross-hatch; gray represents reads discarded by the QC stage; black without cross-hatch represent QC reads taxonomically classified) **(A)**, and relative abundance of species detected by each vendor **(B)**. Samples with a low- (group [L]ow; $<1 \times 10^6$) or high- (group [H]igh; $>1 \times 10^6$) numbers of reads were included. Samples with high numbers of reads associated with Actinobacteria and Firmicutes (group H-F) were also included. Vendor C used KrakenUniq and vendor N used MetaPhlAn2 for classifying the reads and assigning taxa.

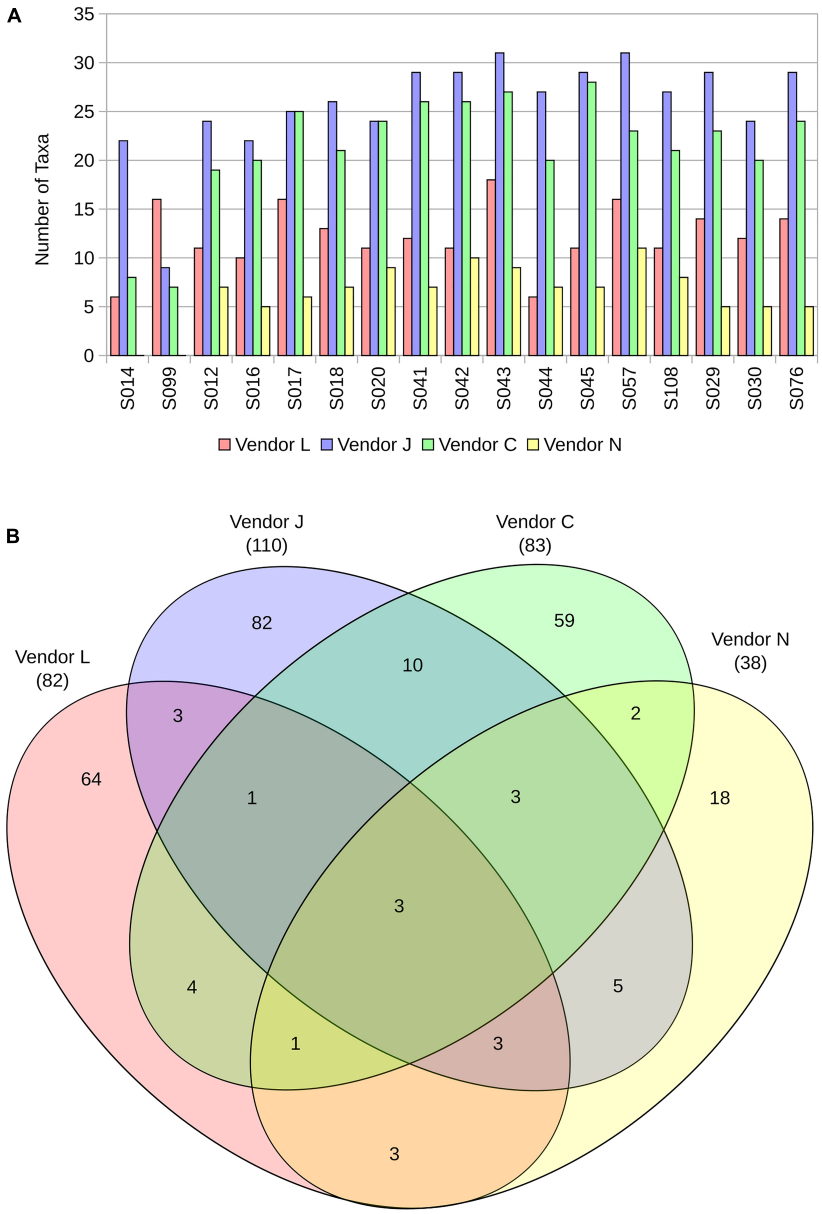


FIGURE 2 | The number of taxa identified in each sample by the four vendors **(A)**, and a Venn diagram representing overlap in taxa detected in SAF floor samples **(B)**.

and to score and assess them for potential assembly into novel genomes of unknown species. However, such metagenome assembled genomes were not described by this vendor for this work. In addition, Vendor L failed to provide the results of their functional characterization.

Vendor J Results

Among the total reads, ∼86% (62,448,263 reads) passed through the QC process (Fastp) of Vendor J (**Table 1**, **Figure 1A**, and **Supplementary Figure 1A**). After QC most of the reads were taxonomically classified as bacteria (85%), followed by eukaryotes (14%), then viruses (<1%), with few archaeal signatures detected (<0.1%). Vendor J detected 110 microbial taxa with high confidence (>1% relative abundance in at least one library) in the shotgun metagenome (**Figure 1B**). Dominating bacterial genera included *Bacillus*, *Acinetobacter*, *Bacteroides*, *Mycobacterium*, *Parabacteroides*, *Sphingomonas*, *Paracoccus*, *Lachnoclostridium*, *Clostridium*, and *Hungatella*. Vendor J paid special attention to detect Firmicutes and Actinobacteria, microorganisms that forms spores or are hardy; also, they are resistant to the extreme conditions (arid, radiation, etc.). *Bacillus*, *Lachnoclostridium*, and *Clostridium* were dominant Firmicutes, while *Mycobacterium*, *Corynebacterium*, and *Microbacterium* dominated *Actinobacteria*. Dominant bacteria detected are part of the human microbiome, which could be the reason for their abundance in the SAF environment.

Vendor J was able to taxonomically resolve to members of eukaryotes with >1% abundance in these samples, even though these cleanrooms were documented to be low in fungal diversity (La Duc et al., 2012). Among the total metagenomic reads of eukaryotes (14%), only half of them were associated with identifiable taxa having >1% relative abundance. Most of these eukaryotic sequences were Fungi, dominated by *Aureobasidium pullulans* and *Coniosporium apollinis*. Both the dominant fungi have been associated with low nutrient biomes (La Duc et al., 2012). Strains of *A. pullulans* have been isolated from SAF (La Duc et al., 2003), and *C. apollinis* is a rock inhibiting lithotroph (Sterflinger et al., 1997) that may be living in the cleanroom floor of the SAF. Similarly, less than 0.02% of archaeal and virus signatures were observed among the predominant metagenomic reads (>1%). Such low abundance of archaea and viruses in cleanrooms are well-known (Moissl et al., 2008).

Vendor J also inspected the genomic and metabolic capability of the microbial community members. To examine the presence of dormancy and sporulation genes, sequence reads from all samples were mapped to individual microbial genes, and then assigned to KEGG, SEED, and eggNOG categories. Genes associated with stationary phase, dormancy, and persistence were detected by Vendor J, including the HipAB system implicated in growth arrest, persistence, and drug tolerance, the Mycobacterial signal transduction system (MprAB) required for persistent infections, and a ribosomal hibernation related cluster. Additionally, multiple genes associated with sporulation were detected, including coat proteins (CotJABC), synthesis of dipicolinate and the exosporium, forespore to mother cell channel proteins, spore germination and germinant receptors, and other sporulation cluster genes like SigEG, SpoIIIAA-SpoIIIAH, SpoVA, and SpoVS (**Supplementary Table 1**). A low-biomass environment tends to have a majority of spore-forming populations (La Duc et al., 2007), and this environment is exhibiting presence of functional genes associated with spore-forming members. A descriptive result of eggNOG categories is presented in **Supplementary Table 1D**, mainly addressing the primary cellular function. KEGG analysis showed a much better input of the functional makeup of the microbial community. The presence of factors for "metabolism of terpenoids and polyketides," "biosynthesis of other secondary metabolites," "xenobiotics biodegradation and metabolism," and "environmental adaptation," represents a highly stressed and competitive environment. A closer look into the environmental adaption reveals the presence of the "bacterial secretion system" which is a strategy utilized by the pathogenic organism to infect the host. Our previous studies have shown that spacecraft associated environment is very similar to the nosocomial setup as both are maintained to near sterile levels using industrial cleaners (data not shown).

Vendor C Results

Before any processing, the 20 samples contained a total of 72,183,076 reads. Of these, just over 10 million reads contained large Illumina adapter sequences. Vendor C removed adapters and low-quality bases using AdapterRemoval, thus reducing the total read count to 61,666,094 reads. Vendor C then removed human reads by aligning to the human reference genome using Bowtie2 (hg38, sensitive), and after removing reads where only one end aligned to the reference, Vendor C was left with 56,086,819 reads, 78% of the total (**Table 1**). Vendor C noted that this is a surprisingly small amount of human DNA for an environmental sample (Mason et al., 2016). Vendor C noted that FC9 and FC2 had noticeably lower complexity than other samples and correctly assumed they were negative controls. A third sample (MC2), was considered ambiguous by Vendor C due to its intermediate complexity. This MC2 sample was a Maxwell reagent control and Vendor C curation deemed correct in deciding which samples were control samples during this blind study.

Vendor C identified 82 relatively abundant taxa (>1%; **Figure 1**) in SAF floor samples, some of which are relevant for PP efforts. By abundance, these species principally came from 7 genera: *Bacillus*, *Acinetobacter*, *Paracoccus*, *Pseudomonas*, *Sphingomonas*, *Methylobacterium*, and *Brevundimonas*. Samples could be broadly clustered into three groups based on their taxonomic profile: the putative negative controls (two samples), *Acinetobacter*-dominated samples (eleven samples), and *Bacillus*-dominated samples (six samples). One sample, S14, was ambiguous. *Acinetobacter*-dominated samples tended to have higher taxonomic diversity than *Bacillus*-dominated samples. Dozens of these organisms are known to persist and survive in deserts, oceans, the arctic, or other harsh terrestrial environments, and are thus likely to be relevant for inter-planetary missions and planning. Vendor C identified a range of predicted phenotypes of PP interest, including: spore-forming, resistance to radiation, resistance to desiccation, halophilic, resistance to extreme pH, able to survive in cold temperatures (including cold waters), resistance to cleaning products, resistance to heavy metals, and an ability to metabolize unusual carbon sources.

Vendor C also used MetaSPAdes (Nurk et al., 2017) to generate assembled contigs from a variety of taxa. Long (>500 kbp) assembled contigs appear to be closely related to *Acinetobacter lwoffii* and *Bacillus cereus,* but do not precisely match any genome in the public databases. The long assembled contigs were covered by an average depth of 100X, indicating sufficient depth for accurate assembly. Contigs longer than 10 kbp were also assembled from a variety of other taxa. They noticed a large amount of *Pinus* sequences, but also assembled sequences from 20 plausible microbial taxa, including *Acinetobacter* and *Bacillus*. No sequence could be assembled from the two putative negative controls. GRiD was able to estimate the replication rate of the *Acinetobacter* assembly for all samples except controls (FC2, FC9, and MC2), and it was also able to estimate *Bacillus* replication rate for all samples with *Bacillus* (**Supplementary Figure 2A**). For both taxa, GRiD showed a high rate of replication (a GRiD score of 2 is a typical threshold for fast replication) in most samples. In samples where scores for both taxa could be obtained the GRiD scores for both taxa were similar, often with overlapping of the 95% confidence interval and similar mean estimates. Excluding two outliers, (S99 and S16) the mean estimate for each sample falls in a tight range. Techniques to estimate bacterial replication from metagenomic

data are novel and have never been applied to low-complexity or clean-room data yet, and as such, thus these results should be viewed as preliminary.

The abundance of different functional pathways in samples was identified using HUMAnN2 (Franzosa et al., 2018) with UniRef90 (Suzek et al., 2014). Dimensionality reduction of the functional pathways showed a group of SAF samples clustering with MetaSUB air samples and a group of samples clustering with surface samples (**Supplementary Figure 2B**). However, these clusters did not correspond at the taxonomic level. Vendor C identified 48 biochemical and biological pathways that were differentially abundant in SAF samples (20 samples) compared to both MetaSUB-surface (2,126 samples) and MetaSUB-air samples (317 samples) ($p < 0.01$, Mann-Whitney-*U* test). These pathways were universally more abundant in SAF samples than MetaSUB samples, and may represent specific adaptation to the cleanroom environment. Enriched pathways were principally for nucleotide and protein biosynthesis, including *de novo* biosynthesis of adenosine, guanosine, L-valine, L-isoleucine, L-lysine, L-proline, L-arginine, and L-methionine. Additionally, pathways associated with metabolism which may be linked to new metabolic sources were also enriched, including glycolysis, pentose-phosphate-pathway, the glyoxylate bypass, acetylene degradation, and fermentation to isobutanol. In particular, nucleotide biosynthesis has been linked to bacterial resistance to radiation and desiccation (Cox and Battista, 2005).

Vendor N Results

Primary quality control evaluation revealed that four of 20 samples (20%) were primarily composed of adapter dimer. One of these samples (S14) yielded no classifiable reads while the other three negative control samples (FC2, FC9, and MC2) showed considerably low number of taxa. Hence, Vendor N correctly concluded that FC2, FC9, and MC2 were negative controls, since the identity of the samples were not revealed to any vendors.

Taxonomy of the samples was determined using three methods: MTSv; MetaPhlAn2, and Bracken. MTSv uses an alignment method, which maps 50-mers from each sample to a local copy of the NCBI GenBank database. Metagenomic Phylogenetic Analysis (MetaPhlAn2) maps shotgun sequence reads to a database of clade-specific marker sequences and additionally determines relative abundance. The MetaPhlAn2 database contains one million genes that represent 17,000 reference genomes (bacterial, archaeal, viral, and eukaryotic). The third method, called Bayesian Reestimation of Abundance with Kraken (Bracken) performs k-mer mapping to a reduced version of the RefSeq database. All three of these methods provided generally similar results with limited but identifiable differences.

The MTSv method identified 189 taxa (155 bacteria, 4 fungi, 18 eukaryotes, and 2 viruses). Of these, 55 were called with very high confidence (40 bacteria, 14 eukaryotes, and 1 virus). MetaPhlAn2 called 121 bacteria, 1 eukaryote, and one virus. However, only 38 microbial taxa were found to be abundant (>1%; **Figure 1**) in SAF floor samples. Most of the bacterial genera and many of the species overlapped, but MTSv called species not identified by MetaPhlAn2 and the latter did not identify plant, *Homo sapiens* and other eukaryotes because they are not in the MetaPhlAn2 database. The Bracken results were more similar to those obtained using MetaPhlAn2, although additional species under the genera *Acinetobacter*, *Bacillus*, *Paracoccus*, *Pseudomonas*, and *Sphingomonas* were reported. A critical difference between read handling for MTSv and MetaPhlAn2 and Bracken is that MTSv reads were not trimmed prior to analysis, likely increasing overall sensitivity and decreasing specificity.

The relative abundance profiles provided by MetaPhlAn2 and Bracken revealed that in most cases, samples were composed of only a few dominant species and that those species were most frequently *Acinetobacter lwoffii*, a common skin inhabitant, or *Bacillus thuringiensis*, a common soil organism. Indeed, most of high abundance organisms identified by these methods were members of the human microbiome. In addition to these high abundance organisms, some low abundance organisms of interest were identified, such as the potentially radiotolerant *Methylobacterium radiotolerans* and *Deinococcus* sp., the UV resistant *Hymenobacter* sp. and the halotolerant *Kocuria rhizophila*. Several sphingomonads that have the potential to degrade hydrocarbons were also identified.

The bacteria identified were annotated using the KEGG and the HUMAnN2 tool. Vendor N examined the annotation tables for features associated with sporulation, peroxide resistance, halotolerance, cold shock, phosphate starvation and antibiotic and heavy metal resistance and discovered genes associated with all these functions (**Supplementary Table 2**). This includes 73 predicted sporulation proteins, including sigma factors, two component regulators, germination proteins, assembly proteins, stage-specific proteins, proteases, and coat proteins. A variety of heat shock proteins, including ECF sigma factors and chaperones, and cold shock proteins were annotated in numerous organisms. Catalase, heme and non-heme peroxidases, glutathione peroxidase and peroxiredoxin, which are all involved in resistance to peroxide were also found. Halotolerance factors, including glycine/betaine transporters or inositol monophosphatase, and genes associated with resistance to cadmium, camphor, chromate, cobalt, copper, fluoride, mercury, and tellurite were predicted. Additionally, genes associated with resistance to at least thirteen antibiotics, including vancomycin were found.

Comparative Analyses of All Vendors

Analyses of taxonomic tables (**Supplementary Table 3**) produced by all four vendors (J, C, N, and L) reveals differences in classification methodologies used. Different QC metrics used by the four vendors resulted in different numbers of sequences analyzed in subsequent stages (**Figure 1A** and **Supplementary Figure 1A**). Analyses of species-level taxon tables using stacked bar plots of relative species abundance (**Figure 1B** and **Supplementary Figure 1B**) reveal the differences in taxonomic classification among vendors. The number of taxa identified in each sample by the four vendors are depicted in **Figure 2A**. Although there is overlap in species identified (**Figure 2B**), no two vendors report the same relative

abundance (**Figure 1**), likely owing to different databases used during analysis. When using the same taxonomic classifier (MetaPhlAn2) with its own built-in database, there is a large similarity in relative abundance of species between vendor C and N that is easily visible in a stacked bar plot (**Supplementary Figure 3**), showing that earlier QC steps likely played a minor role in these different taxonomic classifications.

In low-biomass SAF floor samples (S014 and S099) that resulted in $<1 \times 10^6$ reads (sample group [L]ow), only vendors L, J, and C provided taxonomic classification. Vendor L describes these low-biomass samples as containing mainly human skin commensals, including *A. lwoffii* (16–50%), *A. johnsonii* (11–20%), *B. flexus* (1–23%), *P. acnes* (5–19%), *Phaeobacter gallaeciensis* (0–7%), *S. aureus* (0–4%), and *S. epidermidis* (1–14%). Vendor J describes these samples as containing *A. lwoffii* (<1–8%), *A. johnsonii* (<1-1%), *Cnuella takakiae* (0–9%), *S. aureus* (<1–18%), *Mycobacterium xenopi* (2–53%), *Sphingomonas paucimobilis* (0–15%), *Skermanella aerolata* (0–2%), and *A. baumannii* (1–3%). Vendor C describes these two low-biomass samples as containing *A. johnsonii* (1–25%), *A. schindleri* (<1–24%), *Acinetobacter* sp. NCu2D-2 (0–16%), *Acinetobacter* sp. TTH0-4 (0–7%), *Bullavirinae* (3–33%), *Enterobacteria* phage phiX174 *sensu lato* (3–33%), *Mycobacterium avium* (0–10%), and *Sphingomonas* sp. LK-11 (0–10%).

Taxonomic classifications for high-biomass samples collected from the SAF floor that resulted in $>1 \times 10^6$ reads (sample group H) were successfully provided by all four vendors without fail. Vendor L described these samples as being dominated by a diverse assemblage: *A. lwoffii* (1–55%), *A. johnsonii* (2–22%), *B. flexus* (0–24%), *Bacteroides fragilis* (0–5%), *B. plebeius* (0–10%), *Faecalibacterium prausnitzii* (0–8%), *Paracoccus aminovorans* (0–10%), *P. marcusii* (0–11%), *P. acnes* (<1–22%), *Rothia mucilaginosa* (0–4%), *Ruminococcus gnavus* (0–79%), *Sphingomonas yabuuchiae* (0–14%), *Staphylococcus aureus* (0–33%), *S. epidermidis* (0–19%), *S. haemolyticus* (0–11%), *Streptococcus infantis* (0–8%), and *Veillonella dispar* (0–7%). Vendor J also described these samples as being comprised of a diverse assemblage: an unidentified *Acidobacteria* sp. (0–6%), *A. baumannii* (<1–15%), *A. lwoffii* (<1–16%), *B. cereus* (<1–56%), *B. thuringiensis* (<1–34%), *Bacteroides plebeius* (0–5%), an unidentified *Brevundimonas* sp. (<1–13%), *Cellulomonas cellasea* (0–5%), *Comamonas terrigena* (0–6%), *Mycobacterium tuberculosis* (0–5%), *M. xenopi* (6–69%), *Paracoccus salipaludis* (<1–12%), *P. sediminis* (<1–45%), *Prevotella* sp. CAG:891 (0–24%), *Roseomonas nepalensis* (<1–10%), *Rubellimicrobium* sp. YIM 131921 (<1–10%), *Skermanella aerolata* (<1–15%), *S. aureus* (<1–27%), and *Xanthomonas campestris* (0–13%). Vendor C similarly described a diverse assemblage in the SAF floor samples, containing *A. baumannii* (<1–7%), *A. calcoaceticus* (<1–12%), *A. johnsonii* (<1–16%), *A. schindleri* (<1–15%), *Acinetobacter* sp. NCu2D-2 (<1–6%), *B. cereus* (0–64%), *B. mycoides* (0–22%), *B. thuringiensis* (0–23%), *Bullavirinae* (<1–14%), an unidentified *Dikarya* sp. (0–10%), *Enterobacteria* phage phiX174 *sensu lato* (<1–14%), *Leishmania major* (0–18%), *Mycobacterium avium* (<1–10%), *Negativicoccus massiliensis* (0–5%), *Pseudomonas stutzeri* (<1–21%), *Rothia mucilaginosa* (0–9%), *Streptococcus parasanguinis* (0–7%), and *Xanthomonas campestris* (0–10%). Vendor N described these samples as being composed of mainly *A. lwoffii* (3–91%), *B. thuringiensis* (<1–92%), *Bacteroides coprocola* (0–10%), *Mycobacterium xenopi* (2–39%), *Rickettsia felis* (0–9%), *Ruminococcus gnavus* (0–9%), *S. aureus* (0–13%), *Staphylococcus* phage PVL (0–32%), and an unidentified *Veillonella* sp. (0–12%).

High-biomass samples collected from the SAF floor that resulted in $>1 \times 10^6$ reads and contained a high abundance of Firmicutes or Actinobacteria (sample group H-F) were analyzed by all vendors. Vendor L described these samples as containing *A. johnsonii* (2–11%), *A. lwoffii* (2–19%), *B. flexus* (1–24%), *Faecalibacterium prausnitzii* (0–8%), *P. acnes* (<1–22%), *S. aureus* (4–33%), and *S. epidermidis* (1–19%). Vendor J described these samples as containing *B. cereus* (<1–56%), *B. thuringiensis* (<1–9%), *Mycobacterium xenopi* (6–69%), *Paracoccus sediminis* (<1–6%), and *Skermanella aerolata* (<1–15%) in abundance. Vendor C described these three samples as containing *B. cereus* (0–64%), *B. mycoides* (0–22%), *B. thuringiensis* (0–12%), *Bullavirinae* (<1–14%), an unidentified *Dikarya* sp. (0–10%), *Enterobacteria* phage phiX174 *sensu lato* (<1–14%), and *Mycobacterium avium* (<1–6%). Vendor N described these three samples as being dominated by *A. lwoffii* (3–59%), *B. thuringiensis* (0–92%), and *Mycobacterium xenopi* (2–39%).

A Venn diagram representing overlap in species detected in SAF floor samples by the four vendors (**Figure 2B**) demonstrates the similarity and differences between the methods used. All four vendors detected three microbial species in common from floor samples, including *A. johnsonii*, *P. stutzeri*, and *S. aureus*. Species detected by only three vendors include *A. lwoffii*, *B. thuringiensis*, *Bacteroides plebeius*, *E. coli*, *Haemophilus parainfluenzae*, *Rickettsia felis*, *Rothia mucilaginosa*, and *Ruminococcus gnavus*. All vendors classified species from floor samples not classified by other vendors, with Vendor L uniquely classifying 64 of 82 species (78%), Vendor J uniquely classifying 82 of 110 species (75%), Vendor C uniquely classifying 59 of 83 species (71%), and Vendor N uniquely classifying 18 of 38 species (47%).

By plotting SAF floor and control samples with respect to the species classified by each vendor using non-metric multidimensional scaling (NMDS; **Figure 3**), the differences between methods used becomes very apparent. For the most part, point clusters representing SAF floor samples taxonomically classified by each vendor do not share the same ordination space with other vendors, with disjoint 99% confidence intervals for the sample mean (lightly colored ellipses). Control samples analyzed by all vendors (square glyphs) cluster separately (empty ellipse) from the SAF floor samples. Analysis of similarities of species classification among vendors shows that there is more similarity between samples analyzed by the same vendor than there is a similarity between vendors that is statistically significant (ANOSIM; $R = 0.6056$, $P = 0.001$).

All four vendors detected *S. aureus*, a skin-associated bacterium, as a dominant species in control samples (**Supplementary Figure 1B**). *S. aureus* makes up 34–97%

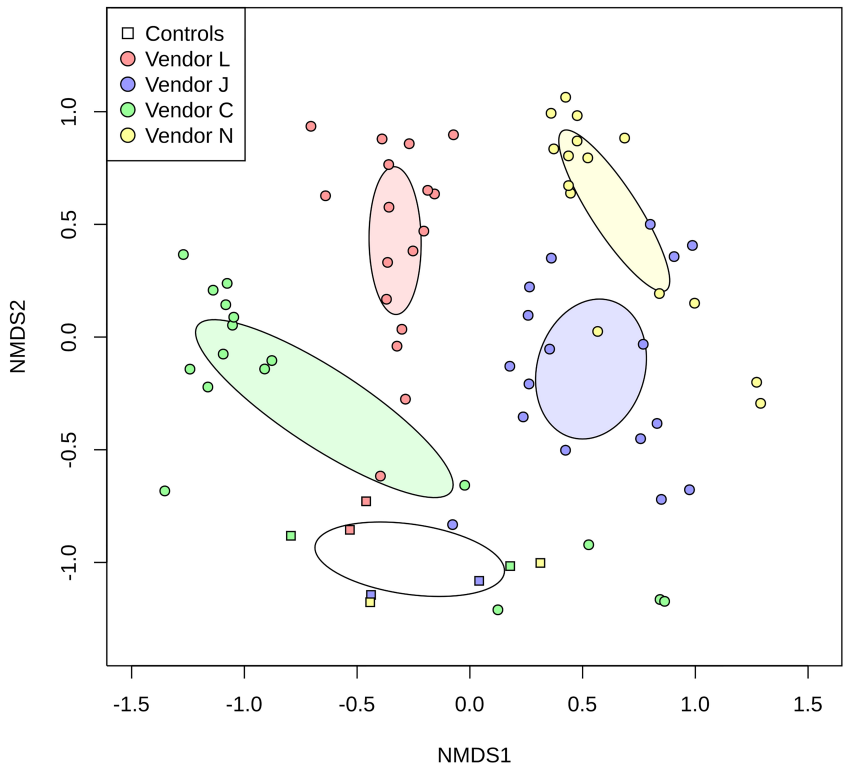


FIGURE 3 | Non-metric multidimensional scaling (NMDS) plot representing differences in species detected in SAF floor samples (circle glyphs) and control samples (square glyphs). Lighter colored ellipses represent the 99% confidence interval for the mean of the SAF floor samples for each vendor. The empty ellipse represents the 99% confidence interval for the mean of the control samples for all vendors.

of handling control sample FC2 (4.2×10^4 reads), and 14–66% of Maxwell reagent control MC2 (1.56×10^5 reads). *B. thuringiensis, B. cereus*, and *B. flexus* which are closely related, spore-forming, soil-associated bacteria, also made up a large proportion (9–66%) of Maxwell reagent control sample MC2. However, Vendor L only identified *B. flexus* in this sample and Vendor N only identified *B. thuringiensis*, while Vendor J and Vendor C identified closely related *B. cereus* and *B. thuringiensis*.

Analysis of overlap in the dominant species (>1%) detected in control samples (kitome) by each vendor (**Supplementary Figure 1C**) reveals limited similarity. However, the number of taxonomically classified reads (>1%) in these libraries were extremely low, ranging from 9 to 57,179 reads. All four vendors detected one species in common, *S. aureus*. Vendor C, Vendor N, and Vendor J additionally detected *B. thuringiensis*. All vendors but Vendor N uniquely identified species in the control samples: Vendor L uniquely identified 13 of 17 species, Vendor J uniquely identified 41 of 50 species, and Vendor C uniquely identified 12 of 19 species.

Comparative Analyses of All Vendors on Planetary Protection Relevant Microbial Species

The Space Studies Board suggested classifying microbial species based on the level of concern that the species could survive unintentional transport via spacecraft and colonize a planetary body such as Mars or Jupiter's icy moon Europa (National Research Council, 2000). To this end, they created four categories: Type A, microorganisms cultivable on the NSA without heat shock; Type B, spore-forming microorganisms (NSA heat-shock); Type C, radiation-resistant (10% survival at 0.8 Mrad), spore-forming microorganisms; and Type D, radiation-resistant (10% survival at 4.0 Mrad), non-spore-forming microorganisms. Although microorganisms classified as Type C and Type D are the most likely to survive space flight, we consider microorganisms classified as Type B, Type C, and Type D, or extremophiles suited to mission-specific environments as potentially PP relevant microbial species for this analysis.

All four vendors detected bacteria that may be relevant for planetary protection purposes (see red nodes in **Figure 4**). Radiation resistant bacteria associated with *Acinetobacter* (*A. baumannii*, *A. johnsonii*, *A. lwoffii*, and *A. schindleri*), *Brevundimonas diminuta*, and an unidentified *Brevundimonas* sp. were detected by multiple vendors in SAF samples. Additionally, Vendor C detected other *Acinetobacter* species in SAF floor samples, including *A. calcoaceticus* and other *Acinetobacter* species (strains LoGeW2-3, NCu2D-2, and TTH0-4). An unidentified *Acidobacteria* sp., which could be an extremophile (Kielak et al., 2016), was detected by Vendor J in SAF floor samples. Notably, all four vendors detected members of *Bacillus*, a genus of microorganisms of PP concern with spore-forming

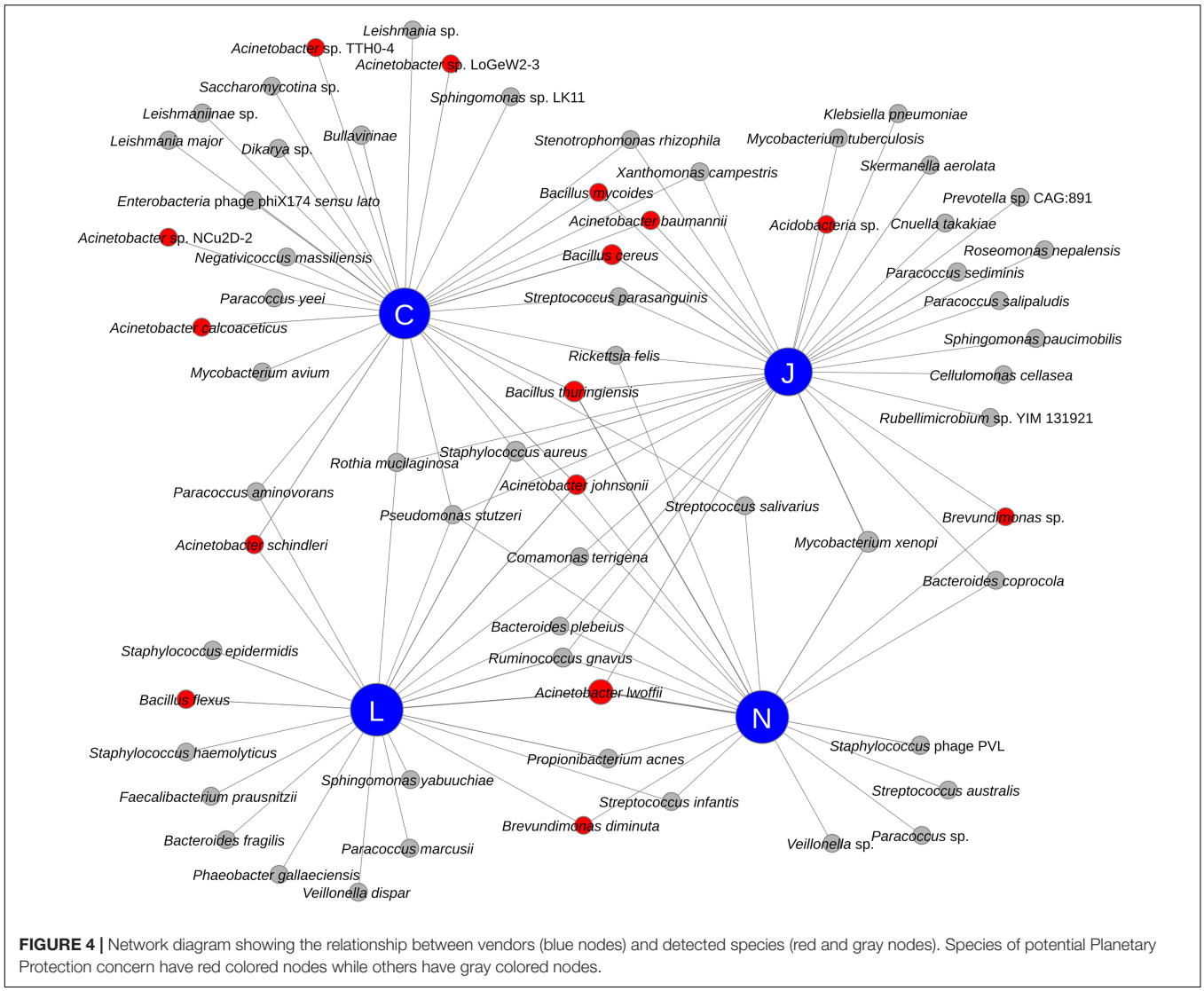


FIGURE 4 | Network diagram showing the relationship between vendors (blue nodes) and detected species (red and gray nodes). Species of potential Planetary Protection concern have red colored nodes while others have gray colored nodes.

capabilities. *B. cereus* was detected by Vendor J and Vendor C; *B. flexus* was detected by Vendor L; *B. mycoides* was detected by Vendor J and Vendor C; and *B. thuringiensis* was detected by Vendors C, N, and J.

Among the functional characteristics, genes associated with dormancy and sporulation pathways were detected by Vendors J and N. These include sigma factors, two component regulators, germination proteins, assembly proteins, stage-specific proteins, proteases, and coat proteins. Additionally, these two vendors detected gene pathways involved in resistance to antimicrobial agents, heavy metals, and resistance to stress (e.g., aerotolerance, heat shock, cold shock, and osmotic stress), potentially showing adaptation of these microorganisms to the clean-room environment.

In field control sample (FC2), spore-forming bacteria including *B. cereus*, *B. flexus*, and *B. thuringiensis* were also detected but in low numbers. In addition, the spore-forming *Anoxybacillus kestanbolensis* was also detected in the Maxwell reagent control sample (MC2) by Vendor L.

DISCUSSION

NASA PP efforts are concerned with hitch-hiking microorganisms that may have the capability to withstand spaceflight conditions and contaminate a planetary body of scientific interest. Concerns regarding the inability of the NSA culture method to detect all microorganisms in a sample has led NASA to look toward state-of-the art molecular methods like shotgun metagenomics to detect a wider range of taxa on spacecraft and associated environments. All four vendors used in this study identified taxa of PP concern in the JPL SAF environment where spacecraft are built, including *Acidobacteria*, *Acinetobacter, Anoxybacillus*, *Bacillus, Brevundimonas*, and *Caulobacteraceae*, demonstrating that modern metagenomic approaches and computational biology can detect these organisms of concern, unlike NSA culture methods.

Although similar species were detected by the four vendors, one issue noted by these analyses is that the relative abundances varied drastically between vendors. Different QC measures likely

played a minor role in this since Vendor C and N used different QC methods but arrived at similar results when using the same taxonomic classifier (MetaPhlAn2; **Supplementary Figure 3**). The largest contributors to this difference in relative abundance is likely due to the different programs and databases used. Although Vendor L, Vendor J, and Vendor C used similar sequence-based analyses, their results varied drastically due to different databases used. Vendor C and Vendor N both used MetaPhlAn2 with its built-in database to arrive at very similar results, but different from the results produced by Vendor J and Vendor L, and Vendor C when using KrakenUniq. MetaPhlAn2 reports relative abundances in terms of the number of cells (Segata et al., 2012) rather than fraction of reads as done by the other analyses, likely contributing to this observed difference.

By ignoring relative abundance calculations and simply examining the overlap in species detected (**Figure 2B**), the difference between vendors becomes clearer. The overlap in species identified in SAF floor samples between the four vendors is small (three species). Vendor J and Vendor C shared the largest overlap in species identified (17 species). Vendor J identified more species than other vendors (110), likely due to the vast size of the database they used. All vendors uniquely identified species not identified by the other vendors. Failure to detect microorganisms present in a sample is concerning for PP purposes, especially if the missed microorganisms could survive on a spacecraft and go on to contaminate a planetary body of scientific interest. Utilization of large databases containing a wide breadth of available reference genome sequences may help to alleviate this issue.

Falsely detected species may also be of concern, especially if the detected species is of potential PP concern. Falsely detecting such a species would needlessly require extra rounds of cleaning, negatively affect the schedule of a spacecraft build, and potentially delay its launch. The large number of uniquely detected species by each vendor is especially concerning in this regard. Some of these uniquely identified species may have simply been miscategorized due to lack of a better alternative in the databases used or sequence fragments aligning to genomic regions with poor phylogenetic resolution. Adding a requirement for minimum coverage of a genome prior to classifying a species as present in a sample may help to limit such false positives. In addition, computational analysis at the genus level may also solve this problem since short reads of metagenome sequences may not have the taxonomic resolution to differentiate taxa at the species level.

Some of the same species seen in SAF floor samples can be seen in control samples, alluding to potential contamination of sampling equipment (e.g., the wipes) or reagents. Simply excising species with an abundant number of reads found in control samples from environmental samples is an option, but this may be problematic for PP purposes given that some of the species detected in the control samples are organisms of potential PP concern. The QC read counts were extremely low in these control samples, and the PP relevant microorganisms detected (*Bacillus* spp. and *Anoxybacillus kestanbolensis*) likely had a minor influence on the species detected in the floor samples. Compared to control samples, QC reads from floor samples were much more abundant (**Table 1**). Among control samples, *S. aureus* constituted 47–97% of a sample, whereas *S. aureus* constituted 1–11% in corresponding floor samples. Furthermore, viability assessment of the microbes would help to eliminate the kitomes from the calculations. In addition to pretreating with propidium monoazide (Vaishampayan et al., 2013), application of novel technique such as GRiD analysis would eliminate naked DNA sequences associated with kitomes. In this study, Vendor C was able to estimate the replication rate of some taxa from SAF samples but such analysis did not show replication of taxa in control samples (FC2, FC9, and MC2).

The variance in relative abundance seen between vendors may make metagenomic assays problematic for PP purposes without additional work on calibration of these metagenomic pipelines. The ability to estimate the number of problematic microorganisms on a spacecraft surface is currently hampered by these large variations between methods and databases used, but is paramount for determining the amount of bioburden on these surfaces. Synthetic metagenomes containing sequence fragments in known quantities have been developed (McIntyre et al., 2017), and can be utilized for this purpose. However, it would be beneficial to develop synthetic metagenomes from the genomes of microorganisms that PP efforts are concerned with, in addition to microorganisms previously detected in PP-relevant samples, to better test and calibrate these pipelines. Since SAF cleanrooms are known to be similar to other built environments with regular human movement (Lax et al., 2014), inclusion of microorganisms known to inhabit these environments in a synthetic metagenome would be important. Additionally, procedural changes that include spiking a known quantity of DNA into samples prior to sequencing should be considered to allow for more precise estimates of microbial abundance (Hardwick et al., 2018; Blackburn et al., 2019).

Only ∼15% of the QC sequences were utilized for identification of species by all vendors except Vendor L (<1%), likely due to the non-availability of annotated genomes in public databases. Public databases are mainly comprised of cultivable microorganisms with annotations, whereas very few genomes of non-cultivable microorganisms are present. To enable the utilization of discarded shotgun metagenome reads, metagenome assembled genomes (MAG) of microorganisms present in NASA cleanroom samples should be generated, allowing for annotation of their genetic capabilities and detection of PP-relevant metabolisms. Since spacecraft associated surfaces are low in biomass (Minich et al., 2018), MAGs should additionally be generated from the samples collected from PP relevant extreme environments such as radionucleotide dumping sites, nuclear accidents, and hot and cold arid regions of the Earth. Additionally, MAGs should be generated from samples collected from built-environments that do not undergo strict cleaning regimes and are directly adjacent to the relevant cleanrooms, to ensure that the genetic repertoire of microorganisms that may hitch-hike between the two are captured. MAGs from such environments would help to identify and annotate these uncultivable microorganisms to decipher the metabolic and functional pathways and expand the understanding of species diversity of spacecraft and associated environments.

This study was performed using short metagenomic sequences (∼150-bp) generated using Illumina technology, which is known to function with low-biomass clean room samples. The vast number of sequences that can be produced by this technology usually offsets its main drawback, the short length of the sequences. Short sequences are often harder to identify with certainty because they align to highly conserved segments of a genome and provide no specific information, or because they represent novelty that is not yet represented in a public database. Longer sequences generated by technologies developed by Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT) provide an opportunity to more accurately identify a larger portion of the sequences generated because the ambiguous segments are often linked to segments that can be identified. Technologies that generate longer sequences should be considered for future metagenomic studies of spacecraft hardware to aid in identification of PP relevant taxa. The issues associated with long read technologies like ONT MinION or PacBio are that these methodologies require higher concentration of DNA as well as longer fragment length. Such issues can be mitigated when larger surface areas are collected and gentle sample processing technologies developed to get longer fragment DNA length. These tasks are not trivial to perform in samples collected from an extremely oligotrophic, clean, low-biomass spacecraft and associated environments.

Each vendor in this study utilized different computational algorithms and pipelines to analyze the provided shotgun metagenome data, and here we attempt to highlight some of the benefits and downfalls of each approach. Vendor C and Vendor N used multiple methods for taxonomic identification (Vendor C used KrakenUniq and MetaPhlAn2; Vendor N used MTSv, MetaPhlAn2, and Kraken/Braken), which allowed them to have more confidence in species identified via multiple methods. However, Vendor N did not provide taxon tables for their MTSv or Kraken/Braken results, so comparative analyses with these pipelines were not possible. The taxon table provided by Vendor L identified <1% of QC sequences, with an anomaly in sample S99 where 12,223 reads were taxonomically classified, but only 2,671 reads passed their QC stage (458% identified), bringing into question the overall quality of their results. Vendor J used a massive database (NCBI-NR) that allowed them to identify species not detected by other vendors, however, the degenerate nature of protein sequences (with to regard to DNA sequences) means that their pipeline would have a hard time differentiating between two closely related microorganisms. Vendors L, C, and N all used databases containing DNA sequences for identification potentially allowing for more specific results, but increased computation time to analyze DNA sequences limited the size of databases that could be used. The custom database used by Vendor L to analyze the shotgun metagenome sequences was too limited in genetic breadth to accurately describe the diversity present in the samples provided to them. Vendor C and Vendor J generated MAGs from the dataset provided, giving these vendors more confidence in the species identified. These MAGs additionally allowed Vendor C to estimate the rate of growth of these microorganisms by using the GRiD software package to map that ratio of DNA close to the origin of replication with that of the terminus region. Although no results were provided, Vendor L claims to be working on novel machine-learning/artificial-intelligence algorithms that could help increase accuracy of results.

CONCLUSION

Although metagenomics offers tantalizing access to the genetic and functional potential of a microbial community, more work needs to be carried out to standardize ever evolving analyses pipelines and databases used to understand big-data generated from a low-biomass environment. To standardize these metagenomics sequence analyses pipelines, synthetic metagenomes containing known quantities should be developed and utilized. Additionally, the use of sequencing technologies that produce longer sequences should be considered to aid in the identification of PP-relevant microorganisms. Even though all the vendors in this study utilized different computational algorithms and pipelines to analyze the provided shotgun metagenome data, Vendor J approaches allowed them to identify species not detected by other vendors. Invariably, all vendors utilized could only resolve about 15% of the shotgun sequences since public databases contain only annotated sequences pertaining to culturable microorganisms. Metagenomics has the potential to be a useful assay for detecting microorganisms of PP concern and explore functional characteristics of yet-to-be cultured microorganisms by constructing MAGs.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: BioProject PRJNA668809.

AUTHOR CONTRIBUTIONS

KV and JB coordinated with all authors in designing the concept, executed the study, implemented the project, involved in data analyses, communicated with all vendors in producing the data, and involved in the logistical efforts of the work. JW and KV wrote the manuscript. JW generated figures and tables for the manuscript and performed all statistical analyses. NS implemented some of the bioinformatics analyses. AS generated a figure for the manuscript. LG and AS curated metadata and organized the sequence data. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the NASA Planetary Protection Office funds to JB which also funded a portion of the fellowship of JW. The funders had no role in study design, data collection and interpretation, the writing of the manuscript, or the decision to submit the work for publication.

ACKNOWLEDGMENTS

The research described in this manuscript was performed at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. The authors would like to thank Kyle Landry (Liberty BioSecurity), Sara Highlander (T-Gen) and Paul Kiem (Northern Arizona University), and Chris Mason (Weill Cornell Medicine) for their support in treating data with various metagenomics pipelines. ©2021 California Institute of Technology. Government sponsorship acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.685254/full#supplementary-material

Supplementary Figure 1 | Abundance of reads in control samples that passed quality control (QC) measures of each vendor (dark- and light-blue; gray represents reads discarded by the QC stage; dark-blue represent QC reads taxonomically classified) **(A)**, relative abundance of species detected by each vendor **(B)**, Venn diagram representing overlap in species detected in control samples **(C)**, and network diagram showing the relationship between vendors (blue nodes) and detected species (red and gray nodes; red nodes are species of potential Planetary Protection concern) in control samples **(D)**. The FC9 sample is not included due to the low number of reads (n = 44) as well as their taxonomic affiliation to the genus and species was not resolved.

Supplementary Figure 2 | Estimated replication score for assembled *Acinetobacter* and *Bacillus* species in each sample **(A)**. Higher scores indicate faster replication, range indicates the 95% confidence interval. Some samples were not scored for one or both species because that species was not detected. Values are similar when reference genomes are substituted for assemblies. Red glyphs are *Acinetobacter* and blue are *Bacillus*. Comparison of cleanroom samples with the MetaSUB database **(B)**. Both panels (left and right) show cleanroom samples plotted with MetaSUB samples as the background. Plotting was performed using UMAP to reduce data to two dimensions. The input data were binary, indicating presence or absence of a feature. Left shows microbial species, right shows microbial metabolic pathways.

Supplementary Figure 3 | Relative abundance of species detected by Vendor C and Vendor N when both vendors use the same taxonomic classification algorithm (MetaPhlAn2).

Supplementary Table 1 | Dormancy and sporulation subsystems **(A)** and all subsystems **(B)** detected by vendor J using the SEED database.

Supplementary Table 2 | Genes and predicted proteins associated with sporulation, heat shock, peroxide resistance, halotolerance, cold shock, phosphate starvation, and antibiotic and heavy metal resistance detected by vendor N using HUMAnN2.

Supplementary Table 3 | Taxonomic classification of SAF floor samples **(A)** and control samples **(B)** performed by each vendor.

REFERENCES

Avila-Herrera, A., Thissen, J., Urbaniak, C., Be, N. A., Smith, D. J., Karouia, F., et al. (2020). Crewmember microbiome may influence microbial composition of ISS habitable surfaces. *PLoS One* 15:e0231838. doi: 10.1371/journal.pone.0231838

Beaty, D. W., Mcsween, H. Y., Carrier, B. L., Czaja, A. D., Goreva, Y. S., Hausrath, E. M., et al. (2018). "Analysis of the Scientific Value of the Mars2020 Spacecraft Genetic Inventory to Mars Sample Return," in *49th Lunar and Planetary Science Conference*, (Houston, TX).

Benardini, J. N., La Duc, M. T., Beaudet, R. A., and Koukol, R. (2014). Implementing planetary protection measures on the mars science laboratory. *Astrobiology* 14, 27–32. doi: 10.1089/ast.2013.0989

Blackburn, J., Wong, T., Madala, B. S., Barker, C., Hardwick, S. A., Reis, A. L. M., et al. (2019). Use of Synthetic DNA spike-in controls (sequins) for human genome sequencing. *Nat. Protoc.* 14, 2119–2151. doi: 10.1038/s41596-019-0175-1

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Breitwieser, F. P., Baker, D. N., and Salzberg, S. L. (2018). KrakenUniq: confident and fast metagenomics classification using unique K-mer Counts. *Genome Biol.* 19:198.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Chung, S., Kern, R., Koukol, R., Barengoltz, J., and Cash, H. (2008). Vapor hydrogen peroxide as alternative to dry heat microbial reduction. *Adv. Space Res.* 42, 1150–1160. doi: 10.1016/j.asr.2008.01.005

Cleary, B., Brito, I. L., Huang, K., Gevers, D., Shea, T., Young, S., et al. (2015). Detection of low-abundance bacterial strains in metagenomic datasets by eigengenome partitioning. *Nat Biotechnol.* 33, 1053–1060. doi: 10.1038/nbt.3329

COSPAR (2011). *COSPAR Planetary Protection Policy*. Houston, TX: World Space Council.

Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., et al. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* 35, 1069–1076.

Cox, M. M., and Battista, J. R. (2005). *Deinococcus radiodurans* — the consummate survivor. *Nat. Rev. Microbiol.* 3, 882–892. doi: 10.1038/nrmicro1264

Emiola, A., and Oh, J. (2018). High throughput *in situ* metagenomic measurement of bbacterial replication at ultra-low sequencing coverage. *Nat. Commun.* 9:4956.

Franzosa, E. A., Mciver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. doi: 10.1038/s41592-018-0176-y

Hardwick, S. A., Chen, W. Y., Wong, T., Kanakamedala, B. S., Deveson, I. W., Ongley, S. E., et al. (2018). Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. *Nat. Commun.* 9:3096.

Hart, Y., Sheftel, H., Hausser, J., Szekely, P., Ben-Moshe, N. B., Korem, Y., et al. (2015). Inferring biological tasks using pareto analysis of high-dimensional data. *Nat. Methods* 12, 233–235. doi: 10.1038/nmeth.3254

Heberle, H., Meirelles, G. V., Da Silva, F. R., Telles, G. P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through venn diagrams. *BMC Bioinform.* 16:169. doi: 10.1186/s12859-015-0611-3

Hendrickson, R., Lundgren, P., Malli Mohan, G. B., Urbaniak, C., Benardini, J. N., and Venkateswaran, K. (2017). "Comprehensive Measurement of Microbial Burden in Nutrient-Deprived Cleanrooms," in *47th International Conference on Environmental Systems*, (Charleston, SC: International Conference on Environmental Systems).

Horneck, G., Moeller, R., Cadet, J., Douki, T., Mancinelli, R. L., Nicholson, W. L., et al. (2012). Resistance of bacterial endospores to outer space for planetary protection purposes—experiment PROTECT of the EXPOSE-E Mission. *Astrobiology* 12, 445–456. doi: 10.1089/ast.2011.0737

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12:e1004957. doi: 10.1371/journal.pcbi.1004957

Huson, D. H., Tappu, R., Bazinet, A. L., Xie, C., Cummings, M. P., Nieselt, K., et al. (2017). Fast and simple protein-alignment-guided assembly of orthologous gene families from microbiome sequencing reads. *Microbiome* 5:11.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., Van Veen, J. A., and Kuramae, E. E. (2016). The ecology of acidobacteria: moving beyond genes and genomes. *Front. Microbiol.* 7:744. doi: 10.3389/fmicb.2016.00744

La Duc, M. T., Dekas, A., Osman, S., Moissl, C., Newcombe, D., and Venkateswaran, K. (2007). Isolation and characterization of bacteria capable of tolerating the extreme conditions of clean room environments. *Appl. Environ. Microbiol.* 73, 2600–2611. doi: 10.1128/aem.03007-06

La Duc, M. T., Nicholson, W., Kern, R., and Venkateswaran, K. (2003). Microbial characterization of the mars odyssey spacecraft and its encapsulation facility. *Environ. Microbiol.* 5, 977–985. doi: 10.1046/j.1462-2920.2003.00496.x

La Duc, M. T., Vaishampayan, P., Nilsson, H. R., Torok, T., and Venkateswaran, K. (2012). Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for mars. *Appl. Environ. Microbiol.* 78, 5912–5922. doi: 10.1128/aem.01435-12

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lax, S., Smith, D. P., Hampton-Marcell, J., Owens, S. M., Handley, K. M., Scott, N. M., et al. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* 345, 1048–1052. doi: 10.1126/science.1254529

Levy, R., and Borenstein, E. (2012). "Reverse Ecology: From Systems to Environments and Back," in *Evolutionary Systems Biology*, ed. O. Soyer (New York, NY: Springer).

Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 3:e104. doi: 10.7717/peerj-cs.104

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of K-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011

Mason, C., Afshinnekoo, E., Ahsannudin, S., Ghedin, E., Read, T., Fraser, C., et al. (2016). The Metagenomics and metadesign of the subways and urban biomes (MetaSUB) international consortium inaugural meeting report. *Microbiome* 4:24.

McInnes, L., Healy, J., and Melville, J. (2020). *UMAP: uniform manifold approximation and projection for dimension reduction*. Ithaca, NY: arXiv e print.

McIntyre, A. B. R., Ounit, R., Afshinnekoo, E., Prill, R. J., Hénaff, E., Alexander, N., et al. (2017). Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biol.* 18:182.

Minich, J. J., Zhu, Q., Janssen, S., Hendrickson, R., Amir, A., Vetter, R., et al. (2018). KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems* 3, e217–e218.

Moissl, C., Bruckner, J. C., and Venkateswaran, K. (2008). Archaeal diversity analysis of spacecraft assembly clean rooms. *ISME J.* 2, 115–119. doi: 10.1038/ismej.2007.98

NASA. (2010). *Handbook for the Microbiological Examination of Space Hardware, NASA-HDBK-6022*. Washington, DC: National Aeronautics and Space Administration.

National Research Council. (2000). *Preventing the Forward Contamination of Europa*. Washington, DC: The National Academies Press.

Nayfach, S., and Pollard, K. S. (2016). Toward accurate and quantitative comparative metagenomics. *Cell* 166, 1103–1116. doi: 10.1016/j.cell.2016.08.007

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., et al. (2013). *Vegan: Community Ecology Package". 2.0-10 ed*. Available online at: https://CRAN.R-project.org/package=vegan (accessed Nov 28, 2020).

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

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740. doi: 10.1126/science.276.5313.734

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., et al. (2011). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40, D284–D289.

Probst, A., Vaishampayan, P., Osman, S., Moissl-Eichinger, C., Andersen, G. L., and Venkateswaran, K. (2010). Diversity of anaerobic microbes in spacecraft assembly clean rooms. *Appl. Environ. Microbiol.* 76, 2837–2845. doi: 10.1128/aem.02167-09

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2008). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 37, D5–D15.

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9:88. doi: 10.1186/s13104-016-1900-2

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shirey, T. B., Schubert, W., and Benardini, J. (2017). "An Overview of Surface Heat Microbial Reduction as a Viable Microbial Reduction Modality for Spacecraft Surfaces," in *47th International Conference on Environmental Systems*, (Charleston, S C: International Conference on Environmental System).

Singh, N. K., Wood, J. M., Karouia, F., and Venkateswaran, K. (2018). Succession and persistence of microbial communities and antimicrobial resistance genes associated with International Space Station environmental surfaces. *Microbiome* 6:214.

Sterflinger, K., De Baere, R., De Hoog, G. S., De Wachter, R., Krumbein, W. E., and Haase, G. (1997). *Coniosporium perforans* and *C. apollinis*, two new rock-inhabiting fungi isolated from marble in the Sanctuary of Delos (Cyclades. Greece). *Antonie van Leeuwenhoek* 72, 349–363.

Suzek, B. E., Wang, Y., Huang, H., Mcgarvey, P. B., Wu, C. H., and The Uniprot Consortium. (2014). UniRef Clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31, 926–932. doi: 10.1093/bioinformatics/btu739

Tamames, J., and Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol*:9:3349. doi: 10.3389/fmicb.2018.03349

Thompson, L. R., Sanders, J. G., Mcdonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals earth's multiscale microbial diversity. *Nature* 551, 457–463.

Vaishampayan, P., Osman, S., Andersen, G., and Venkateswaran, K. (2010). High-Density 16S microarray and clone library–based microbial community composition of the phoenix spacecraft assembly clean room. *Astrobiology* 10, 499–508. doi: 10.1089/ast.2009.0443

Vaishampayan, P., Probst, A. J., La Duc, M. T., Bargoma, E., Benardini, J. N., Andersen, G. L., et al. (2013). New perspectives on viable microbial communities in low-biomass cleanroom environments. *ISME J.* 7, 312–324.

Vaishampayan, P. A., Rabbow, E., Horneck, G., and Venkateswaran, K. J. (2012). Survival of *Bacillus pumilus* spores for a prolonged period of time in real space conditions. *Astrobiology* 12, 487–497. doi: 10.1089/ast.2011.0738

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome Biol.* 20:257.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wood, Singh, Guan, Seuylemezian, Benardini and Venkateswaran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 October 2021
doi: 10.3389/fmicb.2021.740909

Phylogenomics of Haloarchaea: The Controversy of the Genera *Natrinema-Haloterrigena*

*Rafael R. de la Haba[1], Hiroaki Minegishi[2], Masahiro Kamekura[3], Yasuhiro Shimane[4] and Antonio Ventosa[1]**

[1] Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [2] Department of Applied Chemistry, Faculty of Science and Engineering, Toyo University, Kawagoe, Japan, [3] Halophiles Research Institute, Chiba, Japan, [4] Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan

The haloarchaeal genera *Natrinema* and *Haloterrigena* were described almost simultaneously by two different research groups and some strains studied separately were described as different species of these genera. Furthermore, the description of additional species were assigned to either *Natrinema* or *Haloterrigena*, mainly on the basis of the phylogenetic comparative analysis of single genes (16S rRNA gene and more recently *rpoB'* gene), but these species were not adequately separated or assigned to the corresponding genus. Some studies suggested that the species of these two genera should be unified into a single genus, while other studies indicated that the genera should remain but some of the species should be reassigned. In this study, we have sequenced or collected the genomes of the type strains of species of *Natrinema* and *Haloterrigena* and we have carried out a comparative genomic analysis in order to clarify the controversy related to these two genera. The phylogenomic analysis based on the comparison of 525 translated single-copy orthologous genes and the Overall Genome Relatedness Indexes (i.e., AAI, POCP, ANI, and dDDH) clearly indicate that the species *Haloterrigena hispanica*, *Haloterrigena limicola*, *Haloterrigena longa*, *Haloterrigena mahii*, *Haloterrigena saccharevitans*, *Haloterrigena thermotolerans*, and *Halopiger salifodinae* should be transferred to the genus *Natrinema*, as *Natrinema hispanicum*, *Natrinema limicola*, *Natrinema longum*, *Natrinema mahii*, *Natrinema saccharevitans*, *Natrinema thermotolerans*, and *Natrinema salifodinae*, respectively. On the contrary, the species *Haloterrigena turkmenica*, *Haloterrigena salifodinae*, and *Haloterrigena salina* will remain as the only representative species of the genus *Haloterrigena*. Besides, the species *Haloterrigena daqingensis* should be reclassified as a member of the genus *Natronorubrum*, as *Natronorubrum daqingense*. At the species level, *Haloterrigena jeotgali* and *Natrinema ejinorense* should be considered as a later heterotypic synonyms of the species *Haloterrigena* (*Natrinema*) *thermotolerans* and *Haloterrigena* (*Natrinema*) *longa*, respectively. Synteny analysis and phenotypic features also supported those proposals.

Keywords: haloarchaea, *Halobacteria*, *Natrinema*, *Haloterrigena*, comparative genomic analysis, taxophylogenomic analysis

OPEN ACCESS

Edited by:
Richard Allen White III, University of North Carolina at Charlotte, United States

Reviewed by:
Munusamy Madhaiyan, Temasek Life Sciences Laboratory, Singapore
Aharon Oren, Hebrew University of Jerusalem, Israel
Heng-Lin Cui, Jiangsu University, China

***Correspondence:**
Antonio Ventosa
ventosa@us.es

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

Received: *13 July 2021*
Accepted: *25 August 2021*
Published: *07 October 2021*

Citation:
de la Haba RR, Minegishi H, Kamekura M, Shimane Y and Ventosa A (2021) Phylogenomics of Haloarchaea: The Controversy of the Genera Natrinema-Haloterrigena. Front. Microbiol. 12:740909. doi: 10.3389/fmicb.2021.740909

INTRODUCTION

Haloarchaea are a monophyletic group of extremely halophilic archaea affiliated to the single class *Halobacteria*, belonging to the phylum *Euryarchaeota* (Oren et al., 2017). Currently, the class *Halobacteria* comprises three orders (i.e., *Halobacteriales*, *Haloferacales*, and *Natrialbales*), six families (i.e., *Halobacteriaceae*, *Haloarculaceae*, *Halococcaceae*, *Haloferacaceae*, *Halorubraceae*, and *Natrialbaceae*), 72 genera and 289 species whose names have been validly published (Parte et al., 2020), reflecting the high diversity and complex phylogenetic relationships within the haloarchaea. In fact, recent pan-genome analysis and ancestral state reconstruction has brought to light the heterogeneity of this class, which possesses an open pan-genome, and the occurrence of genome expansion and horizontal gene transfer during the evolution of *Halobacteria* (Gaba et al., 2020).

The genera *Natrinema* and *Haloterrigena* are members of the family *Natrialbaceae*. The genus *Natrinema* was described in October 1998 (McGenity et al., 1998), just 3 months earlier than the genus *Haloterrigena* (Ventosa et al., 1999). For that reason, the latter article did not include the recently described strains of *Natrinema* for comparative purposes since the manuscript was submitted for peer-review before the acceptance of the former. Therefore, Ventosa et al. (1999), honestly according to their results, proposed the creation of the new genus *Haloterrigena* with the new species *Htg. turkmenica*, instead of a novel species within the genus *Natrinema*, which would have been more advisable. Since then, several new species affiliated to both genera have been described and, nowadays, the genus *Natrinema* comprises eight validly published species names (Minegishi and Kamekura, 2019b) while *Haloterrigena* harbors 11 species (Chen et al., 2019; Minegishi and Kamekura, 2019a). In addition, other non-validated species names have been proposed, specifically, "*Natrinema ajinwuensis*" (Mahansaria et al., 2018) and "*Natrinema thermophila*" (Kim et al., 2018), as well as isolates not-yet assigned to any existent species (*Natrinema* sp. J7-1, *Natrinema* sp. J7-2, *Haloterrigena* sp. GSL-11, and *Haloterrigena* sp. SGH1) (Post and Al-Harjan, 1988; Zhang et al., 2012; Flores et al., 2020).

Several studies have pointed out the taxonomic problems arising in the genera *Natrinema* and *Haloterrigena* from the fact that molecular markers (i.e., 16S rRNA, *atpB*, *EF-2*, *radA*, *rpoB'*, and *secY* gene sequences) and DNA–DNA hybridization data suggest an overlapping among members of both genera (Oren and Ventosa, 2002; Tindall, 2003; Wright, 2006; Enache et al., 2007; Minegishi et al., 2010; Papke et al., 2011). However, a detailed phylogenomic and comparative genomic study based on whole genome sequences has not been accomplished yet, nor was a formal proposal made to unravel the controversy between the cluster *Natrinema*/*Haloterrigena*. Moreover, the taxonomic status of the closely related genus *Natronorubrum* deserves special attention because 16S rRNA gene phylogenetic reconstructions suggest that the species *Natronorubrum sediminis* might belong to the *Natrinema*/*Haloterrigena* group, as the closest relative to *Haloterrigena daqingensis* (Ruiz-Romero et al., 2013). Since *Natronorubrum sediminis* (Gutiérrez et al., 2010) and *Haloterrigena daqingensis* (Wang et al., 2010) were proposed at almost the same time (only a 2-month gap), their close relationship was not noticed at that time. Additionally, the species *Halopiger salifodinae* seems to be properly affiliated to the genus *Halopiger* according to the 16S rRNA gene-based phylogeny, but complete *rpoB'* gene sequence analysis (which has been demonstrated to be a more advantageous phylogenetic marker than the 16S rRNA gene in the class *Halobacteria*) (Minegishi et al., 2010), indicated its closest relationship with the *Natrinema*/*Haloterrigena* cluster (Minegishi et al., 2016).

In the post-genomic era, it is possible to take advantage of big genome databases and low-cost sequencing to infer phylogenetic relationships among prokaryotes using the core orthologous genes detected in the genomes under study in order to accurately elucidate their evolutionary history (de la Haba et al., 2019). Besides, comparative genomics and Overall Genome Related Indexes (OGRI) have been proposed as approaches to inspect the evolutionary distance among species and to delineate prokaryotic taxa at family, genus and species level (Borriss et al., 2011; Chun and Rainey, 2014; Konstantinidis et al., 2017; Ramírez-Durán et al., 2021) and current taxonomy should benefit from them.

Aimed to resolve the taxonomic issues in the cluster *Natrinema*/*Haloterrigena* and related taxa within the family *Natrialbaceae*, we conducted phylogenomic and comparative genomic analyses using available dataset from public databanks. Additionally, we also obtained the whole genome sequence of a relevant type strain of this family which was missing in data banks. Several taxonomic changes are formally proposed in view of our results.

MATERIALS AND METHODS

Genome Retrieval and Sequencing

All genome sequences from type strains of species of the family *Natrialbaceae* available until May 31st, 2020 in NCBI GenBank database were retrieved. Other additional genomes from reference (non-type) strains of *Natrinema*/*Haloterrigena* genera were also recovered (**Table 1**). Whole genome sequences were annotated following the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Haft et al., 2018) to predict protein-coding genes as well as other functional genome units, such as structural RNAs and tRNAs.

The genome sequence of the type strain of *Haloterrigena longa* was not available in any searched public database (NCBI GenBank, JGI Genome Portal, Global Catalog of Type Strain). Since that sequence data was quite relevant for the present work, we obtained the type material from the Japanese Collection of Microorganisms for the aforementioned strain (JCM 13563) and further processed it in order to obtain its whole genome sequence. High-quality genomic DNA was extracted using the QIAmp DNA Mini Kit (Qiagen) following the manufacturer's instructions. Library preparation was performed using a combination of paired-end and mate pair strategies to generate short-insert and long-insert paired-end DNA libraries, respectively. DNA fragments were sequenced on an Illumina MiSeq platform to obtain 2 × 301-bp short-insert paired-end

TABLE 1 | Main features of genome sequences of strains of the family *Natrialbaceae* used in this study.

Strain	Accession no.	Assembly	Level	Size (Mb)	GC%	Scaffolds	Contigs	CDS	N50	L50
Halobiforma haloterrestris DSM 13078^T	FOKW00000000.1	GCA_900112205.1	Scaffold	4.50	65.4	31	32	4273	375,716	4
Halobiforma lacisalsi AJ5^T	CP019285.1	GCA_000226975.3	Complete	4.38	65.2	3	3	4177	4,161,587	1
Halobiforma nitratireducens JCM 10879^T	AOMA00000000.1	GCA_000337895.1	Contig	3.69	63.7	205	205	3552	47,406	25
Halopiger aswanensis DSM 13151^T	RAPO00000000.1	GCA_003610195.1	Scaffold	4.87	64.4	17	18	4589	1,426,401	2
Halopiger djelfimassiliensis IIH2^T	CBMA00000000.1	GCA_000455365.1	Scaffold	3.78	64.2	6	55	3671	1,082,527	2
Halopiger goleimassiliensis IIH3^T	CBMB00000000.1	GCA_000455345.1	Scaffold	3.91	66.1	3	11	3756	3,025,424	1
Halopiger salifodinae CGMCC 1.12284^T	FOIS00000000.1	GCA_900110455.1	Scaffold	4.27	65.4	8	9	4010	878,349	3
Halopiger xanaduensis SH-6^T	NC_015666.1	GCA_000217715.1	Complete	4.36	65.2	4	4	4178	3,668,009	1
Halostagnicola kamekurae DSM 22427^T	FOZS00000000.1	GCA_900116205.1	Contig	4.11	61.5	16	16	4042	1,202,185	2
Halostagnicola larsenii XH-48^T	CP007055.1	GCA_000517625.1	Complete	4.13	60.9	5	5	3966	2,789,326	1
Haloterrigena daqingensis CGMCC 1.8909^T	FTNP00000000.1	GCA_900156445.1	Contig	3.83	61.4	14	14	3687	859,600	2
Haloterrigena daqingensis JX313^T	CP019327.1	GCA_001971705.1	Complete	3.84	61.3	4	4	3692	3,397,437	1
Haloterrigena hispanica CDM_1	FMZP00000000.1	GCA_900101245.1	Scaffold	3.91	61.0	135	139	3983	148,801	9
Haloterrigena hispanica CDM_6	FOIC00000000.1	GCA_900111485.1	Scaffold	3.96	61.0	92	100	3989	128,565	9
Haloterrigena hispanica DSM 18328^T	SHMP00000000.1	GCA_004217335.1	Contig	4.26	60.7	11	11	4121	1,073,359	2
Haloterrigena jeotgali A29^T	CP031303.1 (chromosome), CP031298.1, CP031299.1, CP031300.1, CP031301.1, CP031302.1, CP031304.1 (plasmids)	GCA_004799625.1	Complete	4.90	65.0	7	7	4967	3,644,881	1
Haloterrigena limicola JCM 13563^T	AOIT00000000.1	GCA_000337475.1	Contig	3.52	61.8	94	94	3512	116,493	9
Haloterrigena longa JCM 13563^T	JAHUQE000000000.1	GCA_020105915.1	Scaffold	4.13	63.8	6	15	4069	3,590,587	1
Haloterrigena mahii H13^T	JHUT00000000.2	GCA_000690595.2	Scaffold	3.79	65.1	24	29	3707	248,588	4
Haloterrigena saccharevitans AB14^T	LWLN00000000.1	GCA_001953745.1	Contig	3.98	65.3	3	3	3921	3,473,758	1
Haloterrigena salifodinae ZY19^T	RQWN00000000.1	GCA_003977755.1	Scaffold	4.96	64.5	11	14	4761	1,204,032	2
Haloterrigena salina JCM 13891^T	AOIS00000000.1	GCA_000337495.1	Contig	4.84	65.2	71	71	4540	151,334	11
Haloterrigena sp. H1	SMZK00000000.1	GCA_005938085.1	Contig	4.26	61.5	9	9	4253	3,035,199	1
Haloterrigena thermotolerans DSM 11552^T	AOIR00000000.1	GCA_000337115.1	Contig	3.90	65.4	68	68	3862	162,183	9
Haloterrigena turkmenica DSM 5511^T	NC_013743.1	GCA_000025325.1	Complete	5.44	64.2	7	7	5167	3,889,038	1
Haloterrigena turkmenica WANU15	LKCV00000000.1	GCA_001483125.1	Contig	2.95	64.0	574	574	3202	15,902	50
Halovivax asiaticus JCM 14624^T	AOIQ00000000.1	GCA_000337515.1	Contig	3.24	64.5	24	24	3115	327,817	4
Halovivax ruber XH-70^T	NC_019964.1	GCA_000328525.1	Complete	3.23	64.3	1	1	3099	3,223,876	1
Natrarchaeobaculum aegyptiacum JW/NM-HA 15^T	CP019893.1	GCA_002156705.1	Complete	3.93	64.1	1	1	3745	3,930,546	1
Natrarchaeobaculum sulfurireducens AArc1^T	CP024047.1	GCA_003430825.1	Complete	3.79	62.4	3	3	3576	3,521,804	1
Natrarchaeobius chitinivorans AArcht4^T	REGA00000000.1	GCA_003841505.1	Contig	4.57	61.9	48	48	4382	170,161	8
Natrarchaeobius halalkaliphilus AArcht-SI^T	REFY00000000.1	GCA_003841485.1	Contig	3.51	61.1	12	12	3409	639,802	3

(Continued)

TABLE 1 | (Continued)

Strain	Accession no.	Assembly	Level	Size (Mb)	GC%	Scaffolds	Contigs	CDS	N50	L50
Natrialba aegyptia DSM 13077^T	AOIP00000000.1	GCA_000337535.1	Contig	4.62	62.0	66	66	4429	145,225	7
Natrialba asiatica DSM 12278^T	AOIO00000000.1	GCA_000337555.1	Contig	4.40	62.4	49	49	4188	174,934	7
Natrialba chahannaoensis JCM 10990^T	AOIN00000000.1	GCA_000337135.1	Contig	4.31	60.4	106	106	4030	129,612	13
Natrialba hulunbeirensis JCM 10989^T	AOIM00000000.1	GCA_000337575.1	Contig	4.16	61.7	48	48	3834	159,578	8
Natrialba magadii ATCC 43099^T	NC_013922.1	GCA_000025625.1	Complete	4.44	61.0	4	4	4154	3,751,858	1
Natrialba swarupiae ESP3B_9^T	VTAW00000000.1	GCA_008245225.1	Contig	4.20	62.5	99	99	3969	129,190	11
Natrialba taiwanensis DSM 12281^T	AOIL00000000.1	GCA_000337595.1	Contig	4.64	61.5	70	70	4399	199,614	9
Natrinema altunense AJ2^T	JNCS00000000.1	GCA_000731985.1	Contig	3.77	64.6	20	20	3688	425,349	4
Natrinema altunense 1A4-DGR	JXAN00000000.1	GCA_000815265.1	Contig	3.72	64.8	215	215	5159	33,862	34
Natrinema altunense 4.1R	SHMR00000000.1	GCA_004209855.1	Scaffold	3.67	64.9	12	81	3631	1,929,556	1
Natrinema altunense JCM 12890^T	AOIK00000000.1	GCA_000337155.1	Contig	3.77	64.5	52	52	3698	184,807	7
Natrinema ejinorense JCM 13890^T	NXNI00000000.1	GCA_002494345.1	Contig	4.48	63.9	3	3	4337	3,988,345	1
Natrinema gari JCM 14663^T	AOIJ00000000.1	GCA_000337175.1	Contig	4.02	63.7	88	88	3997	126,340	11
Natrinema pallidum BOL6-1	CP040637.1	GCA_005890195.1	Complete	3.78	64.3	3	3	3723	3,503,953	1
Natrinema pallidum DSM 3751^T	AOII00000000.1	GCA_000337615.1	Contig	3.92	63.7	115	115	3852	88,603	17
Natrinema pellirubrum DSM 15624^T	NC_019962.1 (chromosome), NC_019963.1, NC_019967.1 (plasmids)	GCA_000230735.3	Complete	4.35	64.0	3	3	4249	3,790,479	1
Natrinema salaciae DSM 25055^T	FOFD00000000.1	GCA_900110865.1	Scaffold	4.86	65.0	11	15	4634	865,606	3
"Natrinema thermophila" CBA1119	PDBS00000000.1	GCA_002572525.1	Contig	5.06	62.3	9	9	4965	4,087,412	1
Natrinema sp. J7-1	AJVG00000000.1	GCA_000493245.1	Contig	3.67	64.4	42	42	3632	196,646	6
Natrinema sp. J7-2	NC_018224.1	GCA_000281695.1	Complete	3.79	64.1	2	2	3706	3,697,626	1
Natrinema versiforme BOL5-4	CP040330.1	GCA_005576615.1	Complete	4.67	63.4	5	5	4514	3,747,116	1
Natrinema versiforme JCM 10478^T	AOID00000000.1	GCA_000337195.1	Contig	4.19	64.0	72	72	4146	121,463	13
Natronobacterium gregoryi SP2^T	NC_019792.1	GCA_000230715.3	Complete	3.79	62.2	1	1	3710	3,788,356	1
Natronobacterium texcoconense DSM 24767^T	FNLC00000000.1	GCA_900104065.1	Scaffold	4.01	62.9	9	10	3976	1,245,734	2
Natronococcus amylolyticus DSM 10524^T	AOIB00000000.1	GCA_000337675.1	Contig	4.42	64.4	44	44	4320	232,276	7
Natronococcus jeotgali DSM 18795^T	AOIA00000000.1	GCA_000337695.1	Contig	4.50	64.4	170	170	4458	76,066	20
Natronococcus occultus SP4^T	NC_019974.1	GCA_000328685.1	Complete	4.31	64.6	3	3	4174	4,013,216	1
Natronolimnobius baerhuensis CGMCC 1.3597^T	MWPH00000000.1	GCA_002177135.1	Contig	3.91	60.2	8	8	3745	1,261,254	2
Natronolimnohabitans innermongolicus JCM 12255^T	AOHZ00000000.1	GCA_000337215.1	Contig	4.59	64.3	121	121	4384	96,333	18
Natronorubrum aibiense 7-3^T	CP045488.1	GCA_009392895.1	Complete	4.35	61.3	4	4	4130	3,352,994	1
Natronorubrum bangense JCM 10635^T	AOHY00000000.1	GCA_000337715.1	Contig	4.11	60.4	62	62	3982	138,654	10
Natronorubrum sediminis CGMCC 1.8981^T	FNWL00000000.1	GCA_900108095.1	Scaffold	3.78	61.1	6	9	3583	1,300,740	2
Natronorubrum sulfidifaciens JCM 14089^T	AOHX00000000.1	GCA_000337735.1	Contig	3.46	61.8	63	63	3408	225,522	5
Natronorubrum texcoconense B4^T	FNFE00000000.1	GCA_900100335.1	Contig	4.64	63.6	11	11	4423	457,630	3
"Natronorubrum thiooxidans" HArc	FTNR00000000.1	GCA_900156475.1	Scaffold	4.21	60.9	62	63	4067	250,216	6
Natronorubrum tibetense GA33^T	ARPH00000000.1	GCA_000383975.1	Scaffold	4.93	62.3	5	10	4649	4,057,512	1

reads (SIPERs) and 2 × 301-bp long-insert paired-end reads (LIPERs). Downstream analyses were carried out as previously described (Ramírez-Durán et al., 2020). In brief, sequencing reads were quality filtered and trimmed using BBTools v.38.44 (Bushnell, 2020) and then assembled with SPAdes v.3.13.0 (Bankevich et al., 2012) using combined SIPERs and LIPERs as input. Automatic annotation of the draft genome was achieved using PGAP (Haft et al., 2018) as indicated above, that includes prediction of protein-coding genes, as well as other functional genome units such as structural RNAs, tRNAs, small RNAs and pseudogenes using a combination of *ab initio* gene prediction algorithms with homology based methods.

Phylogenetic and Phylogenomic Treeing

The 16S rRNA gene sequences from the type strains of species of the family *Natrialbaceae* were downloaded from GenBank/EMBL/DDBJ databases or extracted from the whole genome sequences and then, aligned and used to calculate similarity matrixes and to construct neighbor-joining (NJ) (Saitou and Nei, 1987), maximum-parsimony (MP) (Fitch, 1971), and maximum-likelihood (MP) (Felsenstein, 1981) phylogenetic trees in ARB v.6.0.5 software package (Westram et al., 2011). Jukes-Cantor model of DNA evolution (Jukes and Cantor, 1969) was selected to correct the distance matrix. General Time Reversible model (Tavaré, 1986) with gamma-distribution and proportion of invariant sites to estimate rate heterogeneity over sites (GTR + Γ + I) was used to infer ML phylogeny. Branch support was assessed by 1,000 bootstrap pseudo-replicates (Felsenstein, 1985).

Since 16S rRNA gene-based phylogenies have been demonstrated not to be reliable to determine in-depth evolutionary relationships within the class *Halobacteria* and their results must be regarded with caution and carefully checked (Papke, 2009; Corral et al., 2018; de la Haba et al., 2018; Infante-Domínguez et al., 2020), a more robust and accurate phylogenomic approach was attempted. Firstly, pan- and core-genome datasets were determined using an all-vs.-all Blastp comparison among the translated CDS features of the annotated genomes under study, as previously described (de la Haba et al., 2019). Then, translated single-copy core gene sequences were individually aligned with Muscle (Edgar, 2004) and concatenated into a super-protein alignment, which was further analyzed to generate the phylogenomic tree by means of the approximately maximum-likelihood algorithm implemented in FastTreeMP v.2.1.8 (Price et al., 2010). Jones-Taylor-Thornton model of amino acid evolution (Jones et al., 1992) with a single rate for each site (JTT + CAT) was applied for phylogenomic reconstruction. Tree branch support was inferred using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999).

Both, 16S rRNA gene-based and phylogenomic trees, were managed, displayed and annotated using the online tool iTOL v.5.7 (Letunic and Bork, 2021).

Comparative Genomic Analyses

Overall Genome Relatedness Indexes (OGRI) were calculated for all-vs.-all genome pairs. Specifically, Orthologous Average Nucleotide Identity (OrthoANI) was determined using the OrthoANIu Tool (Yoon et al., 2017) which depends on USEARCH v8.1.1861, the digital DNA-DNA hybridization (dDDH) was inferred by means of the Genome-to-Genome Distance Calculator (GGDC) (formula 2) (Meier-Kolthoff et al., 2013), the Average Amino-acid Identity (AAI) was estimated using the aai.rb script from the Enveomics collection (Rodriguez-R and Konstantinidis, 2016) and, finally, the Percentage Of Conserved Proteins (POCP) was calculated with a homemade Perl script as described elsewhere (Qin et al., 2014).

Synteny analysis among selected representative genomes within the family *Natrialbales* was carried out to detect conservation of homologous genes and gene order across closed relatives. Because synteny can be affected by sequence fragmentation (Liu et al., 2018), draft genome contigs were reordered prior to infer synteny blocks using a gold standard genome (i.e., complete genome sequence) of a closely related species as a reference, using the Mauve Contig Mover functionality (Rissman et al., 2009). Conserved blocks were identified after Blastn pairwise comparisons (e-value $\leq 10^{-3}$) between the rearranged genomes and synteny plots were visualized using Easyfig v.2.2.3 (Sullivan et al., 2011).

RESULTS AND DISCUSSION

The 16S rRNA Gene Sequence Analysis Unveils the Taxonomic Problems Arising Within the Genera *Haloterrigena* and *Natrinema*

To gain a general overview of the current taxonomic situation of the family *Natrialbaceae* we reconstructed a phylogeny based on the 16S rRNA gene sequences (the most widely used molecular marker in modern prokaryotic systematics) including all type strains of the species with validly published names within that family (**Figure 1**). As hinted at previous studies (Tindall, 2003; Wright, 2006; Gupta et al., 2016), our results confirm that neither the genus *Natrinema* nor the genus *Haloterrigena* constituted monophyletic groups, but the constituent species of both genera were intermingled into a single monophyletic cluster, with the exception of the species *Haloterrigena daqingensis* which clustered together to *Natronorubrum sediminis* and *Natronococcus roseus*, distantly related to the rest of the species of *Natrinema*/*Haloterrigena*. Other problematic (polyphyletic or paraphyletic) genera within this family were *Halovivax*, *Natrialba*, *Natronococcus*, and *Natronorubrum* (**Figure 1**).

The 16S rRNA gene sequence similarities among the type species within the genera *Natrinema* and *Haloterrigena*, independently considered, ranged between 99.5–95.3% and 99.0–94.4%, respectively, while the sequence similarities between both genera varied from 99.0 to 94.6%, by far above the threshold value for differentiating prokaryotic genera (<94.5%) (Yarza et al., 2014). Therefore, intra- and inter-genera sequence similarities overlap almost entirely, which indicates a rather fuzzy delineation between those two genera. With regards to species delineation, the following monophyletic groups sharing equal or more than 98.65%

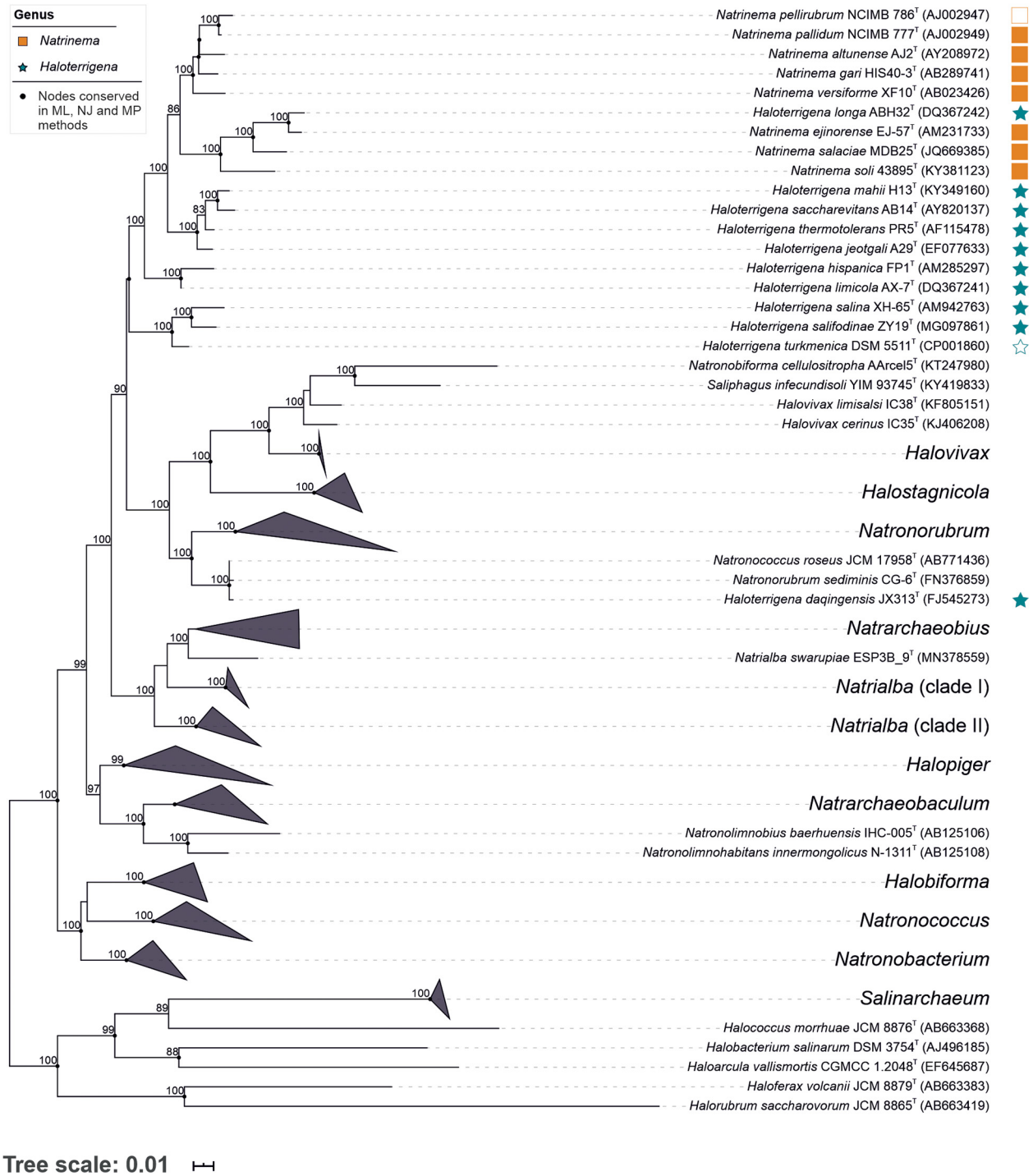


FIGURE 1 | Maximum-likelihood phylogenetic tree based on the 16S rRNA gene sequence comparison of members of the genera *Natrinema* and *Haloterrigena* and representatives of the most closely related genera of the family *Natrialbaceae*. Bootstrap values ≥ 70% (based on 1,000 *pseudo*-replicates) are shown above the branches. Bar, 0.01 changes per nucleotide position. Empty square and star indicate the type species of the corresponding genus.

sequence similarity (generally accepted as the prokaryotic species cutoff value) (Kim et al., 2014) could be observed: *Natrinema pellirubrum—Natrinema pallidum*; *Natrinema ejinorense—Haloterrigena longa*; *Haloterrigena mahii—Haloterrigena saccharevitans—Haloterrigena thermotolerans*; *Haloterrigena hispanica—Haloterrigena limicola*; *Haloterrigena daqingensis—Natronococcus roseus—Natronorubrum sediminis*. Besides, other potential species synonymy could be detected in the family

Natrialbaceae: *Halobiforma haloterrestris—Halobiforma lacisalsi*; *Halopiger aswanensis—Halopiger thermotolerans—Halopiger xanaduensis*; *Halostagnicola alkaliphila—Halostagnicola bangensis*; *Halovivax asiaticus—Halovivax ruber*; and *Natrialba aegyptia—Natrialba taiwanensis—Natrialba asiatica*. Despite that different species could sometimes share values above the indicated threshold, the groups mentioned here should be carefully checked to detect the existence of synonymy.

Taxophylogenomics and Overall Genome Related Indexes Values Prove the Proposal to Keep the Genera *Natrinema* and *Haloterrigena* as Separated Taxa

To confirm the results noted after 16S rRNA gene sequence analysis, a more robust and determinative phylogenomic analysis was carried out. For that purpose, all genome sequences from type strains of the species of the family *Natrialbaceae* as well as other non-type strains of the genera *Natrinema* and *Haloterrigena* available in NCBI GenBank database at the time of the study were recovered. Since the genome data for the type strain of *Haloterrigena longa* could not be retrieved and because this species requested special attention given its close relationship to *Natrinema ejinorense* (as indicated above), we sequenced and analyzed it. A total of ∼0.32 and ∼2.12 Gb from paired-end and mate pair libraries, respectively, were obtained after trimming and filtering. Average insert size was computationally estimated to be ∼550 bp for paired-end and ∼2,000 bp for mate pair datasets. Assembly yielded a 4.13 Mb, 6 scaffolds genome with a N50 of 3,590,587 bp and a coverage of 78X. **Table 1** shows all the genome sequences used in this study, as well as their main features.

Phylogenomic trees inferred from the concatenation of the 525 amino acid sequences of the orthologous single-copy genes present in the type strain genomes (**Figure 2**) and in all the genomes under study (**Supplementary Figure 1**) were obtained. A previous study focused on the evolution of the class *Halobacteria* has also reported phylogenomic core trees that concurs with our results, although those phylogenetic reconstructions were based on only 45 orthologous core genes and did not include all the representative genomes within the *Natrialbales* (Gaba et al., 2020). As might be expected, the topology of our phylogenomic tree was not totally in agreement to the 16S rRNA gene phylogeny, but the clusters obtained were better supported in the phylogenomic tree, with 100% bootstrap in almost all bifurcations. Most significantly, the strains from the cluster *Natrinema*/*Haloterrigena* (which could not be distinguished either from each other, as in the 16S rRNA tree) did not form a monophyletic group, even when excluding the species *Haloterrigena daqingensis*. In particular, *Haloterrigena turkmenica* (the type species of the genus), *Haloterrigena salifodinae*, and *Haloterrigena salina* clustered together and separated from the other *Natrinema*/*Haloterrigena* members. Therefore, our results indicate that merging both genera is not convenient, while transferring some current *Haloterrigena* species (i.e., *Htg. hispanica*, *Htg. jeotgali*, *Htg. limicola*, *Htg. longa*, *Htg. mahii*, *Htg. saccharevitans*, *Htg. thermotolerans*) to the genus *Natrimena* seems more appropriate. That way the genus *Haloterrigena* would remain composed of the species *Htg. turkmenica*, *Htg. salina*, and *Htg. salifodinae*. Furthermore, the species *Haloterrigena daqingensis* formed a monophyletic group with all the *Natronorubrum* species, thus suggesting its reclassification as a member of the latter genus. It is worth noting that the species *Halopiger salifodinae* did not affiliate with the other *Halopiger* species but was closely related to the *Natrinema*/*Haloterrigena* group. This taxon might belong, indeed, to the latter group or, alternatively, it might constitute a new separate genus within the *Natrialbaceae*. In order to unravel this issue, an in-depth analysis of OGRI values may be determinative.

The reference non-type strains whose genome sequences were included in this study showed that all of them clustered together to their respective type strain, except for the strain *Haloterrigena turkmenica* WANU15, which might be part of the genus *Natronolimnohabitans*; however, it must be noted that the genome sequence of *Haloterrigena turkmenica* WANU15 has been confirmed to be contaminated (Lee et al., 2017) and, thus, this result must be observed with caution. Concerning the unnamed strains analyzed (i.e., *Natrinema* sp. J-1, *Natrinema* sp. J-2, and *Haloterrigena* sp. H1), the two first probably belong to the species *Natrinema gari*, whereas the latter might be regarded as a new species into the *Natrinema*/*Haloterrigena* archaeal set. Phylogenomic tree also uncover some other taxonomic problems arising within the *Natrialbaceae*, such as the polyphyly of the genera *Natrialba* and *Halopiger*, but they are out of the scope of this study.

Aimed to shed light on the classification of the *Natrialbaceae*, several OGRI types were calculated, in particular those mostly accepted to delineate taxa at the prokaryotic genus and species level. Methods to demarcate genera have been proposed that are based on either AAI (Konstantinidis and Tiedje, 2007) or the POCP (Qin et al., 2014). The former approach sets a cutoff value for genus demarcation of 65% AAI (Konstantinidis et al., 2017); however, this threshold cannot be universally employed for all bacterial and archaeal lineages. In fact, if we would use the 65% AAI cutoff all the genera within the *Natrialbaceae*, apart from the genus *Halovivax*, should be merged in a single one since they shared AAI values equal or above 67% (**Figure 3** and **Supplementary Figure 2**). Therefore, AAI values might be useful for genera demarcation in this family, but a different boundary needs to be established for it. Previous studies have pointed out the convenience to set lineage specific OGRI limits to define prokaryotic genera (Barco et al., 2020). Genus demarcation boundaries were determined for the family *Natrialbaceae* after detailed inspection of AAI values for all pairwise genome comparisons (**Figure 3** and **Supplementary Figure 2**), in agreement with the phylogenomic trees (**Figure 2** and **Supplementary Figure 1**), to avoid the existence of polyphyletic genera. Thus, we propose a cutoff value of $\leq$ 76% AAI to differentiate genera within the family *Natrialbaceae*, a robust and consistent threshold according to the observed evolutionary relationships among members of this family. By using this threshold, the species *Haloterrigena turkmenica*,

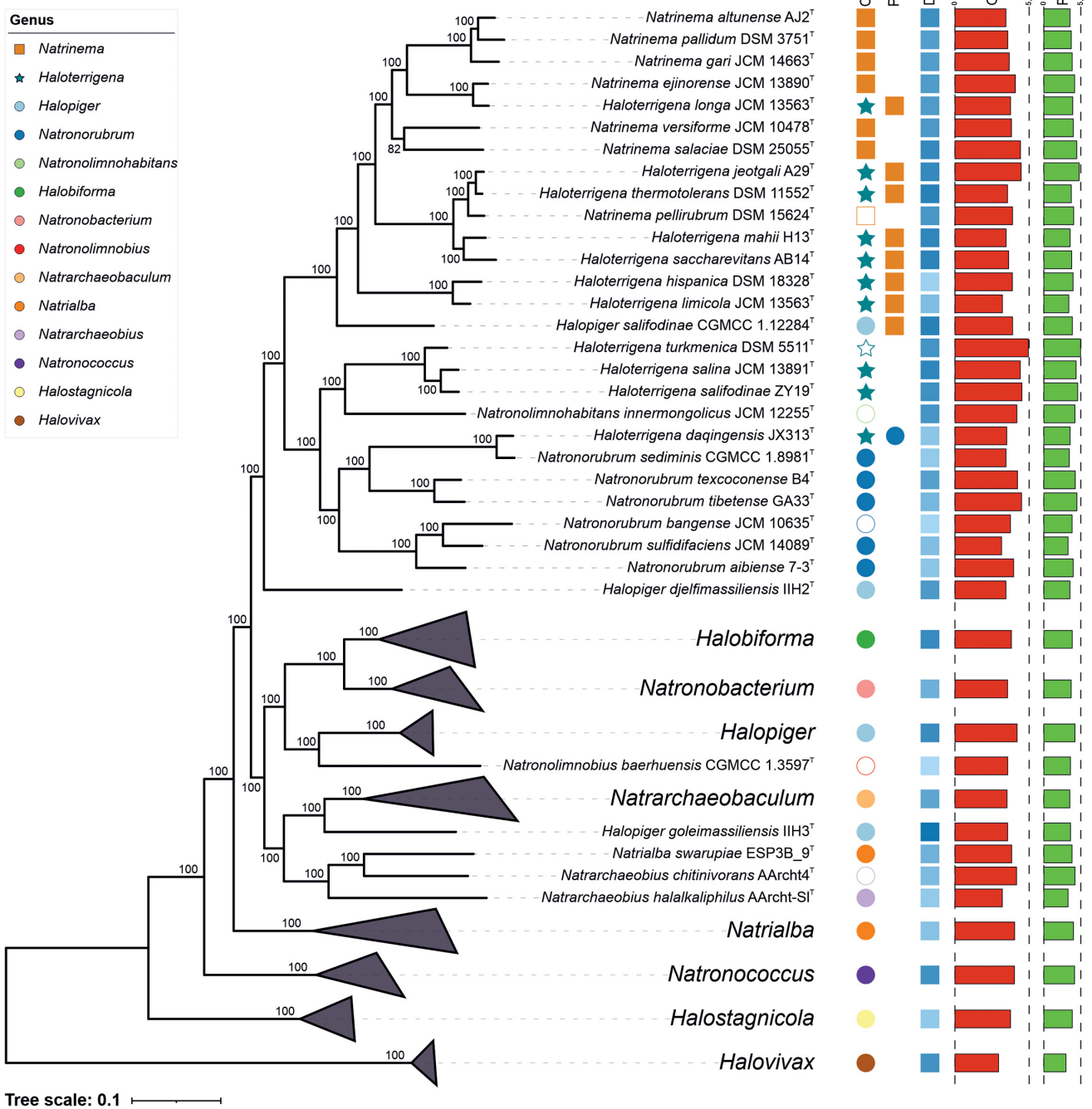


FIGURE 2 | Approximate maximum-likelihood phylogenomic tree based on the concatenation of the translated sequence of the 525 single-copy genes shared by the type strains of members of the genera *Natrinema* and *Haloterrigena* and related taxa of the family *Natrialbaceae* under study. Bootstrap values ≥ 70% (based on Shimodaira-Hasegawa-like local support) are shown above the branches. Bar, 0.1 changes per nucleotide position. Empty symbols indicate the type species of the corresponding genus.

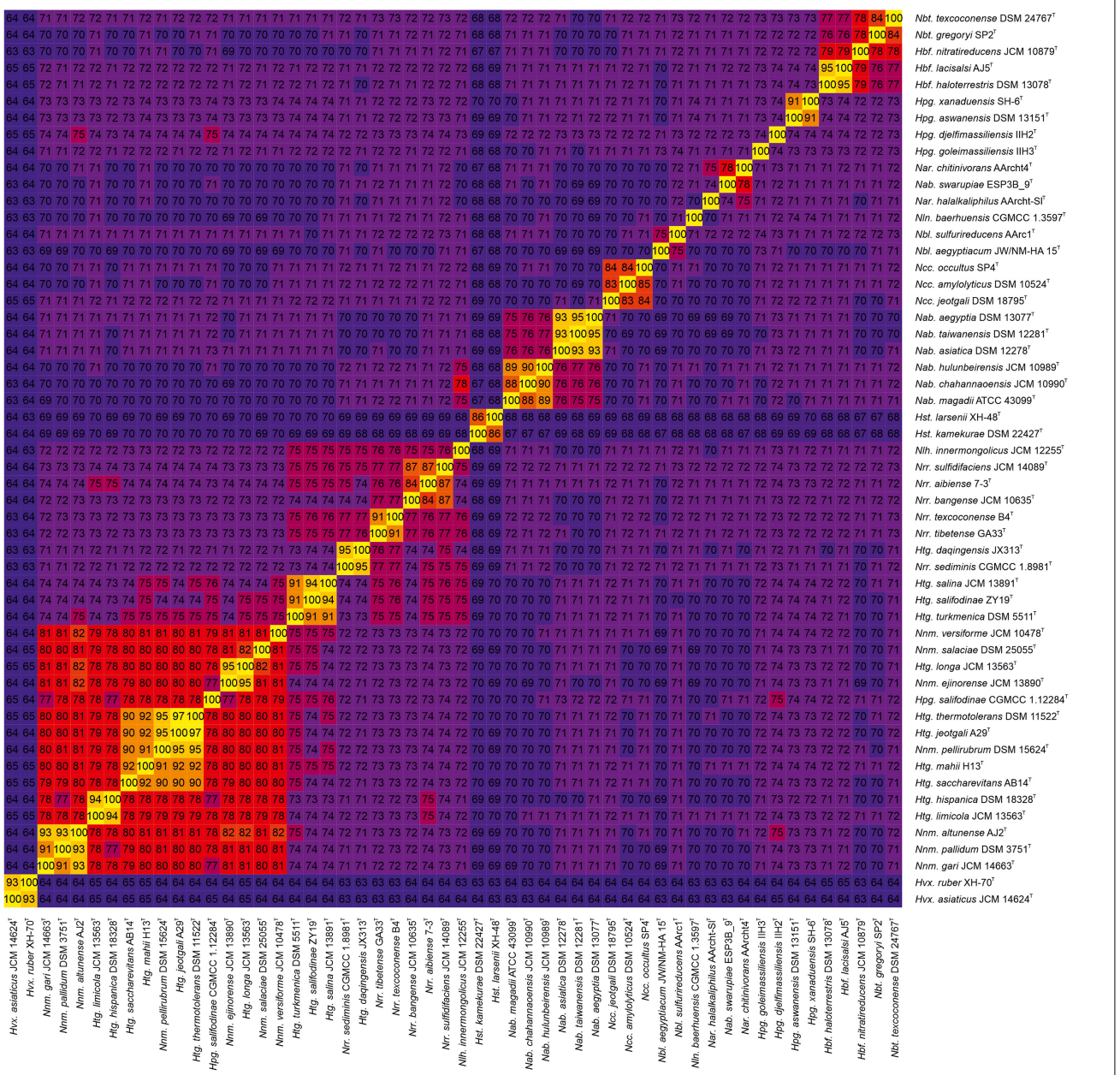


FIGURE 3 | Heatmap of AAI relatedness among the type strains of members of the genera *Natrinema* and *Haloterrigen*a and representatives of the other genera of the family *Natrialbaceae*.

Haloterrigena salifodinae, and *Haloterrigena salina* will be retained as the only members of *Haloterrigena*. Moreover, the species *Haloterrigena daqingensis* should be transferred to the genus *Natronorubrum*. Finally, the remaining species of *Haloterrigena*, the species *Halopiger salifodinae* and all the *Natrinema* species should be joined together into a single genus. Since the genus *Natrinema* has priority over the other two, all the aforementioned species should be reclassified as members of *Natrinema*. Our proposed genus limit should also have consequences in the taxonomic status of other genera of the family *Natrialbaceae*, such as the convenience to merge the genera *Halobiforma* and *Natronobacterium*, the transfer of *Natrialba swarupia* into the genus *Natrarchaeobius* and the need to revisit the affiliation of *Halopiger goleimassiliensis* and *Halopiger djelfimassiliensis* outside the genus *Halopiger*. Nevertheless, additional studies including all the type strains of the species of those genera is required, which is beyond the subject of the present article.

On the other hand, the POCP method sets a genus boundary at a value of 50% (Qin et al., 2014). Nevertheless, that limit cannot be applied to the family *Natrialbaceae* since all the constituent genera shared values above it. It has been discussed that this cutoff value was arbitrarily established (Barco et al., 2020), so, according to our results (**Supplementary Figure 3**) we can propose a threshold at a POCP value of < 66% for genus demarcation in this family. Nevertheless, this genomic index seems not to be as accurate as AAI and in borderline cases interpretation of results may be unclear. For example, the group formed by *Haloterrigena turkmenica*, *Haloterrigena salifodinae*, and *Haloterrigena salina* (which seemed to constitute an independent genus as explained above) could not be clearly separated from the *Haloterrigena daqingensis*/*Natronorubrum* spp. cluster, from the genus *Natronolimnohabitans*, or from the rest of the strains of the *Natrinema*/*Haloterrigena* clade using our proposed POCP-based genus cutoff. Another outlier was the low POCP values of the strain *Natrinema altunense* 1A4-DGR with respect to most of the strains within the *Natrinema*/*Haloterrigena* cluster, indicating some confidence issues for this index. Besides, other genera within the family *Natrialbaceae* that could not be distinguished using POCP index but whose unification is not supported by phylogenomic tree were *Natrarchaeobaculum—Natrarchaeobius—Natronolimnobius*; *Halobiforma—Halopiger*. Hence, we discourage taxonomist from using POCP method to define genera within the family *Natrialbaceae*.

A longer list of OGRI has been proposed to be useful for prokaryotic species delineation (Palmer et al., 2020), such as AAI (Konstantinidis and Tiedje, 2005) –which can also be employed for genus demarcation–, ANIb (Goris et al., 2007), ANIm, TETRA (Richter and Rossello-Mora, 2009), MUMi (Deloger et al., 2009), dDDH (Meier-Kolthoff et al., 2013), gANI, alignment fraction (Varghese et al., 2015), OrthoANI (Lee et al., 2016), and FastANI (Jain et al., 2018). Among them, two of the most widely used for taxonomic purposes at species level are dDDH and OrthoANI, with widely accepted cutoff values of 70% (Auch et al., 2010) and 95–96% (Goris et al., 2007; Richter and Rossello-Mora, 2009; Chun and Rainey, 2014), respectively. We calculated these two OGRI for the family *Natrialbaceae* (**Figure 4** and **Supplementary Figure 4**) with the aim to identify the existence of synonymy between recognized species names and to properly affiliate unnamed strains to a species. A first glimpse of OrthoANI/dDDH results showed several borderline genome pairs (94% OrthoANI and ∼55% dDDH) in our dataset, in particular *Haloterrigena hispanica* DSM 18328^T/*Haloterrigena limicola* JCM 13563^T, *Haloterrigena daqingensis* JX313^T/CGMCC 1.8909^T/*Natronorubrum sediminis* CGMCC 1.8961^T, and *Haloterrigena salifodinae* ZY19^T/*Haloterrigena salina* JCM 13891^T, but they cannot be regarded as synonyms because they are still below the species threshold values and might indicate a recent speciation event. Following this criterion, the non-type strains *Haloterrigena hispanica* CDM_1 and *Haloterrigena hispanica* CDM_6 seemed to be misclassified and they should be described as a separated species from *Haloterrigena hispanica*, although a further descriptive characterization is required for this purpose. Unfortunately, none of both strains are available in public microbial culture collections. Similarly, the strain *Haloterrigena* sp. H1, sharing ≤ 89% OrthoANI and ≤ 39% dDDH values with respect to any of the analyzed strains in the family *Natrialbaceae*, constitutes a novel species within the cluster *Natrinema*/*Haloterrigena*, but access to the biological resource is needed before to make any formal proposal. Our study also indicated that the species "*Natrinema thermophila*" (Kim et al., 2018) and "*Natronorubrum thiooxidans*" (Sorokin et al., 2005) (names effectively but not validly published) should be unequivocally considered as novel taxa within their respective genera, although those names need to be validated beforehand. More uncertain was the taxonomic differentiation of several genome pairs within the fuzzy zone (95% OrthoANI and 60–63% dDDH), specifically *Natrinema pellirubrum* DSM 15624^T/*Haloterrigena jeotgali* A29^T, *Natrinema pellirubrum* DSM 15624^T/*Haloterrigena thermotolerans* DSM 11522^T, and *Natrinema ejinorense* JCM 13890^T/*Haloterrigena longa* JCM 13563^T. Additionally, it must be noted that when using formula 1 and 3 (instead of formula 2) for dDDH calculation the results for the aforementioned genome pairs were 64–65%, 67–68%, and 70%, respectively, making more challenging their proper taxonomic classification. In those cases, the sole use of OGRI values was not discriminative enough as to make a decision on their taxonomy and additional genomic and phenotypic data must be provided. On the other hand, OrthoANI and dDDH values doubtlessly indicate that each of the following groups of strains belongs to the same species: *Natrinema gari* JCM 14663^T/*Natrinema* sp. J7-1/*Natrinema* sp. J7-2, *Natrinema pallidum* DSM 3751^T/*Natrinema pallidum* BOL6-1, *Natrinema altunense* JCM 12890^T/*Natrinema altunense* AJ2^T/*Natrinema altunense* 4.1R/*Natrinema altunense* 1A4-DGR, *Haloterrigena hispanica* CDM_1/*Haloterrigena hispanica* CDM_6, *Haloterrigena jeotgali* A29^T/*Haloterrigena thermotolerans* DSM 11522^T, *Haloterrigena daqingensis* JX313^T/*Haloterrigena daqingensis* CGMCC 1.8909^T, and *Natronolimnohabitans innermongolicus* JCM 12255^T/*Haloterrigena turkmenica* WANU15. Therefore, the species *Haloterrigena jeotgali* should be considered as a later heterotypic synonym of *Haloterrigena thermotolerans* and the strains *Natrinema* sp. J7-1, *Natrinema* sp. J7-2, and *Haloterrigena turkmenica* WANU15 should be renamed as *Natrinema gari* J7-1, *Natrinema gari* J7-2, and *Natronolimnohabitans innermongolicus* WANU15, respectively. Other putative ambiguous synonyms were detected, such as those for the species *Natrialba aegyptia*/*Natrialba taiwanensis* and *Halobiforma haloterrestris*/*Halobiforma lacisalsi*, but the convenience to be merged or not should be accomplished in future studies.

Synteny Analysis Applied to Elucidation of Uncertain Synonyms Into the *Natrinema/Haloterrigena*

The evolutionary processes that lead to diversity, chromosomal dynamics, and rearrangement rates between species can be assessed by means of the analysis of the synteny among two or more genomes, that is, the spatial distribution of locally collinear

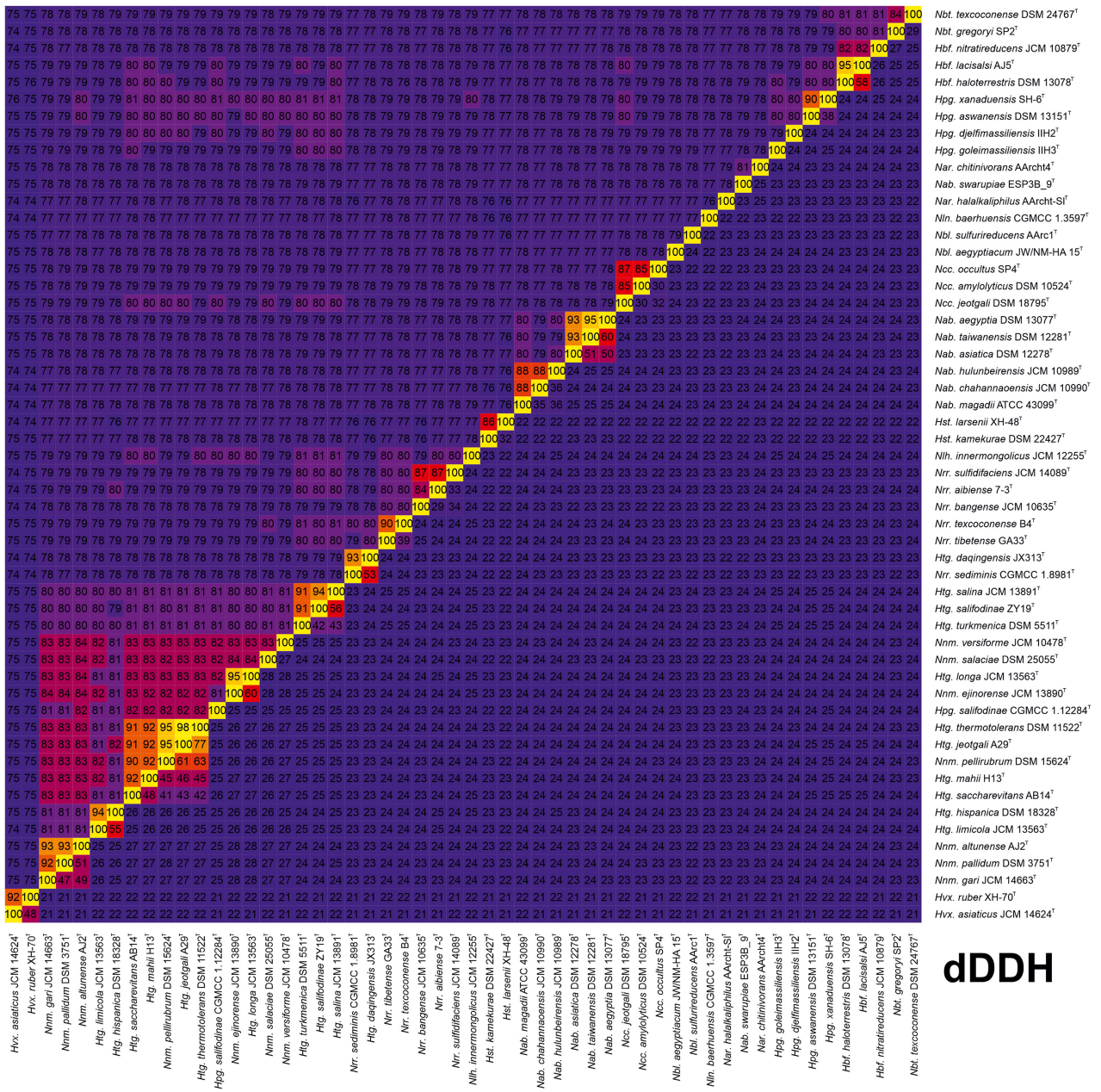


FIGURE 4 | Heatmap of OrthoANIu **(upper triangle)** and dDDH **(lower triangle)** relatedness among the type strains of members of the genera *Natrinema* and *Haloterrigena* and representatives of the other genera of the family *Natrialbaceae*.

blocks (Bhutkar et al., 2006). Thus, an approach to gain insight into the evolutionary distance between two species is to inspect the synteny of the genome sequences under study (Borriss et al., 2011; Ramírez-Durán et al., 2021). As indicated in the previous section, OGRI values equal to the species cutoffs were not able to reliably solve the taxonomic status of several species and so, the synteny analysis might shed light to elucidate the affiliation of those uncertain taxa.

Specifically, we have evaluated, on the one hand, the synteny between *Natrinema ejinorense* JCM 13890^T and *Haloterrigena longa* JCM 13563^T and, on the other hand, the synteny among *Natrinema pellirubrum* DSM 15624^T, *Haloterrigena jeotgali* A29^T, and *Haloterrigena thermotolerans* DSM 11522^T (**Figure 5**). As can be observed, although some genomic rearrangements could be evidenced, all comparisons showed high levels of conservation of locally collinear blocks. It

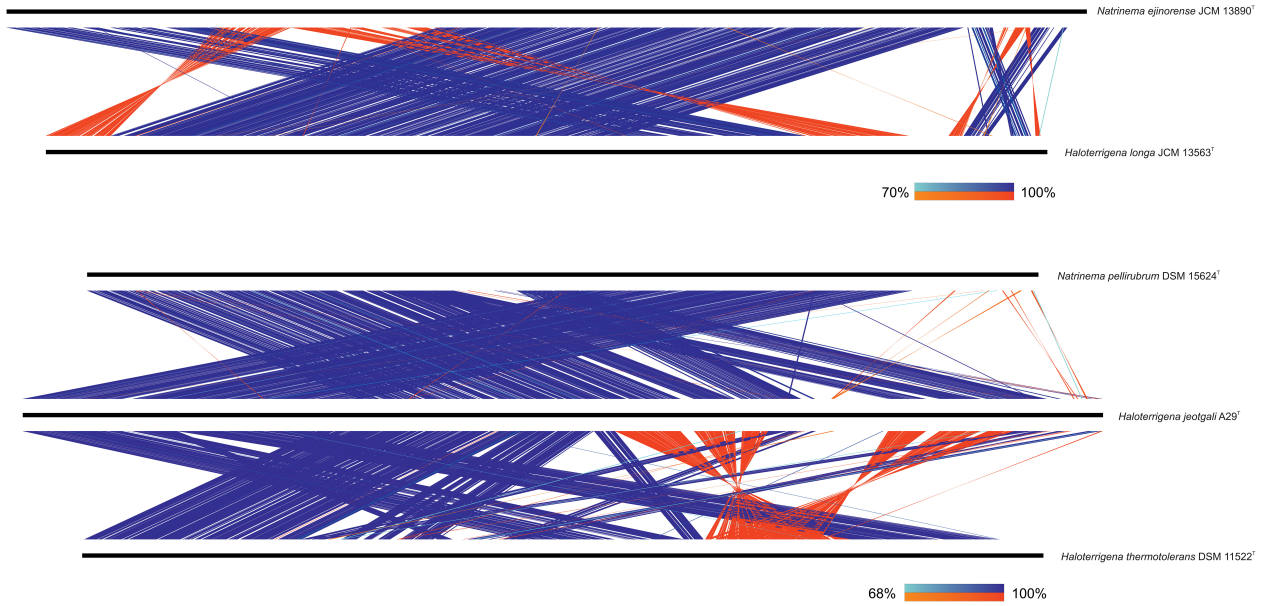


FIGURE 5 | Synteny plot between the genomes of *Natrinema ejinorense* and *Haloterrigena longa* **(above)** and among the genomes of *Natrinema pellirubrum*, *Haloterrigena jeotgali*, and *Haloterrigena thermotolerans* **(below)**. Only matches with ≥ 500 bp alignment length and ≥ 90% identity are shown.

must be noted that the synteny between *Haloterrigena jeotgali* A29^T and *Haloterrigena thermotolerans* DSM 11522^T seemed to be more disorganized than that for other genome pairs; however, this fact is due to the elevated fragmentation of the genome sequence from *Haloterrigena thermotolerans* DSM 11522^T (68 scaffolds and a N50 of 162,183 bp), which reduces the robustness of the synteny analysis. In any case, the synteny results are not so relevant for such genome pair since OGRI values undoubtedly demonstrated the synonym between those species, as stated earlier. The other genome sequences analyzed here for synteny comparisons possessed high-quality, with a minimum N50 of 3.59 Mb and, therefore, they met the requirements to be confidently used for this purpose (Liu et al., 2018).

Our results concerning the study of regions of local collinearity support the union of *Natrinema ejinorense* and *Haloterrigena longa* and of *Natrinema pellirubrum* and *Haloterrigena jeotgali*/*Haloterrigena thermotolerans* as a single species, respectively. Nevertheless, phenotypic features should also be considered before those proposals can be formulated.

Phenotypic Characteristics Endorse the Taxonomic Rearrangements for the Genera *Natrinema* and *Haloterrigena*

For an accurate classification of a taxon, three major premises should be fulfilled: (i) monophyly, (ii) genomic coherence, and (iii) phenotypic coherence (Rosselló-Móra and Amann, 2015). In the previous sections we have examined the two first criteria (phylogenetic/phylogenomic trees and OGRI/synteny), but any formal taxonomic proposal should also be supported by phenotypic characters.

The species *Haloterrigena turkmenica*, *Haloterrigena salifodinae*, and *Haloterrigena salina*, which we propose to be retained as members of the genus *Haloterrigena*, shared a bunch of characteristics (**Table 2**), such as the coccoid morphology, the red pigmentation, the resistance to lysis in distilled water, the high salt concentration for optimal growth [>15% (w/v) NaCl], the inability to produce gas from nitrate, to form indole and H_2S, and to hydrolyze starch, gelatin and Tween 80, the ability to use D-glucose, D-mannose and lactose as a sole carbon and energy sources, the presence of phosphatidylglycerol (PG), phosphatidylglycerol phosphate methyl ester (PGP-Me) and mannose-2,6-disulfate (1→2)-glucose glycerol diether (S_2-DGD-1) as membrane polar lipids, and the lack of phosphatidylglycerol sulfate (PGS). On the other hand, the species *Haloterrigena daqingensis*, which formed a monophyletic cluster with the species of the genus *Natronorubrum*, showed some phenotypic similarities with the species of the latter genus, remarkably, the haloalkaliphilic behavior, the inability to hydrolyze casein and to assimilate D-ribose, D-mannitol and sorbitol, the presence of PG and PGP-Me, and the absence of PGS (**Table 2**). Finally, the remaining species of *Haloterrigena* together to the species of the genus *Natrinema* and *Halopiger salifodinae* lysed in distilled water, grew optimally in media with 15–29% (w/v) NaCl, utilized acetate but not D-mannitol as the only carbon and energy sources, and possessed PG and PGP-Me as major polar lipids (**Table 2**). Some differences in the minor polar lipid composition of this *Natrinema*/*Haloterrigena*/*Halopiger salifodinae* group can be observed, in particular, the presence of S_2-DGD-1 glycolipid in some species and its absence in others, and the lack of PGS in several taxa but not in all of them. Although previous studies have shown that there are

TABLE 2 | Main comparative phenotypic features among members of the genera *Natrinema* (including the species *Halopiger salifodinae*), *Haloterrigena,* and *Natronorubrum*.

Characteristics	*Nnm. altunense*[a]	*Nnm. ejinorense*[b]	*Nnm. gari*[c]	*Nnm. pallidum*[d]	*Nnm. pellirubrum*[d]	*Nnm. salaciae*[e]	*Nnm. soli*[f]	*Nnm. versiforme*[g]	*Htg. hispanica*[h]
Morphology	Rods	Pleomorphic	Rods	Rods	Rods	Pleomorphic	Oval	Pleomorphic	Coccoid
Cell size	0.8–1.2 × 3.0–7.0	0.8–2.0 × 1.5–4.0	0.5–0.8 × 2.0–3.0	0.7–1.0 × 1.5–6.0	0.6–1.0 × 1.0–4.0	0.8–1.5 × 1.0–3.0	1.2–1.6	ND	1.5–2.0
Motility	+	−	+	+	+	−	−	−	−
Colony pigmentation	Orange or red	Light red	Pale orange	Pale orange, beige or almost colorless	Light red or orange	Red	Cream	Light red	Light red
Cells lyse in distilled water	ND	ND	ND	ND	ND	ND	ND	ND	+
NaCl range, % (w/v) (optimum)	>10 (17.5–25)	>10.5 (20)	>10 (15–20)	> 10 (20–25)	>12 (20–25)	10–30 (15–20)	17.5–26 (23)	>9 (20–25)	13–23 (20)
$MgCl_2$ requirement	+	−	−	ND	ND	−	−	+	−
Temperature range (optimum)	20–60 (37–40)	25–50 (37)	20–60 (37–40)	25–60 (37–40)	20–45 (30–37)	30–52.5 (45)	25–45 (40)	20–53 (37–46)	37–60 (50)
pH range (optimum)	6.0–8.0 (7.0–7.7)	6.0–8.5 (7.0)	5.5–8.5 (6.0–6.5)	6.0–8.4 (7.2–7.6)	6.0–8.6 (7.2–7.8)	6.5–9.0 (7.0–8.0)	6.0–8.0 (7.0)	6.0–8.0 (6.5–7.0)	6.5–8.5 (7.0)
Anaerobic growth in presence of:									
Nitrate	+	−	−	+	−	+	−	+	ND
L-arginine	ND	−	ND	ND	ND	−	−	ND	ND
Oxidase	+	+	+	+	−	+	−	+	+
Catalase	+	+	+	ND	ND	+	+	ND	+
Nitrate reduction to nitrite	+	+	−	+	+	+	−	+	+
Gas from nitrate	+	+	−	−	−	ND	ND	+	ND
Indole production	−	−	−	−	−	ND	−	+	+
H_2S production	+	−	ND	−	−	ND	−	+	+
Hydrolysis of:									
Starch	−	+	−	−	−	+	−	±	−
Casein	−	−	−	ND	ND	+	−	−	−
Gelatin	+	+	+	+	+	+	−	−	−
Tween 80	+	+	−	+	−	ND	+	+	−

(Continued)

TABLE 2 | (Continued)

Characteristics	*Nnm. altunense*[a]	*Nnm. ejinorense*[b]	*Nnm. gari*[c]	*Nnm. pallidum*[d]	*Nnm. pellirubrum*[d]	*Nnm. salaciae*[e]	*Nnm. soli*[f]	*Nnm. versiforme*[g]	*Htg. hispanica*[h]
Utilization as sole carbon and energy source of:									
Acetate	+	+	ND	ND	ND	+	ND	ND	+
D-Glucose	+	+	+	+	+	−	+	+	−
D-Fructose	−	+	ND	+	+	−	ND	+	−
D-Galactose	−	−	ND	ND	ND	−	+	+	−
D-Mannose	+	−	−	ND	ND	−	−	+	−
D-Ribose	−	−	−	−	+	−	−	+	−
D-Xylose	−	−	−	ND	ND	−	−	+	−
Sucrose	−	+	−	ND	ND	−	−	+	−
Maltose	+	+	−	ND	ND	−	+	+	−
Lactose	−	−	−	+	+	−	−	−	ND
Glycerol	+	+	+	ND	ND	−	ND	+	+
Sorbitol	ND	−	−	ND	ND	−	−	ND	ND
D-Mannitol	ND	−	−	ND	ND	−	−	ND	ND
Acid production from:									
D-Mannose	+	−	ND	ND	ND	+	−	ND	ND
D-Glucose	+	−	ND	ND	ND	+	−	ND	ND
Lipids:									
Major polar lipids	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me
Major glycolipids	Unidentified glycolipid	S_2-DGD-1	Unidentified glycolypids	Unidentified glycolypids	Unidentified glycolypids	S_2-DGD-1	S_2-DGD-1	Unidentified glycolypids	S-DGD-1
Presence of PGS	+	−	+	+	+	+	−	+	−

(Continued)

TABLE 2 | (Continued)

Characteristics	*Htg. jeotgali*[i]	*Htg. limicola*[j]	*Htg. longa*[j]	*Htg. mahii*[k]	*Htg. saccharevitans*[l]	*Htg. thermotolerans*[m]	*Hpg. salifodinae*[n]	*Htg. salifodinae*[o]	*Htg. salina*[p]	*Htg. turkmenica*[q]
Morphology	Rods	Rods	Rods	Rods	Rods	Rods	Pleomorphic	Coccoid or oval	Coccoid	Coccoid
Cell size	0.4 × 1.0	0.6–0.8 × 1.8–3.6	0.5–0.6 × 2.8–11.0	3.0–10.0	0.4–1.0 × 3.0–10.0	0.7–1.0 × 4.0–13.0	0.1–0.3	1.1–1.5 × 1.1–1.7	1.2–1.6	1.5–2.0
Motility	−	+	−	−	+	−	−	−	−	+
Colony pigmentation	Light red	Red	Red	Deep red	Light red	Pale red	Cream	Pale red	Light red	Red or light pink
Cells lyse in distilled water	+	+	+	+	+	+	+	−	−	−
NaCl range, % (w/v) (optimum)	10–30 (15–20)	10–30 (18)	10–30 (18)	12–30 (20–29)	>10 (17.5–20)	12–25 (17.5–20)	11–32 (17–20)	10–30 (20–25)	15–29 (20)	>12 (15–20)
$MgCl_2$ requirement	−	+	−	ND	−	−	−	−	−	+
Temperature range (optimum)	17–50 (37–45)	30–61 (40–45)	30–56 (41–45)	35–55 (45–55)	24–58 (42–45)	25–60 (50)	25–50 (37–45)	20–55 (42)	25–50 (37)	20–55 (45)
pH range (optimum)	6.5–8.5 (7.0–7.5)	6.5–9.0 (7.0)	6.5–9.0 (7.0–7.5)	6.0–9.0 (6.5–8.2)	6.5–8.5 (7.5)	6.5–8.2 (7.0–7.5)	6.0–8.0 (7.0)	6.0–9.5 (7.5–8.0)	6.0–9.0 (7.0–8.0)	ND
Anaerobic growth in presence of:										
Nitrate	+	−	−	ND	+	−	−	−	−	ND
L-arginine	ND	−	−	ND	ND	−	−	−	−	ND
Oxidase	−	+	+	−	+	+	+	−	+	−
Catalase	+	+	+	+	+	+	+	+	+	ND
Nitrate reduction to nitrite	−	+	−	+	+	+	−	−	+	+
Gas from nitrate	ND	ND	ND	ND	−	−	ND	−	−	−
Indole production	+	−	+	+	−	−	−	−	−	−
H_2S production	ND	+	+	+	+	+	+	−	−	−

(Continued)

TABLE 2 | (Continued)

Characteristics	*Htg. jeotgali*[i]	*Htg. limicola*[j]	*Htg. longa*[j]	*Htg. mahii*[k]	*Htg. saccharevitans*[l]	*Htg. thermotolerans*[m]	*Hpg. salifodinae*[n]	*Htg. salifodinae*[o]	*Htg. salina*[p]	*Htg. turkmenica*[q]
Hydrolysis of:										
Starch	−	−	−	−	−	−	−	−	−	−
Casein	+	−	−	−	−	−	−	ND	−	ND
Gelatin	−	−	−	−	−	+	−	−	−	−
Tween 80	+	−	−	+	+	+	−	−	−	−
Utilization as sole carbon and energy source of:										
Acetate	+	+	+	ND	+	ND	+	+	+	ND
D-Glucose	−	−	+	+	−	−	+	+	+	+
D-Fructose	+	−	−	+	−	−	−	−	+	+
D-Galactose	ND	−	−	ND	−	−	−	+	+	ND
D-Mannose	ND	−	−	ND	−	−	+	+	+	+
D-Ribose	ND	−	−	−	−	−	−	−	−	+
D-Xylose	ND	−	−	−	−	−	−	−	+	−
Sucrose	−	−	+	+	−	−	ND	+	−	+
Maltose	ND	−	+	ND	−	−	−	−	+	ND
Lactose	+	−	−	+	−	−	−	+	+	+
Glycerol	ND	ND	ND	ND	+	−	−	+	+	ND
Sorbitol	ND	−	−	ND	−	ND	+	−	−	ND
D-Mannitol	ND	−	−	ND	−	−	−	−	−	ND
Acid production from:										
D-Mannose	ND	−	−	−	−	−	+	−	−	+
D-Glucose	ND	−	+	−	−	−	+	−	−	+
Lipids:										
Major polar lipids	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me
Major glycolipids	S_2-DGD-1	S_2-DGD-1	S_2-DGD-1	S_2-DGD-1	S_2-DGD-1	S_2-DGD-1	S_2-DGD-1	S-DGD-1, S_2-DGD-1	S-DGD-1, S_2-DGD-1	S_2-DGD-1
Presence of PGS	−	−	−	−	−	−	−	−	−	−

(Continued)

TABLE 2 | (Continued)

Characteristics	*Nrr. aibiense*[r]	*Nrr. bangense*[s]	*Nrr. halophilum*[t]	*Nrr. sediminis*[u]	*Nrr. sulfidifaciens*[v]	*Nrr. texcoconense*[w]	*Nrr. tibetense*[s]	*Htg. daqingensis*[x]
Morphology	Rods	Pleomorphic	Pleomorphic	Pleomorphic	Pleomorphic	Pleomorphic	Pleomorphic	Coccoid
Cell size	0.8–1.0 × 1.4–3.6	ND	ND	0.8–1.0 × 4.0–6.0	ND	0.8–1.0 × 1.0–5.0	ND	0.8–1.3
Motility	+	−	+	−	+	−	−	−
Colony pigmentation	Red	Red	Red	Pink	Red	Pink-red	Red	Orange
Cells lyse in distilled water	+	+	+	+	ND	+	+	−
NaCl range, % (w/v) (optimum)	12–25 (15–18)	12–25 (22.5)	8–28 (15–18)	15–29 (20)	12–28 (18)	10–25 (15–20)	12–30 (20)	10–32 (12–15)
$MgCl_2$ requirement	−	ND	−	−	+	−	ND	ND
Temperature range (optimum)	20–50 (45)	25–55 (45)	25–42 (37)	25–50 (37)	20–55 (44–47)	25–45 (37)	25–55 (45)	20–50 (35)
pH range (optimum)	6.5–9.5 (8.0)	8.0–11.0 (9.5)	5.5–9.5 (7.0-7.5)	8.0–11.0 (9.0)	8.0–10.0 (8.7–9.2)	8.0–10.5 (9.0)	8.5–11.0 (9.0)	8.0–10.5 (10.0)
Anaerobic growth in presence of:								
Nitrate	−	−	+	−	−	−	−	−
L-arginine	−	ND	−	−	−	−	ND	−
Oxidase	+	+	+	+	+	−	+	−
Catalase	+	+	+	+	+	+	+	+
Nitrate reduction to nitrite	+	−	+	+	+	−	−	−
Gas from nitrate	+	−	v	−	+	−	−	ND
Indole production	+	+	+	−	+	−	+	−
H_2S production	−	−	v	−	+	−	−	+
Hydrolysis of:								
Starch	−	−	+	−	−	−	−	−
Casein	−	−	v	−	−	−	−	−
Gelatin	−	−	v	−	−	−	+	−
Tween 80	−	−	v	+	−	−	−	+
Utilization as sole carbon and energy source of:								
Acetate	ND	+	−	−	+	ND	+	+
D-Glucose	+	+	v	+	+	+	+	−

(Continued)

TABLE 2 | (Continued)

Characteristics	*Nrr. aibiense*[r]	*Nrr. bangense*[s]	*Nrr. halophilum*[t]	*Nrr. sediminis*[u]	*Nrr. sulfidifaciens*[v]	*Nrr. texcoconense*[w]	*Nrr. tibetense*[s]	*Htg. daqingensis*[x]
D-Fructose	−	+	−	+	−	ND	+	−
D-Galactose	+	−	+	−	−	ND	−	−
D-Mannose	−	−	v	+	−	+	−	−
D-Ribose	−	ND	−	−	−	ND	ND	−
D-Xylose	−	−	−	+	−	ND	−	−
Sucrose	+	+	+	ND	+	+	+	−
Maltose	+	+	−	−	+	−	+	−
Lactose	−	+	−	−	−	+	+	−
Glycerol	ND	ND	+	−	+	ND	ND	+
Sorbitol	−	ND	−	−	−	ND	ND	−
D-Mannitol	−	−	−	−	−	−	−	−
Acid production from:								
D-Mannose	−	ND	+	ND	ND	ND	ND	−
D-Glucose	+	ND	+	ND	ND	+	ND	−
Lipids:								
Major polar lipids	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me	PG, PGP-Me
Major glycolipids	S_2-DGD-1, TGD-1	None	S_2-DGD-1, TGD-1	None	None	None	None	S_2-DGD-1
Presence of PGS	−	−	−	−	−	−	−	−

+, positive; −, negative; ND, not determined; ±, doubtful; v, variable. Species that should be regarded as member of the genera Natrinema, Haloterrigena, or Natronorubrum are marked in light orange, light blue, and light green, respectively.

[a]*Data from Xu et al. (2005b).*
[b]*Data from Castillo et al. (2006).*
[c]*Data from Tapingkae et al. (2008).*
[d]*Data from McGenity et al. (1998).*
[e]*Data from Albuquerque et al. (2012).*
[f]*Data from Rasooli et al. (2017).*
[g]*Data from Xin et al. (2000).*
[h]*Data from Romano et al. (2007).*
[i]*Data from Roh et al. (2009).*
[j]*Data from Cui et al. (2006b).*
[k]*Data from Ding et al. (2017).*
[l]*Data from Xu et al. (2005a).*
[m]*Data from Montalvo-Rodríguez et al. (2000).*
[n]*Data from Zhang et al. (2013).*
[o]*Data from Chen et al. (2019).*
[p]*Data from Gutiérrez et al. (2008).*
[q]*Data from Zvyagintseva and Tarasov (1987) and Ventosa et al. (1999).*
[r]*Data from Cui et al. (2006a).*
[s]*Data from Xu et al. (1999).*
[t]*Data from Tao et al. (2020).*
[u]*Data from Gutiérrez et al. (2010).*
[v]*Data from Cui et al. (2007).*
[w]*Data from Ruiz-Romero et al. (2013).*
[x]*Data from Wang et al. (2010).*

differences in the polar lipid composition between the species of the genera *Natrinema* and *Haloterrigena* –with *Natrinema* species harboring PGS but not S_2-DGD-1 (McGenity et al., 1998; Xin et al., 2000; Xu et al., 2005b; Tapingkae et al., 2008), while *Haloterrigena* representatives containing S_2-DGD-1 and lacking PGS (Montalvo-Rodríguez et al., 2000; Xu et al., 2005a; Cui et al., 2006b; Roh et al., 2009; Ding et al., 2017)–, we observed that minor polar lipid profiles are not genus-specific. For example, *Natrinema ejinorense* and *Natrinema soli* possessed S_2-DGD-1 and lacked PGS, and *Natrinema salaciae* contained S_2-DGD-1 (characteristic profiles of *Haloterrigena* species). On the contrary, *Haloterrigena hispanica* did not hold S_2-DGD-1 (typical profile of *Natrinema* species). Therefore, those differences in minor polar lipid composition cannot be regarded as phenotypic incoherence within the *Natrinema*/*Haloterrigena*/*Halopiger salifodinae* cluster, whose species should be merged into the single genus *Natrinema*.

With respect to genera differentiation, the genuine genus *Haloterrigena* (*Haloterrigena turkmenica*, *Haloterrigena salifodinae*, and *Haloterrigena salina*) can be distinguished from the now expanded genus *Natrinema* (*Natrinema*/*Haloterrigena*/*Halopiger salifodinae* group) by the resistance to cell lysis in distilled water of the former but not of the latter. Likewise, members of the genus *Natronorubrum* (now also including the species *Haloterrigena daqingensis*) are haloalkalophiles, in contrast to their *Haloterrigena* and *Natrinema* counterparts which better thrive at almost neutral pH values (**Table 2**).

At the species level, phenotypic features can also shed light on uncertain taxa. This is the case of the cluster *Natrinema pellirubrum*/*Haloterrigena jeotgali*/*Haloterrigena thermotolerans* and the cluster *Natrinema ejinorense*/*Haloterrigena longa*, for which OGRI values fell in the fuzzy zone and synteny analysis agreed with the possibility of merging the species within each cluster. A careful inspection of the phenotypic characteristics of *Natrinema pellirubrum*, *Haloterrigena jeotgali*, and *Haloterrigena thermotolerans* demonstrated a similar profile for the two latter, whereas the former showed significant differences as to be considered as a separated species, such as the cell motility, the absence of S_2-DGD-1 glycolipid and the presence of PGS (**Table 2**). On the contrary, phenotypic profile for the species *Natrinema ejinorense* and *Haloterrigena longa* was quite similar, with only minor strain-specific differences (**Table 2**), thus supporting the unification of both taxa into a single species.

Taxonomic Consequences

After having completed detailed phylogenomic, genomic and phenotypic comparative analyses in the family *Natrialbaceae*, and more specifically in the genera *Natrinema* and *Haloterrigena*, we have demonstrated that the species *Haloterrigena jeotgali* and *Natrinema ejinorense* should be considered as later heterotypic synonyms of the species *Haloterrigena thermotolerans* and *Haloterrigena longa*, respectively, according to Rule 23a of the International Code of Nomenclature of Prokaryotes (Parker et al., 2019). Additionally, the species *Haloterrigena hispanica*, *Haloterrigena limicola*, *Haloterrigena longa*/*Natrinema ejinorense*, *Haloterrigena mahii*, *Haloterrigena saccharevitans*, *Haloterrigena thermotolerans*/*Haloterrigena jeotgali*, and *Halopiger salifodinae* should be transferred to the genus *Natrinema*, as *Natrinema hispanicum*, *Natrinema limicola*, *Natrinema longum*, *Natrinema mahii*, *Natrinema saccharevitans*, *Natrinema thermotolerans*, and *Natrinema salifodinae*, respectively. On the contrary, the species *Haloterrigena turkmenica*, *Haloterrigena salifodinae*, and *Haloterrigena salina* will remain as the only representative species of the genus *Haloterrigena*. Besides, the species *Haloterrigena daqingensis* should be reclassified as a member of the genus *Natronorubrum*, as *Natronorubrum daqingense*.

With regards to non-type or unnamed strains, our study indicates that the strains *Natrinema* sp. J7-1, *Natrinema* sp. J7-2, and *Haloterrigena turkmenica* WANU15 should be renamed as *Natrinema gari* J7-1, *Natrinema gari* J7-2, and *Natronolimnohabitans innermongolicus* WANU15, respectively, although it is worth mentioning that the genome sequence of *Haloterrigena turkmenica* WANU15 has been identified as contaminated in a previous study (Lee et al., 2017). Moreover, the strains *Haloterrigena hispanica* CDM_1 and *Haloterrigena hispanica* CDM_6 should not be longer affiliated to the species *Haloterrigena* (*Natrinema*) *hispanica* and, thus, they should be referred as *Natrinema* sp. CDM_1 and *Natrinema* sp. CDM_6, respectively.

On the basis of these data, we propose the following taxonomic re-arrangements.

Description of *Natrinema hispanicum* comb. nov.

Natrinema hispanicum (his.pa'ni.cum. L. neut. adj. *hispanicum* of Hispania, from where the organism was originally isolated)

Basonym: *Haloterrigena hispanica* Romano et al., 2007, 1501.

The description is identical to that of *Haloterrigena hispanica* as given previously (Romano et al., 2007) with the following amendments: the G + C content of the type strain genome is 60.7 mol%, its approximate size 4.26 Mb, and its GenBank Assembly accession number is GCA_004217335.1.

The type strain is FP1^T (= ATCC BAA-1310^T = DSM 18328^T).

Description of *Natrinema limicola* comb. nov.

Natrinema limicola (li.mi'co.la. L. masc. n. *limus* mud; L. suff. *-cola* from L. masc. or fem. n. *incola* dweller; N.L. n. *limicola* mud-dweller)

Basonym: *Haloterrigena limicola* Cui et al., 2006b, 1839.

The description is identical to that of *Haloterrigena limicola* as given previously (Cui et al., 2006b) with the following amendments: the G + C content of the type strain genome is 61.8 mol%, its approximate size 3.52 Mb, and its GenBank Assembly accession number is GCA_000337475.1.

The type strain is AX-7^T (= CGMCC 1.5333^T = JCM 13563^T).

Description of *Natrinema longum* comb. nov.

Natrinema longum (lon'gum. L. neut. adj. *longum* long, referring to the production of long rods in liquid medium)

Basonym: *Haloterrigena longa* Cui et al., 2006b, 1838.

The description is identical to that of *Haloterrigena longa* as given previously (Cui et al., 2006b) with the amendments as follows. Cells are rod-shaped or pleomorphic (0.5–2.0 × 1.5–11.0 μm). Aerobic growth occurs at pH 6.0–9.0 and 25–56°C. Optimal NaCl concentration and temperature for growth are 18–20% (w/v) and 37–45°C, respectively. Nitrate reduction to nitrite is variable. Indole and H_2S formation are variable. Hydrolysis of starch, gelatin and Tween 80 is variable. Assimilation of fructose as carbon and energy sources is variable. Acid production from glucose and sucrose is variable. Phosphatidylglycerol sulfate polar lipid is absent or below detection limit. The DNA G + C content is 61.8–63.9 mol% (genome).

The type strain is ABH32^T (= CGMCC 1.5334^T = JCM 13562^T). The G + C content of the type strain genome is 61.8 mol%, its approximate size 3.52 Mb, and its GenBank Assembly accession number is GCA_020105915.1.

Natrinema ejinorense EJ-57 (= CECT 7144 = CGMCC 1.6202 = DSM 18194 = JCM 13890) is an additional strain of *Natrinema longa*. The G + C content of this reference strain genome is 63.9 mol%, its approximate size 4.48 Mb, and its GenBank Assembly accession number is GCA_002494345.1.

Description of *Natrinema mahii* comb. nov.

Natrinema mahii (mah'i.i. N.L. gen. n. *mahii* of Mah, in honor of R.A. Mah at UCLA for his noteworthy research in the areas of archaea isolation and classification, and also for initiating the solar saltern sampling in the original description)

Basonym: *Haloterrigena mahii* Ding et al., 2017, 1337.

The description is identical to that of *Haloterrigena mahii* as given previously (Ding et al., 2017) with the following amendments: the G + C content of the type strain genome is 65.1 mol%, its approximate size 3.79 Mb, and its GenBank Assembly accession number is GCA_000690595.2.

The type strain is H13^T (= BCRC 910151^T = NBRC 111885^T).

Description of *Natrinema saccharevitans* comb. nov.

Natrinema saccharevitans (sac.char.e.vi'tans. L. neut. n. *saccharon*, *-i* a kind of sugar; L. pres. part. *evitans* shunning, avoiding; N.L. part. adj. *saccharevitans* sugar-avoiding, because it uses very few sugars)

Basonym: *Haloterrigena saccharevitans* Xu et al., 2005a, 2541.

The description is identical to that of *Haloterrigena saccharevitans* as given previously (Xu et al., 2005a) with the following amendments: the G + C content of the type strain genome is 65.3 mol%, its approximate size 3.98 Mb, and its GenBank Assembly accession number is GCA_001953745.1.

The type strain is AB14^T (= AS 1.3730^T = JCM 12889^T).

Description of *Natrinema thermotolerans* comb. nov.

Natrinema thermotolerans (ther.mo.to'le.rans. Gr. fem. n. *therme* heat; L. pres. part. *tolerans* tolerating; N.L. part. adj. *thermotolerans* heat-tolerant)

Basonym: *Haloterrigena thermotolerans* Montalvo-Rodríguez et al., 2000, 1070.

The description is identical to that of *Haloterrigena thermotolerans* as given previously (Montalvo-Rodríguez et al., 2000) with the amendments as follows. Cells are 0.4–1.0 × 1.0–13.0 μm. Aerobic growth occurs in the presence of 10–30% (w/v) NaCl, pH 6.5–8.5 and 17–60°C. Optimal NaCl concentration and temperature for growth are 15–20% (w/v) and 37–50°C, respectively. Anaerobic growth in the presence of nitrate is variable. Oxidase activity, reduction of nitrate to nitrite and indole formation are variable. Hydrolysis of casein and gelatin is variable. Assimilation of fructose and lactose as carbon and energy sources is variable. The DNA G + C content is 65.0–65.4 mol% (genome).

The type strain is PR5^T (= ATCC 700275^T = DSM 11552^T). The G + C content of the type strain genome is 65.4 mol%, its approximate size 3.90 Mb, and its GenBank Assembly accession number is GCA_000337115.1.

Haloterrigena jeotgali A29 (= CECT 7218 = DSM 18794 = JCM 14585 = KCTC 4020) is an additional strain of *Natrinema thermotolerans*. The G + C content of this reference strain genome is 65.0 mol%, its approximate size 4.90 Mb, and its GenBank Assembly accession number is GCA_004799625.1.

Description of *Natrinema salifodinae* comb. nov.

Natrinema salifodinae (sa.li.fo.di'nae. N.L. gen. fem. n. *salifodinae* of a saltpit, salt mine)

Basonym: *Halopiger salifodinae* Zhang et al., 2013, 3565.

The description is identical to that of *Halopiger salifodinae* as given previously (Zhang et al., 2013) with the following amendments: the G + C content of the type strain genome is 65.4 mol%, its approximate size 4.27 Mb, and its GenBank Assembly accession number is GCA_900110455.1.

The type strain is KCY07-B2^T (= CGMCC 1.12284^T = DSM 26231^T = JCM 18547^T).

Description of *Natronorubrum daqingense* comb. nov.

Natronorubrum daqingense (da.qing.en'se. N.L. neut. adj. *daqingense* pertaining to Daqing, north-east China, where the type strain was isolated)

Basonym: *Haloterrigena daqingensis* Wang et al., 2010, 2270.

The description is identical to that of *Haloterrigena daqingensis* as given previously (Wang et al., 2010) with the following amendments: the G + C content of the type strain genome is 61.3–61.4 mol%, its approximate size 3.83–3.84 Mb, and its GenBank Assembly accession numbers are GCA_900156445.1 and GCA_001971705.1.

The type strain is JX313^T (= CGMCC 1.8909^T = NBRC 105739^T).

Emended description of the genus *Natrinema*

Natrinema (Na.tri.ne'ma. N.L. n. *natrium* sodium; Gr. neut. n. *nema* a thread; N.L. neut. n. *Natrinema* the sodium thread, referring to the high sodium ion requirement, and the cell shape)

Cells are rods, coccoid or pleomorphic. Cells lyse at low NaCl concentration (<1.0 M). Colonies are red, light

orange-red, pale orange-red, or cream pigmented. Chemo-organotroph. Some species are strict aerobes, whereas others show anaerobic growth with nitrate. Catalase positive. Grows on a wide range of substrates, including single and complex carbon sources. Extremely halophilic, requiring at least 9–10% (w/v) NaCl for growth, with optimum at 15–29% (w/v) NaCl. Grows at pH values of 5.5–9.0, with optimum pH at 6.0–8.2. Temperature supporting growth ranges from 17 to 61°C, with optimum at 30–55°C. Possesses $C_{20}C_{20}$ and $C_{20}C_{25}$ diether core lipids. The major polar lipids consist of phosphatidylglycerol and phosphatidylglycerol-phosphate-methyl ester, with some species also containing phosphatidylglycerol sulfate. Most species possess the glycolipid S_2-DGD-1, while some species possess S-DGD-1 or unidentified glycolipids. The DNA G + C content is in the range of 60.7–65.4 mol% (genome). The genus is a member of the family *Natrialbaceae*, order *Natrialbales*, class *Halobacteria*. The recommended three-letter abbreviation is *Nnm*. The type species is *Natrinema pellirubrum*.

Emended description of the genus *Haloterrigena*

Haloterrigena (Ha.lo.ter.ri'ge.na. Gr. n. *hals* halos the sea, salt; L. fem. adj. *terrigena* born from the earth; N.L. fem. n. *Haloterrigena* salt (-requiring) and born from the earth).

Cells are Gram-strain-negative, coccoid, or oval-shaped, and 1.1–2.0 μm in size. Colonies are colored light red or light pink due to the presence of bacterioruberin carotenoids. Cells are non-motile or motile and aerobic. Catalase-positive and oxidase-variable. Extremely halophilic, with growth occurring in media containing 10–30% (w/v) NaCl, with optimum at 15–25% (w/v) NaCl. Cells lyse in distilled water. Species may require or not magnesium to grow. Grows at pH values of 6.0–9.5, with optimum pH at 7.0–8.0. Temperature supporting growth ranges from 20 to 55°C, with optimum at 37–45°C. Some species reduce nitrate to nitrite but they do not form gas from nitrate. Indole formation and H_2S production are negative. Hydrolysis of starch, gelatin and Tween 80 is negative. Chemo-organotrophic. All species use sugars, some of them with the production of acids. The major polar lipids are $C_{20}C_{20}$ and $C_{20}C_{25}$ glycerol diether derivatives of phosphatidylglycerol and phosphatidylglycerol-phosphate-methyl ester as well as the glycolipid S_2-DGD-1. Some species may also contain the glycolipid S-DGD-1. Phosphatidylglycerol sulfate is absent. The DNA G + C content is between 64.5 and 65.4 mol% (genome). The genus is a member of the family *Natrialbaceae*, order *Natrialbales*, class *Halobacteria*. The recommended three-letter abbreviation is *Htg*. The type species is *Haloterrigena turkmenica*.

Emended description of the genus *Natronorubrum*

Natronorubrum (Na.tro.no.ru'brum. Gr. n. *natron* derived from Arabic *natrun* soda (sodium carbonate); L. neut. adj. *rubrum* red; N.L. neut. n. *Natronorubrum* the red of soda).

Cells are Gram-strain-negative, rods, coccoid or pleomorphic (flat, triangular, square, disc and other polygonal shapes). Colonies are red, pink, or orange pigmented. Cells are non-motile or motile, aerobic or facultative anaerobic. Catalase-positive and oxidase-variable. Extremely halophilic, with growth occurring in media containing 8–32% (w/v) NaCl, with optimum at 12–22.5% (w/v) NaCl. Cells from most species are lysed in distilled water, but others are not. Alkaliphilic or neutrophilic, growing at pH values of 5.5–11.0, with optimum pH at 7.0–10.0. Temperature supporting growth ranges from 20 to 55°C, with optimum at 35–47°C. Chemo-organotrophic. Many substrates are utilized, sometimes with acid production. The major polar lipids are $C_{20}C_{20}$ and $C_{20}C_{25}$ derivatives of phosphatidylglycerol and phosphatidylglycerol-phosphate-methyl ester. Phosphatidylglycerol sulfate is absent. Cells may also contain S_2-DGD-1, TGD-1 and other unidentified glycolipids. The DNA G + C content is in the range of 60.4–63.6 mol% (genome). The genus is a member of the family *Natrialbaceae*, order *Natrialbales*, class *Halobacteria*. The recommended three-letter abbreviation is *Nrr*. The type species is *Natronorubrum bangense*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

RRH, HM, MK, YS, and AV: conceptualization, investigation, writing – review and editing. RRH, HM, and YS: methodology, formal analysis. RRH, HM, and MK: validation. HM, MK, YS, and AV: resources, project administration, and funding acquisition. RRH and HM: data curation, writing – original draft preparation, and visualization. MK, YS, and AV: supervision. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by Junta de Andalucía, Spain (grants US-1263771 [US/JUNTA/FEDER, UE], P20_01066 and BIO-213, which included FEDER funds), and FEDER/Spanish Ministry of Science and Innovation-State Research Agency (projects CGL2017-83385-P and PID2020-118136GB-I00).

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.2021.740909/full#supplementary-material

REFERENCES

Albuquerque, L., Taborda, M., La Cono, V., Yakimov, M., and da Costa, M. S. (2012). *Natrinema salaciae* sp. nov., a halophilic archaeon isolated from the deep, hypersaline anoxic Lake Medee in the Eastern Mediterranean Sea. *Syst. Appl. Microbiol.* 35, 368–373. doi: 10.1016/j.syapm.2012.06.005

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

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

Barco, R. A., Garrity, G. M., Scott, J. J., Amend, J. P., Nealson, K. H., and Emerson, D. (2020). A genus definition for *Bacteria* and *Archaea* based on a standard genome relatedness index. *mBio* 11:e02475-19. doi: 10.1128/mBio.02475-19

Bhutkar, A., Russo, S., Smith, T. F., and Gelbart, W. M. (2006). Techniques for multi-genome synteny analysis to overcome assembly limitations. *Genome Inform.* 17, 152–161. doi: 10.11234/gi1990.17.2_152

Borriss, R., Rueckert, C., Blom, J., Bezuidt, O., Reva, O., and Klenk, H.-P. (2011). "Whole genome sequence comparisons in taxonomy," in *Methods in Microbiology. Taxonomy of Prokaryotes*, eds F. Rainey and A. Oren (London: Academic Press), 409–436. doi: 10.1016/B978-0-12-387730-7.00018-8

Bushnell, B. (2020). *BBMap Short Read Aligner, and Other Bioinformatic Tools*. Available online at: https://sourceforge.net/projects/bbmap (accessed May 31, 2020).

Castillo, A. M., Gutiérrez, M. C., Kamekura, M., Xue, Y., Ma, Y., Cowan, D. A., et al. (2006). *Natrinema ejinorense* sp. nov., isolated from a saline lake in Inner Mongolia, China. *Int. J. Syst. Evol. Microbiol.* 56, 2683–2687. doi: 10.1099/ijs.0.64421-0

Chen, S., Xu, Y., Sun, S., and Chen, F. (2019). *Haloterrigena salifodinae* sp. nov., an extremely halophilic archaeon isolated from a subterranean rock salt. *Antonie van Leeuwenhoek* 112, 1317–1329. doi: 10.1007/s10482-019-01264-w

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

Corral, P., de la Haba, R. R., Infante-Domínguez, C., Sánchez-Porro, C., Amoozegar, M. A., Papke, R. T., et al. (2018). *Halorubrum chaoviator* Mancinelli *et al.* 2009 is a later, heterotypic synonym of *Halorubrum ezzemoulense* Kharroub *et al.* 2006. Emended description of *Halorubrum ezzemoulense* Kharroub *et al.* 2006. *Int. J. Syst. Evol. Microbiol.* 68, 3657–3665. doi: 10.1099/ijsem.0.003005

Cui, H.-L., Tohty, D., Zhou, P.-J., and Liu, S.-J. (2006b). *Haloterrigena longa* sp. nov. and *Haloterrigena limicola* sp. nov., extremely halophilic archaea isolated from a salt lake. *Int. J. Syst. Evol. Microbiol.* 56, 1837–1840. doi: 10.1099/ijs.0.64372-0

Cui, H.-L., Tohty, D., Feng, J., Zhou, P.-J., and Liu, S.-J. (2006a). *Natronorubrum aibiense* sp. nov., an extremely halophilic archaeon isolated from Aibi salt lake in Xin-Jiang, China, and emended description of the genus *Natronorubrum*. *Int. J. Syst. Evol. Microbiol.* 56, 1515–1517. doi: 10.1099/ijs.0.64222-0

Cui, H.-L., Tohty, D., Liu, H.-C., Liu, S.-J., Oren, A., and Zhou, P.-J. (2007). *Natronorubrum sulfidifaciens* sp. nov., an extremely haloalkaliphilic archaeon isolated from Aiding salt lake in Xin-Jiang, China. *Int. J. Syst. Evol. Microbiol.* 57, 738–740. doi: 10.1099/ijs.0.64651-0

de la Haba, R. R., Corral, P., Sánchez-Porro, C., Infante-Domínguez, C., Makkay, A. M., Amoozegar, M. A., et al. (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

de la Haba, R. R., López-Hermoso, C., Sánchez-Porro, C., Konstantinidis, K. T., and Ventosa, A. (2019). Comparative genomics and phylogenomic analysis of the genus *Salinivibrio*. *Front. Microbiol.* 10:2104. doi: 10.3389/fmicb.2019.02104

Deloger, M., El Karoui, M., and Petit, M.-A. (2009). A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J. Bacteriol.* 191, 91–99. doi: 10.1128/JB.01202-08

Ding, J.-Y., Chen, S.-C., Lai, M.-C., and Liao, T.-L. (2017). *Haloterrigena mahii* sp. nov., an extremely halophilic archaeon from a solar saltern. *Int. J. Syst. Evol. Microbiol.* 67, 1333–1338. doi: 10.1099/ijsem.0.001811

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

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

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Biol.* 20, 406–416. doi: 10.1093/sysbio/20.4.406

Flores, N., Hoyos, S., Venegas, M., Galetoviæ, A., Zúñiga, L. M., Fábrega, F., et al. (2020). *Haloterrigena* sp. strain SGH1, a bacterioruberin-rich, perchlorate-tolerant halophilic archaeon isolated from halite microbial communities, Atacama Desert, Chile. *Front. Microbiol.* 11:324. doi: 10.3389/fmicb.2020.00324

Gaba, S., Kumari, A., Medema, M., and Kaushik, R. (2020). Pan-genome analysis and ancestral state reconstruction of class halobacteria: probability of a new super-order. *Sci. Rep.* 10:21205. doi: 10.1038/s41598-020-77723-6

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, 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

Gutiérrez, M. C., Castillo, A. M., Corral, P., Minegishi, H., and Ventosa, A. (2010). *Natronorubrum sediminis* sp. nov., an archaeon isolated from a saline lake. *Int. J. Syst. Evol. Microbiol.* 60, 1802–1806. doi: 10.1099/ijs.0.015602-0

Gutiérrez, M. C., Castillo, A. M., Kamekura, M., and Ventosa, A. (2008). *Haloterrigena salina* sp. nov., an extremely halophilic archaeon isolated from a salt lake. *Int. J. Syst. Evol. Microbiol.* 58, 2880–2884. doi: 10.1099/ijs.0.2008/001602-0

Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., et al. (2018). RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860. doi: 10.1093/nar/gkx1068

Infante-Domínguez, C., de la Haba, R. R., Corral, P., Sanchez-Porro, C., Arahal, D. R., and Ventosa, A. (2020). Genome-based analyses reveal a synonymy among *Halorubrum* distributum Zvyagintseva and Tarasov 1989; Oren and Ventosa 1996, *Halorubrum terrestre* Ventosa *et al.* 2004, *Halorubrum arcis* Xu *et al.* 2007 and *Halorubrum litoreum* Cui et al., 2007. Emended description of *Halorubrum* distributum Zvyagintseva and Tarasov 1989; Oren and Ventosa 1996. *Int. J. Syst. Evol. Microbiol.* 70, 1698–1705. doi: 10.1099/ijsem.0.003956

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

Jukes, T. H., and Cantor, C. R. (1969). "Evolution of protein molecules," in *Mammalian Protein Metabolism*, Vol. III, ed. H. N. Munro (New York, NY: Academic Press), 21–132. doi: 10.1016/B978-1-4832-3211-9.50009-7

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0

Kim, Y. B., Kim, J. Y., Song, H. S., Lee, C., Ahn, S. W., Lee, S. H., et al. (2018). Novel haloarchaeon *Natrinema thermophila* having the highest growth temperature among haloarchaea with a large genome size. *Sci. Rep.* 8:7777. doi: 10.1038/s41598-018-25887-7

Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi: 10.1038/ismej.2017.113

Konstantinidis, K. T., and Tiedje, J. M. (2005). 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

Lee, I., Chalita, M., Ha, S.-M., Na, S.-I., Yoon, S.-H., and Chun, J. (2017). ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 67, 2053–2057. doi: 10.1099/ijsem.0.001872

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

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Liu, D., Hunt, M., and Tsai, I. J. (2018). Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics* 19:26. doi: 10.1186/s12859-018-2026-4

Mahansaria, R., Dhara, A., Saha, A., Haldar, S., and Mukherjee, J. (2018). Production enhancement and characterization of the polyhydroxyalkanoate produced by *Natrinema ajinwuensis* (as synonym) = *Natrinema altunense* strain RM-G10. *Int. J. Biol. Macromol.* 107, 1480–1490. doi: 10.1016/j.ijbiomac.2017.10.009

McGenity, T. J., Gemmell, R. T., and Grant, W. D. (1998). Proposal of a new halobacterial genus *Natrinema* gen. nov., with two species *Natrinema pellirubrum* nom. nov. and *Natrinema pallidum* nom. nov. *Int. J. Syst. Bacteriol.* 48, 1187–1196. doi: 10.1099/00207713-48-4-1187

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

Minegishi, H., and Kamekura, M. (2019a). "*Haloterrigena*," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. De Vos, et al. (Hoboken, NJ: John Wiley & Sons, Inc., in association with Bergey's Manual Trust), 1–11. doi: 10.1002/9781118960608.gbm00488.pub2

Minegishi, H., and Kamekura, M. (2019b). "*Natrinema*," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. De Vos, et al. (Hoboken, NJ: John Wiley & Sons, Inc., in association with Bergey's Manual Trust), 1–11. doi: 10.1002/9781118960608.gbm00490.pub2

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

Minegishi, H., Shimogaki, R., Enomoto, S., Echigo, A., Kondo, Y., Nagaoka, S., et al. (2016). *Halopiger thermotolerans* sp. nov., a thermo-tolerant haloarchaeon isolated from commercial salt. *Int. J. Syst. Evol. Microbiol.* 66, 4975–4980. doi: 10.1099/ijsem.0.001455

Montalvo-Rodríguez, R., López-Garriga, J., Vreeland, R. H., Oren, A., Ventosa, A., and Kamekura, M. (2000). *Haloterrigena thermotolerans* sp. nov., a halophilic archaeon from Puerto Rico. *Int. J. Syst. Evol. Microbiol.* 50, 1065–1071. doi: 10.1099/00207713-50-3-1065

Oren, A., and Ventosa, A. (2002). International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of *Halobacteriaceae*. Minutes of the meetings, 24 September 2001, Seville, Spain. *Int. J. Syst. Evol. Microbiol.* 52, 289–290. doi: 10.1099/00207713-52-1-289

Oren, A., Ventosa, A., and Kamekura, M. (2017). "*Halobacteria*," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. De Vos, et al. (Hoboken, NJ: John Wiley & Sons, Inc., in association with Bergey's Manual Trust), 1–5. doi: 10.1002/9781118960608.cbm00026.pub2

Palmer, M., Steenkamp, E. T., Blom, J., Hedlund, B. P., and Venter, S. N. (2020). All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. *Int. J. Syst. Evol. Microbiol.* 70, 2937–2948. doi: 10.1099/ijsem.0.004124

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., White, E., Reddy, P., Weigel, G., Kamekura, M., Minegishi, H., et al. (2011). A multilocus sequence analysis approach to the phylogeny and taxonomy of the *Halobacteriales*. *Int. J. Syst. Evol. Microbiol.* 61, 2984–2995. doi: 10.1099/ijs.0.029298-0

Parker, C. T., Tindall, B. J., and Garrity, G. M. (2019). International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* 69, S1–S111. doi: 10.1099/ijsem.0.000778

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Post, F. J., and Al-Harjan, F. A. (1988). Surface activity of halobacteria and potential use in microbially enhanced oil recovery. *Syst. Appl. Microbiol.* 11, 97–101. doi: 10.1016/S0723-2020(88)80055-9

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

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

Ramírez-Durán, N., de la Haba, R. R., Vera-Gargallo, B., Sánchez-Porro, C., Alonso-Carmona, S., Sandoval-Trujillo, H., et al. (2020). Draft genome sequence of *Saccharomonospora piscinae* KCTC 19743^T, an actinobacterium containing secondary metabolite biosynthetic gene glusters. *Microbiol. Resour. Announc.* 9:e01588-19. doi: 10.1128/MRA.01588-19

Ramírez-Durán, N., de la Haba, R. R., Vera-Gargallo, B., Sánchez-Porro, C., Alonso-Carmona, S., Sandoval-Trujillo, H., et al. (2021). Taxogenomic and comparative genomic analysis of the genus *Saccharomonospora* focused on the identification of biosynthetic clusters PKS and NRPS. *Front. Microbiol.* 12:603791. doi: 10.3389/fmicb.2021.603791

Rasooli, M., Naghoni, A., Amoozegar, M. A., Mirfeizi, L., Moshtaghi Nikou, M., Shahzadeh Fazeli, S. A., et al. (2017). *Natrinema soli* sp. nov., a novel halophilic archaeon isolated from a hypersaline wetland. *Int. J. Syst. Evol. Microbiol.* 67, 2142–2147. doi: 10.1099/ijsem.0.001909

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

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

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

Roh, S. W., Nam, Y.-D., Chang, H.-W., Kim, K.-H., Sung, Y., Kim, M.-S., et al. (2009). *Haloterrigena jeotgali* sp. nov., an extremely halophilic archaeon from salt-fermented food. *Int. J. Syst. Evol. Microbiol.* 59, 2359–2363. doi: 10.1099/ijs.0.008243-0

Romano, I., Poli, A., Finore, I., Huertas, F. J., Gambacorta, A., Pelliccione, S., et al. (2007). *Haloterrigena hispanica* sp. nov., an extremely halophilic archaeon from Fuente de Piedra, southern Spain. *Int. J. Syst. Evol. Microbiol.* 57, 1499–1503. doi: 10.1099/ijs.0.64895-0

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

Ruiz-Romero, E., Valenzuela-Encinas, C., López-Ramírez, M. P., de los Angeles Coutiño-Coutiño, M., Marsch, R., and Dendooven, L. (2013). *Natronorubrum texcoconense* sp. nov., a haloalkaliphilic archaeon isolated from soil of the former lake Texcoco (Mexico). *Arch. Microbiol.* 195, 145–151. doi: 10.1007/s00203-012-0852-8

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

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

Sorokin, D. Y., Tourova, T. P., and Muyzer, G. (2005). Oxidation of thiosulfate to tetrathionate by an haloarchaeon isolated from hypersaline habitat. *Extremophiles* 9, 501–504. doi: 10.1007/s00792-005-0465-0

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Tao, C.-Q., Ding, Y., Zhao, Y.-J., and Cui, H.-L. (2020). *Natronorubrum halophilum* sp. nov. isolated from two inland salt lakes. *J. Microbiol.* 58, 105–112. doi: 10.1007/s12275-020-9514-8

Tapingkae, W., Tanasupawat, S., Itoh, T., Parkin, K. L., Benjakul, S., Visessanguan, W., et al. (2008). *Natrinema gari* sp. nov., a halophilic archaeon isolated from fish sauce in Thailand. *Int. J. Syst. Evol. Microbiol.* 58, 2378–2383. doi: 10.1099/ijs.0.65644-0

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* 17, 57–86.

Tindall, B. J. (2003). Taxonomic problems arising in the genera *Haloterrigena* and *Natrinema*. *Int. J. Syst. Evol. Microbiol.* 53, 1697–1698. doi: 10.1099/ijs.0.02529-0

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., et al. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 43, 6761–6771. doi: 10.1093/nar/gkv657

Ventosa, A., Gutiérrez, M. C., Kamekura, M., and Dyall-Smith, M. L. (1999). Proposal to transfer *Halococcus turkmenicus*, *Halobacterium trapanicum* JCM 9743 and strain GSL-11 to *Haloterrigena turkmenica* gen. nov., comb. nov. *Int. J. Syst. Bacteriol.* 49, 131–136. doi: 10.1099/00207713-49-1-131

Wang, S., Yang, Q., Liu, Z.-H., Sun, L., Wei, D., Zhang, J.-Z., et al. (2010). *Haloterrigena daqingensis* sp. nov., an extremely haloalkaliphilic archaeon isolated from a saline-alkaline soil. *Int. J. Syst. Evol. Microbiol.* 60, 2267–2271. doi: 10.1099/ijs.0.013995-0

Westram, R., Bader, K., Prüsse, E., Kumar, Y., Meier, H., Glöckner, F. O., et al. (2011). "ARB: a software environment for sequence data," in *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, ed. F. J. de Bruijn (Hoboken, NJ: Wiley-Blackwell), 399–406. doi: 10.1002/9781118010518.ch46

Wright, A.-D. G. (2006). Phylogenetic relationships within the order *Halobacteriales* inferred from 16S rRNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 56, 1223–1227. doi: 10.1099/ijs.0.63776-0

Xin, H., Itoh, T., Zhou, P., Suzuki, K., Kamekura, M., and Nakase, T. (2000). *Natrinema versiforme* sp. nov., an extremely halophilic archaeon from Aibi salt lake, Xinjiang, China. *Int. J. Syst. Evol. Microbiol.* 50, 1297–1303. doi: 10.1099/00207713-50-3-1297

Xu, X.-W., Liu, S.-J., Tohty, D., Oren, A., Wu, M., and Zhou, P.-J. (2005a). *Haloterrigena saccharevitans* sp. nov., an extremely halophilic archaeon from Xin-Jiang, China. *Int. J. Syst. Evol. Microbiol.* 55, 2539–2542. doi: 10.1099/ijs.0.63761-0

Xu, X.-W., Ren, P.-G., Liu, S.-J., Wu, M., and Zhou, P.-J. (2005b). *Natrinema altunense* sp. nov., an extremely halophilic archaeon isolated from a salt lake in Altun Mountain in Xinjiang, China. *Int. J. Syst. Evol. Microbiol.* 55, 1311–1314. doi: 10.1099/ijs.0.63622-0

Xu, Y., Zhou, P., and Tian, X. (1999). Characterization of two novel haloalkaliphilic archaea *Natronorubrum bangense* gen. nov., sp. nov. and *Natronorubrum tibetense* gen. nov., sp. nov. *Int. J. Syst. Bacteriol.* 49, 261–266. doi: 10.1099/00207713-49-1-261

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

Yoon, S.-H., Ha, S.-M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Zhang, W.-Y., Meng, Y., Zhu, X.-F., and Wu, M. (2013). *Halopiger salifodinae* sp. nov., an extremely halophilic archaeon isolated from a salt mine. *Int. J. Syst. Evol. Microbiol.* 63, 3563–3567. doi: 10.1099/ijs.0.050971-0

Zhang, Z., Liu, Y., Wang, S., Yang, D., Cheng, Y., Hu, J., et al. (2012). Temperate membrane-containing halophilic archaeal virus SNJ1 has a circular dsDNA genome identical to that of plasmid pHH205. *Virology* 434, 233–241. doi: 10.1016/j.virol.2012.05.036

Zvyagintseva, I. S., and Tarasov, A. L. (1987). Extreme halophilic bacteria from saline soils. *Mikrobiologiya* 56, 839–844.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 de la Haba, Minegishi, Kamekura, Shimane 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: 22 October 2021
doi: 10.3389/fmicb.2021.729977

Flavobacterium flabelliforme sp. nov. and *Flavobacterium geliluteum* sp. nov., Two Multidrug-Resistant Psychrotrophic Species Isolated From Antarctica

Stanislava Králová[1], Hans-Jürgen Busse[2], Matěj Bezdíček[3,4], Megan Sandoval-Powers[5], Markéta Nykrýnová[6], Eva Staňková[1], Daniel Krsek[7] and Ivo Sedláček[1]*

[1] Department of Experimental Biology, Czech Collection of Microorganisms, Faculty of Science, Masaryk University, Brno, Czechia, [2] Institut für Mikrobiologie, Veterinärmedizinische Universität Wien, Vienna, Austria, [3] Department of Internal Medicine – Hematology and Oncology, University Hospital Brno, Brno, Czechia, [4] Department of Internal Medicine – Hematology and Oncology, Masaryk University, Brno, Czechia, [5] Department of Biological Sciences, Auburn University, Auburn, AL, United States, [6] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czechia, [7] NRL for Diagnostic Electron Microscopy of Infectious Agents, National Institute of Public Health, Prague, Czechia

OPEN ACCESS

Edited by:
Rafael R. de la Haba, University of Sevilla, Spain

Reviewed by:
Celia Hugo, University of the Free State, South Africa
Paula Maria Tribelli, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

***Correspondence:**
Stanislava Králová
kralova.s@sci.muni.cz

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

Received: *24 June 2021*
Accepted: *27 September 2021*
Published: *22 October 2021*

Citation:
Králová S, Busse H-J, Bezdíček M, Sandoval-Powers M, Nykrýnová M, Staňková E, Krsek D and Sedláček I (2021) Flavobacterium flabelliforme sp. nov. and Flavobacterium geliluteum sp. nov., Two Multidrug-Resistant Psychrotrophic Species Isolated From Antarctica. Front. Microbiol. 12:729977. doi: 10.3389/fmicb.2021.729977

Despite unfavorable Antarctic conditions, such as cold temperatures, freeze-thaw cycles, high ultraviolet radiation, dryness and lack of nutrients, microorganisms were able to adapt and surprisingly thrive in this environment. In this study, eight cold-adapted *Flavobacterium* strains isolated from a remote Antarctic island, James Ross Island, were studied using a polyphasic taxonomic approach to determine their taxonomic position. Phylogenetic analyses based on the 16S rRNA gene and 92 core genes clearly showed that these strains formed two distinct phylogenetic clusters comprising three and five strains, with average nucleotide identities significantly below 90% between both proposed species as well as between their closest phylogenetic relatives. Phenotyping revealed a unique pattern of biochemical and physiological characteristics enabling differentiation from the closest phylogenetically related *Flavobacterium* spp. Chemotaxonomic analyses showed that type strains P4023^T and P7388^T were characterized by the major polyamine *sym*-homospermidine and a quinone system containing predominantly menaquinone MK-6. In the polar lipid profile phosphatidylethanolamine, an ornithine lipid and two unidentified lipids lacking a functional group were detected as major lipids. These characteristics along with fatty acid profiles confirmed that these species belong to the genus *Flavobacterium*. Thorough genomic analysis revealed the presence of numerous cold-inducible or cold-adaptation associated genes, such as cold-shock proteins, proteorhodopsin, carotenoid biosynthetic genes or oxidative-stress response genes. Genomes of type strains surprisingly harbored multiple prophages, with many of them predicted to be active. Genome-mining identified biosynthetic gene clusters in type strain genomes with a majority not matching any known clusters which supports further exploratory research possibilities involving these psychrotrophic bacteria. Antibiotic susceptibility testing revealed a pattern of multidrug-resistant phenotypes that were correlated with *in silico*

antibiotic resistance prediction. Interestingly, while typical resistance finder tools failed to detect genes responsible for antibiotic resistance, genomic prediction confirmed a multidrug-resistant profile and suggested even broader resistance than tested. Results of this study confirmed and thoroughly characterized two novel psychrotrophic *Flavobacterium* species, for which the names *Flavobacterium flabelliforme* sp. nov. and *Flavobacterium geliluteum* sp. nov. are proposed.

Keywords: Antarctica, psychrotrophic bacteria, cold-adaptation, phylogenomics, systematics, *Flavobacterium flabelliforme* sp. nov., *Flavobacterium geliluteum* sp. nov.

INTRODUCTION

For a long time, Antarctica was considered an inhospitable environment with low biodiversity. Although it is a part of Earth's cryosphere which covers about 20% of the Earth's surface (Boetius et al., 2015), it has been of marginal scientific interest due to its difficult accessibility. Until recently, the majority of information on cold-adapted microorganisms was driven from studies originated in Alpine and Arctic areas. Fortunately, advances in technology have paved the way for scientific research in Antarctica which became an attractive place to conduct polar research. As a result, our knowledge on the Antarctic biodiversity predominated by microorganisms has exponentially raised over the recent decades.

Microbial diversity in Antarctica is surprisingly high despite harsh, extreme conditions that living organisms must endure. Many psychrophilic and psychrotrophic taxa have been successfully isolated from Antarctic environments with high abundance of members of the phylum *Bacteroidetes* (Aislabie et al., 2008; Li et al., 2019). The genus *Flavobacterium* is one of the most frequently isolated genera from this phylum, with an extensive number of cold-adapted species. The genus *Flavobacterium* has been constantly growing over the past years and currently comprises 250 established species (LPSN, accessed 27.05.2021) (Parte et al., 2020). Its members are widely distributed, colonizing significantly diverse niches (stream water, lakes, glaciers, soils, rhizosphere or plants) with an obvious preference to cold environments, including Arctic and Antarctic areas (Bernardet and Bowman, 2015). In addition to their ecological importance, some members of the genus *Flavobacterium* are pathogenic to fish and responsible for considerable economic losses (Wahli and Madsen, 2018).

Wide distribution of *Flavobacterium* spp. in Antarctica is supported by their adaptation and coping mechanisms required for survival in harsh conditions including low temperatures, lack of nutrients or intense UV exposure, and freeze-thaw cycles that exert a strong evolutionary pressure on microbial cells (Kumar et al., 2013). Understanding the strategy these microbes employ to survive in cold environments is now increasingly possible due to genomic comparisons of cold-adapted bacteria and their mesophilic relatives. Genes related to the adaptation in psychrophilic and psychrotrophic *Flavobacterium* spp. encode proteorhodopsins, ice-binding proteins, extracellular polysaccharides, and proteins removing reactive oxygen species or cold-shock proteins (Liu et al., 2019). Products of these genes were found useful in biotechnological applications, food-industry, or bioremediation (Cavicchioli et al., 2011) which proportionally increases research interests in cold-adapted microorganisms. The fact that Antarctica represents an extreme and understudied habitat opens many possibilities to explore not only microbial diversity, but also microbial cold-adaptation mechanisms and potential biotechnological applications of its native microbiota. Taxonomic, ecological, physiological and biotechnology-related questions make unique Antarctic *Flavobacterium* spp. an attractive object to study. In addition, the unique resistome of flavobacterial isolates cultivated from a pristine Antarctic environment makes these microorganisms an important model for studying antibiotic resistance beyond the clinical context (González-Aravena et al., 2016; Jara et al., 2020). As such, Antarctic species with novel mechanisms of resistance, including the two novel *Flavobacterium* species described in this study, represent an important source of antibiotic resistance genes (González-Aravena et al., 2016).

MATERIALS AND METHODS

Isolation, Preservation, Culture Conditions

The present study describes a taxonomic investigation of eight *Flavobacterium* strains (**Supplementary Table 1**) isolated in the frame of the microbiological research of the Czech Antarctic Research Program (CARP[1]). Microbiological sampling is conducted yearly on the James Ross Island (near the north-eastern extremity of the Antarctic Peninsula), Antarctica. Sampling is performed during the summer period when the majority of the island becomes an ice-free area. Antarctic summer season allows sampling of upper soil layers, permafrost, lakes and proglacial streams, which are all areas of intense microbial activities (Kopalová et al., 2013; Nedbalová et al., 2013). *Flavobacterium* spp. analyzed in this study were isolated from environmental materials (different water, organic and soil sources) sampled during the years 2010–2019 (**Supplementary Table 1**). Water samples were spread on Reasoner's 2A agar plates (R2A, Oxoid) (150 μl) and cultivated at 15°C for 5 days. Soil samples were processed by dispersing 1 g of soil and/or stone fragment material in 5 ml of sterile saline solution followed by dispersing of 100 μl of obtained solution on R2A

[1]https://polar.sci.muni.cz/en/antarctica

agar plates and cultivated at 15°C for 7 days. Large numbers of well-separated yellow pigmented colonies were randomly selected, hand-picked by sterile inoculation sticks and purified by subculturing. Pure cultures were maintained on the R2A agar slants until transportation to the Czechia. After transportation, subcultures were checked for purity and maintained at −70°C for long-term storage. The type strains of the phylogenetically closest relatives *Flavobacterium hercynium* CCM 9054^T, *Flavobacterium branchiicola* CCM 9061^T, *Flavobacterium chilense* CCM 7940^T, *Flavobacterium araucananum* CCM 7939^T, *Flavobacterium psychroterrae* CCM 8827^T and *Flavobacterium saccharophilum* CCM 8770^T were obtained from the Czech Collection of Microorganisms (CCM[2]) for parallel testing and comparison purposes. All strains were routinely cultivated on R2A agar plates at 20°C for 48 hrs.

Phylogenetic Analyses

Genomic DNA was extracted and purified using High Pure PCR Template Preparation Kit (Roche Diagnostics) according to the manufacturer's recommendations. The 16S rRNA genes were amplified and sequenced using universal bacterial primers pA (5′-AGAGTTTGATCCTGGCTCAG-3′) and pH (5′-AAGGAGGTGATCCAGCCGCA-3′) (Edwards et al., 1989). PCR products were purified using High Pure PCR Product Purification Kit (Roche Diagnostics). Sequencing was performed by the Eurofins MWG Operon sequencing facility[3]. The 16S rRNA sequences were submitted to the EzBioCloud server (Yoon et al., 2017) for initial identification and determination of the closest phylogenetic neighbors. The 16S rRNA sequences of the phylogenetically closest and validly named species were downloaded from the GenBank database[4] and used for multiple sequence alignment using the Molecular Evolutionary Genetics Analysis (MEGA) software (v7.0) (Tamura et al., 2013). The maximum-likelihood (ML) method with the Tamura-Nei gamma distance model was used to calculate the genetic distances and to build the phylogenetic tree.

Whole Genome Sequencing

Genomic DNA was extracted using a High Pure PCR Template Preparation Kit as described above. Sequencing libraries were prepared using KAPA HyperPlus Kits (Roche, Switzerland) with 7 min of enzymatic fragmentation and final fragments length within the range of 550–650 bp. Sequencing was performed on Illumina MiSeq platform with MiSeq Reagent Kit v2 (500-cycles) (Illumina, United States). The St. Petersburg genome assembler (SPAdes v3.11.1) (Bankevich et al., 2012) was used for *de novo* genome assembly. The MismatchCorrector was run to reduce short indels and the number of mismatches. The read coverage cut-off value was automatically computed by the software resulting in contigs and scaffolds. Quality assessment was performed by the QUAST tool (Gurevich et al., 2013) and Bowtie2 v2.4.2 (Langmead and Salzberg, 2012).

Phylogenomics and Genomics

The whole genome sequence (WGS) data were analyzed for further confirmation of the taxonomic status of the analyzed strains and for comparison with genomes of related *Flavobacterium* spp. The WGS data of type strains P4023^T and P7388^T were submitted to the Type Strain Genome server (TYGS) (Meier-Kolthoff and Göker, 2019) and to the Microbial Genome Atlas (MiGA) (Rodriguez-R et al., 2018). To further confirm the taxonomic status of the analyzed strains, type strain genomes along with genomes of the closest related and validly described *Flavobacterium* spp. were subjected to the UBCG pipeline (Na et al., 2018) with default parameters to calculate a phylogenetic tree based on 92 concatenated core genomes. The whole genome sequences of the closest *Flavobacterium* spp. were downloaded from the GenBank database and the genomes of *F. branchiicola* CCM 9061^T and *F. psychroterrae* CCM 8827^T were sequenced and assembled in this study as described above and deposited in the GenBank database under accession numbers GCA_018383905.1 and GCA_018380615.1, respectively. Average nucleotide identity (ANI) values were calculated using the ANI calculator on the Kostas Laboratory website[5] using reciprocal best hits (two-way ANI) (Rodriguez-R and Konstantinidis, 2016). Protein-coding sequences were calculated using the GeneMarkS prediction tool (Besemer et al., 2001) and subjected to the amino acid average identity values (AAI) calculator on the Kostas Laboratory website[6] (Rodriguez-R and Konstantinidis, 2016). Digital DNA-DNA hybridization (dDDH) values were generated using the Genome to Genome Distance Calculator (GGDC) version 2.1 (Meier-Kolthoff et al., 2013).

Additional genome annotation was performed using the Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016; Li et al., 2021) and Operon-Mapper (Taboada et al., 2018). Functional annotation was performed with EggNOG-mapper v5.0 (Huerta-Cepas et al., 2019). CRISPRs were identified by CRISPR Detect (Biswas et al., 2016). Presence of putative prophages was predicted by PHASTER (Arndt et al., 2016) and Prophage Hunter (Song et al., 2019) tools. Genome mining for potential secondary-metabolite associated biosynthetic gene clusters (BGCs) was determined using antiSMASH version 6.0.0 (Blin et al., 2021). Annotations of predicted open reading frames (ORFs) and genes/domains from candidate BGCs were visualized using the R package "gggenes" (Wilkins and Kurtz, 2019) and "ggplot2" (Wickham, 2016). ABRicate v1.0.1[7] was used for screening of known antibiotic and metal resistance genes using the ResFinder database (Zankari et al., 2012), NCBI Bacterial Antimicrobial Resistance Reference Gene Database, CARD (Jia et al., 2017), MEGARes database (Doster et al., 2020) and ARG-ANNOT tool (Gupta et al., 2014). Further *in silico* prediction of resistomes and putative resistance genes was performed using Resistance Gene Identifier (RGI, v5.1.1)[8] with CARD v3.1.1 database (Jia et al., 2017). The enzymatic potential for carbohydrate degradation was evaluated by the web-based tools

[2]http://www.sci.muni.cz/ccm
[3]https://eurofinsgenomics.eu/en
[4]https://www.ncbi.nlm.nih.gov/
[5]http://enve-omics.ce.gatech.edu/ani/
[6]http://enve-omics.ce.gatech.edu/aai/
[7]https://github.com/tseemann/abricate
[8]https://github.com/arpcard/rgi

dbCAN2 and dbCAN-PUL for automated carbohydrate-active enzyme annotation (Yin et al., 2012; Zhang et al., 2018; Ausland et al., 2021).

Morphology

The colony morphology was determined on R2A agar after 48 h cultivation at 20°C. The cellular morphology of all strains was observed by light microscopy after Gram staining. The cellular morphology of the type strains, P4023^T and P7388^T, was also examined by transmission electron microscopy using a Morgagni 268D Philips (FEI Company) electron microscope.

Temperature, NaCl and pH Tolerance

The growth at different temperatures (1–40°C, in 5°C increments) and the salt tolerance (0.5, 1.0, 2.0, 3.0, 4.0 of NaCl) were determined on R2A agar plates. The pH tolerance was assessed in Trypticase Soya Broth (TSB) inoculated with two drops of cell suspension of concentration as McFarland 2.0 with pH 5–10 at intervals of 1.0 pH unit at 20°C adjusted with the following buffer systems: pH 5.0–8.0, 0.1 M KH_2PO_4/0.1 M NaOH; pH 9.0–10.0, 0.1 M $NaHCO_3$/0.1 M Na_2CO_3. The pH value of the TSB was confirmed after autoclaving.

Biochemical and Physiological Characteristics

All eight isolates along with reference strains *F. hercynium* CCM 9054^T, *F. branchiicola* CCM 9061^T, *F. chilense* CCM 7940^T, *F. araucananum* CCM 7939^T, *F. psychroterrae* CCM 8827^T and *F. saccharophilum* CCM 8770^T were phenotypically characterized by the most relevant tests recommended for description of new taxa within the family *Flavobacteriaceae* (Bernardet et al., 2002). These tests comprised of: activities of oxidase (OXItest; Erba-Lachema) and catalase (Atlas, 2010) (ID color Catalase; bioMérieux) performed according to the manufacturer's instructions; production of urease (Christensen, 1946); production of arginine dihydrolase, ornithine and lysine decarboxylase production (Brooks and Sodeman, 1974), oxidation-fermentation test (Hugh and Leifson, 1953); production of acid from carbohydrates aerobically (Gilardi, 1985), production of H_2S on Triple Sugar Iron Agar (HiMedia), hydrolysis of aesculin, starch (Barrow and Feltham, 1993), gelatine, Tween 80 (Pacova and Kocur, 1984), casein, tyrosine (Kurup and Babcock, 1979), cellulose (R2A broth with strip of Whatman paper No.1) (Ali et al., 2009), carboxymethylcellulose (CMC) (Ibrahim et al., 2021) and O-nitrophenyl-β-D-galactopyranoside (ONPG) (Lowe, 1962), phospholipinase activity in the egg-yolk reaction (Owens, 1974), nitrate and nitrite reduction, production of indole, utilization of citrate in Simmon's citrate agar (Barrow and Feltham, 1993), utilization of acetamide (Oberhofer and Rowen, 1974) and sodium malonate (Ewing, 1960). The presence of flexirubin-type pigments was tested using 20% (w/v) KOH solution and the presence of capsule was identified by the Congo red adsorption test (Bernardet et al., 2002).

Capability of growth on different media was tested on R2A agar (Oxoid), Nutrient agar (Oxoid), Plate Count agar (Oxoid), Tryptone Soya agar (Oxoid), marine agar (Lapage et al., 1970), MacConkey and Endo agar (HiMedia), Brain Heart Infusion agar (Oxoid), and Mueller-Hinton agar (Oxoid) under aerobic conditions. Anaerobic growth was tested using Anaerocult A system (Merck) and microaerophilic growth in atmosphere of 5% CO_2, both on R2A agar plates. All above listed tests were inoculated using cells of analyzed strains grown on R2A agar at 20°C for 48 hrs. All tests were read daily for up to 7 days with exception of the tyrosine hydrolysis test (up to 10 days). The utilization of carbon sources, enzyme activities and other additional phenotypic characteristics were further tested using identification test kits GEN III MicroPlateTN (Biolog) with the protocol A, API 20 NE and API ZYM (bioMérieux) according to the manufacturer's instructions.

Antimicrobial Susceptibility

In vitro antimicrobial susceptibility was assessed using the disk diffusion method. Cell suspensions of concentration as 0.5 McFarland were prepared from cultures cultivated on R2A agar. Suspension containing each strain (100 μl) was spread on the Mueller-Hinton agar plates. Tested antibiotic disks (Antimicrobial Susceptibility Disks, Oxoid) were as follows: ampicillin (10 μg), aztreonam (30 μg), carbenicillin (100 μg), cefixime (5 μg), ceftazidime (10 μg), cephalotin (30 μg), ciprofloxacin (5 μg), gentamicin (10 μg), chloramphenicol (30 μg), imipenem (10 μg), kanamycin (30 μg), co-trimoxazole (25 μg), piperacillin (30 μg), polymyxin B (300 U), streptomycin (10 μg) and tetracycline (30 μg). CLSI (The Clinical and Laboratory Standards Institute) and EUCAST (The European Committee on Antimicrobial Susceptibility Testing) standards were strictly followed for cultivation and inhibition zone diameter reading (CLSI, 2015; EUCAST, 2017).

Chemotaxonomic Analyses

Cellular fatty acid methyl esters analysis was performed from biomass of all strains including the reference type strains incubated strictly under the same conditions. All cultures were grown on R2A agar at 20 ± 2°C for 48 hrs to reach the late-exponential stage of growth according to the four quadrant streak method (Sasser, 1990). Fatty acids were identified using an Agilent 7890B gas chromatograph according to the Standard Protocol of the Sherlock Identification System (MIDI Sherlock version 6.2, MIDI database RTSBA 6.21). Further chemotaxonomic analyses including polar lipids, quinones, and polyamines were performed using freeze-dried biomass prepared from cells grown in R2A broth at 20°C for 48 hrs. Extraction of all above mentioned chemotaxonomic markers was performed according to the previously described protocols (Busse and Auling, 1988; Tindall, 1990b,a; Altenburger et al., 1996; Stolz et al., 2007). Analysis of the polyamine patterns was performed using high-performance liquid chromatography (HPLC) with conditions described by Busse et al. (1997) using the HPLC apparatus as described previously (Stolz et al., 2007).

RESULTS AND DISCUSSION

Phylogenetic Analysis

Initial identification based on the 16S rRNA gene sequences classified all Antarctic isolates to the genus *Flavobacterium*. Based on the 16S rRNA gene sequence similarity, Antarctic isolates formed two groups for which strains P4023^T and P7388^T were assigned as type strains. The closest related species to proposed type strain P4023^T were *F. saccharophilum* DSM 1811^T (98.54%), *F. hercynium* DSM 18292^T (98.54%) and *F. chilense* LMG 26360^T (98.39%). A proposed type strain for the second cluster, P7388^T, showed the highest sequence similarities to the *F. branchiicola* 59B-3-09^T (99.12%), *F. araucananum* DSM 24704^T (99.08%) and *F. psychroterrae* CCM 8827^T (99.04%). The 16S rRNA gene sequence similarities of strain P4023^T fell below the cut-off value 98.65%, whereas strain P7388^T reached similarities above this threshold. Nevertheless, several studies have already proved that the standard 16S rRNA gene sequence cut-off value is not sufficient for some genera, such as *Streptomyces, Streptococcus* or *Bacillus* (Fox et al., 1992; Clarridge, 2004; Guo et al., 2008). Indeed, the genus *Flavobacterium* and its species with validly published names also do not follow the 98.65% threshold, which therefore requires additional in-depth phylogenetic and phylogenomic analyses for designation of *Flavobacterium* spp. taxonomic positions and assessment of their novelty (Yi and Chun, 2006; Romanenko et al., 2015).

To evaluate phylogenetic relationships between Antarctic *Flavobacterium* spp. and their closest phylogenetic neighbors as identified using the EzBioCloud database, the ML phylogenetic tree was calculated and clearly showed well-supported clades comprising of strains P4023^T and P7388^T along with two and four non-type isolates, respectively (**Figure 1**). Strains P4023^T, P4911 and CCM 9063 formed a separate cluster with the closest phylogenetic neighbor *F. hercynium* DSM 18292^T supporting the 16S rRNA gene sequence similarities results. Strains P7388^T, P9670, P7381, P7475, and CCM 9065 formed a well-supported cluster close to *Flavobacterium resistens* BD-b365^T and *F. branchiicola* 59B-3-09^T.

Phylogenomics

As mentioned above, clear taxonomic delineation of *Flavobacterium* spp. requires additional comprehensive analyses. For this reason, draft genomes of proposed type strains P4023^T and P7388^T were initially submitted to TYGS and MiGA platforms for genomic comparisons. TYGS results assigned both strains as novel species. Genome-based phylogeny calculated by TYGS showed *F. hercynium* DSM 18292^T as the closest phylogenetic neighbor of strain P4023^T, and *F. araucananum* DSM 27404^T forming a monophyletic cluster with strain P7388^T (**Figure 2**). MiGA results confirmed novelty of both draft genomes and assigned *Flavobacterium weaverense* DSM 19727^T and *Flavobacterium pectinovorum* ATCC 19366^T as the closest related species to P4023^T and P7388^T, respectively. Further phylogenomic analysis through the up-to-date bacterial core gene UBCG pipeline placed strain P4023^T into a cluster shared with *F. weaverense* DSM 19727^T, while strain P7388^T formed a distinct phylogenetic lineage within the genus *Flavobacterium* (**Supplementary Figure 1**).

In silico comparison of genomic distances and calculation of the dDDH values between the draft genomes of strains P4023^T, P7388^T and their closest phylogenetic neighbors resulted in values below 70%, the cut-off established as a threshold value for species delineation by dDDH (Meier-Kolthoff et al., 2013). The ANI values between strain P4023^T and its closest relatives reached 77.60–82.70% with the highest nucleotide similarity to *F. weaverense* DSM 19727^T. Comparison between strain P7388^T and its closest phylogenetic relatives showed ANI values of 78.17–82.06% with *F. araucananum* DSM 24704^T with the highest genomic relatedness. Calculated ANI values were well below the 95–96% threshold for species delineation and further supported novelty of the proposed species (Ciufo et al., 2018; Jain et al., 2018). This was also consistent with the calculated AAI values ranging from 71.24–84.84% and 72.77–82.09% for strains P4023^T and P7388^T, respectively (**Table 1**).

Genomic Analyses

General and Functional Features of P4023^T Genome

The final genome assembly of strain P4023^T contained 57 scaffolds with a draft genome size of 3,631,245 bp and the genomic G + C content 31.2 mol% (**Supplementary Table 2**). PGAP predicted 3,178 genes in total, among which 3,090 belonged to protein-coding genes (CDSs) with 2,525 (81.72%) assigned to a function with high confidence and 565 (18.28%) assigned as hypothetical proteins. The genome contained 48 tRNAs, seven rRNAs (three 5S rRNAs, three 16S rRNAs, one 23S rRNA) and three ncRNAs. The genome does not contain plasmids or CRISPR arrays. Interestingly, seven prophages with four assigned as active were detected by Prophage Hunter and one additional prophage was found by the PHASTER tool.

Functional annotation of the genomes revealed that 359 genes (11.60%) could not be assigned to a specific class of orthologous genes clusters (COGs) or with specific function (667 genes, 21.55%) (**Supplementary Table 3**). However, the remaining majority of genes (67.25%) were assigned to specific COGs classes with the most abundant classes as (M) Cell wall/membrane/envelope biogenesis, (E) Amino acid transport and metabolism, (J) Translation, ribosomal structure and biogenesis, (K) Transcription and (L) Replication, recombination and repair, followed by (C) Energy production and conversion, (P) Inorganic ion transport and metabolism, and (H) Coenzyme transport and metabolism. In total, 76 carbohydrate-active enzymes (CAZymes) were predicted comprising 17 families of glycoside hydrolases (GHs), four families of carbohydrate esterases (CEs), seven families of glycosyltransferases (GTs), and single carbohydrate-binding modules (CBMs) family. Two polysaccharide utilization loci (PULs), PUL0344 and PUL0487, were further predicted, both associated with chitin degradation of *Flavobacterium johnsoniae* (McBride et al., 2009; Larsbrink et al., 2016).

General Features of P7388^T Genome

The final genome assembly of strain P7388^T contained 112 scaffolds with a draft genome size of 4,387,206 bp and the

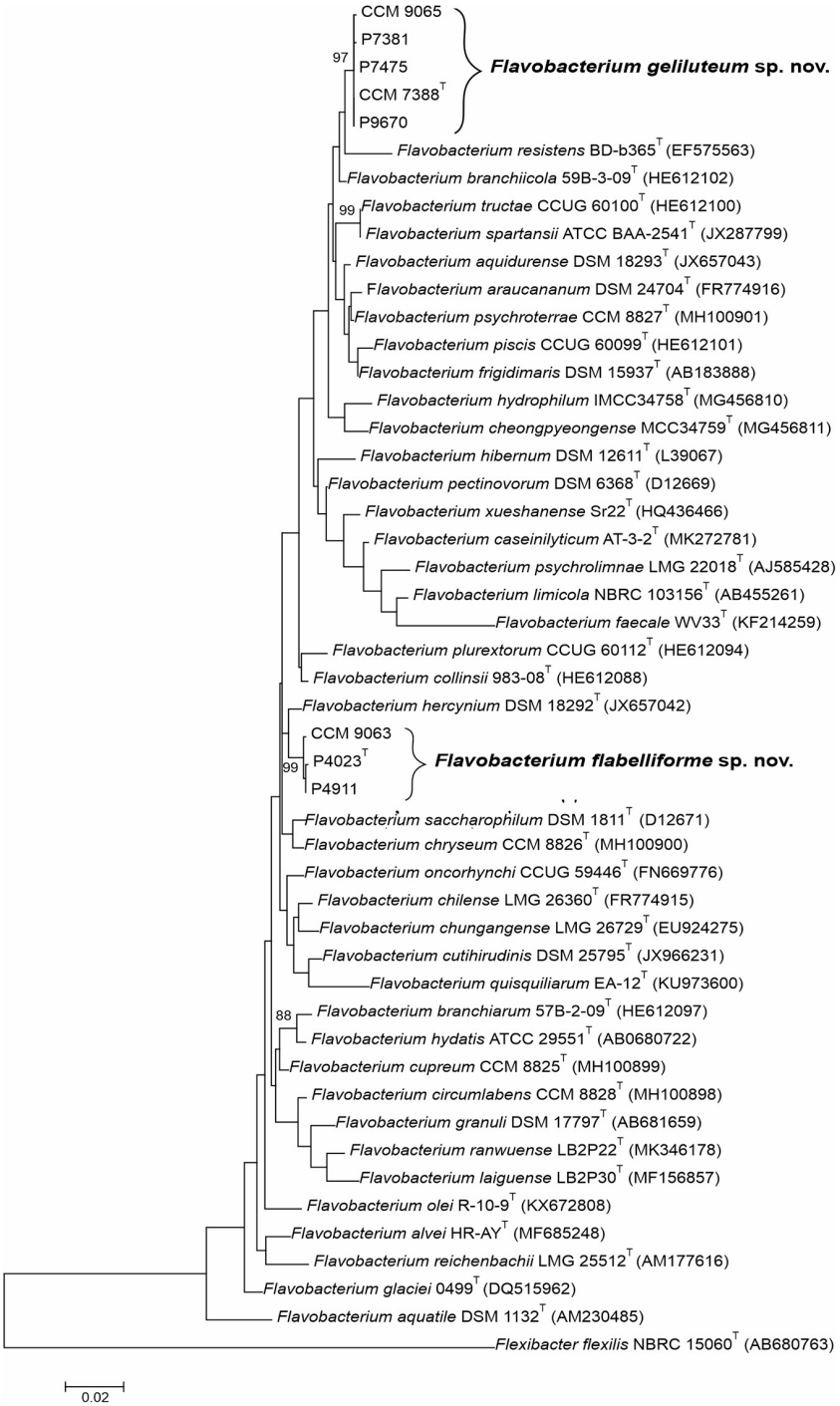


FIGURE 1 | Phylogenetic tree based on 16S rRNA gene sequences comparison showing the phylogenetic positions of *Flavobacterium flabelliforme* sp. nov. and *Flavobacterium geliluteum* sp. nov. among their closest related species within the genus *Flavobacterium*. The evolutionary history was inferred by using the maximum likelihood method based on Tamura-Nei distance with the gamma model. All positions with less than 95% site coverage were eliminated. Bootstrap probability values (percentages of 1,000 tree replications) greater than 70% are indicated at branch points. *Flexibacter flexilis* NBRC 15060^T (AB680763) was used as an outgroup. Bar, 0.02 substitutions per nucleotide position.

genomic G + C content 34.5 mol% (**Supplementary Table 2**). PGAP predicted 3,829 genes in total, among which 3,728 belonged to protein-coding genes (CDSs) with 2,928 (78.54%) assigned to a function with high confidence and 800 (21.46%) assigned as hypothetical proteins. The genome contained 46 tRNAs, three rRNAs (one of each 5S rRNA, 16S rRNA and

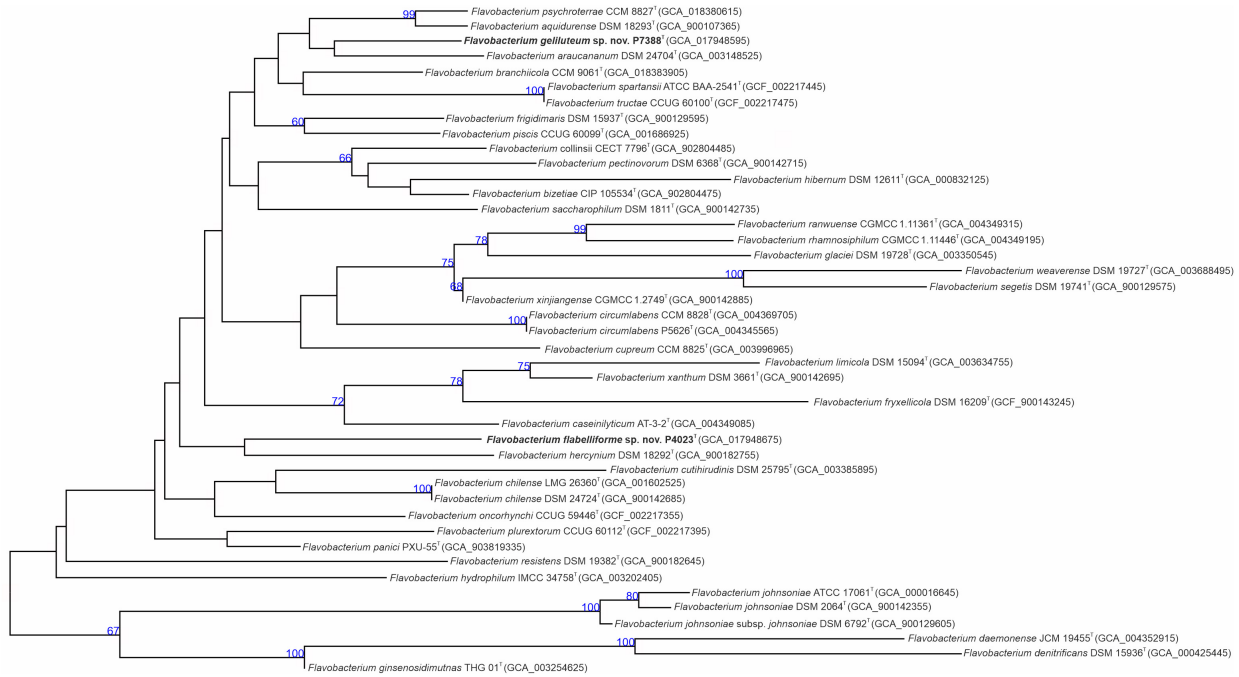


FIGURE 2 | Phylogenomic tree inferred with FastME 2.1.6.1 (Lefort et al., 2015) from GBDP distances calculated from genome sequences. Branch lengths are scaled in terms of GBDP distance formula d_5. Numbers above branches indicate GBDP pseudo-bootstrap support values from 100 replications with the average branch support of 85.1%.

TABLE 1 | ANI, AAI, and dDDH genomic comparisons between strains P4023^T, P7388^T and their closest phylogenetic neighbors.

Reference genomes	GenBank accession number	ANI value (%)		AAI value (%)		dDDH	
		P4023^T	P7388^T	P4023^T	P7388^T	P4023^T	P7388^T
F. hercynium DSM 18292^T	GCA_002217285.1	79.29	81.30	71.87	80.46	20.8	23.3
F. chilense DSM 24724^T	GCA_900142685.1	78.84	81.48	71.24	81.10	20.4	23.4
F. saccharophilum DSM 1811^T	GCA_900142735.1	79.40	82.02	72.32	82.09	21.1	23.9
F. branchiicola CCM 9061^T	GCA_018383905.1	79.22	81.35	71.24	80.05	21.0	24.4
F. araucananum DSM 24704^T	GCA_003148525.1	79.15	82.06	72.13	81.88	21.0	24.4
F. psychroterrae CCM 8827^T	GCA_018380615.1	79.50	82.01	72.09	81.25	21.4	22.9
F. weaverense DSM 19727^T	GCA_003688495.1	82.70	78.72	84.84	72.77	24.6	20.4
F. pectinovorum DSM 6368^T	GCA_900142715.1	79.45	82.03	71.98	81.74	20.6	24.5
F. aquatile LMG 4008^T	GCA_000757385.1	77.60	78.17	68.34	67.49	19.6	19.6
F. geliluteum sp. nov. P7388^T	GCA_017948595.1	79.44	–	78.89	–	20.6	–
F. flabelliforme sp. nov. P4023^T	GCA_017948675.1	–	79.44	–	78.89	–	20.6

23S rRNA) and three ncRNAs. No plasmids were found in the genome. One CRISPR array was identified by CRISPR Detect with medium confidence. A high number of prophages were predicted in the P7388^T genome with 14 prophages detected by Prophage Hunter, six of which were predicted to be active. Six additional incomplete prophage sequences were detected by the PHASTER tool.

Functional annotation revealed that 485 genes (12.97%) could not be assigned to any COGs or with specific function (796 genes, 21.29%) (**Supplementary Table 3**). The remaining majority of genes (74.74%) were assigned to specific COGs classes with the most abundant classes annotated as (M) Cell wall/membrane/envelope biogenesis, (E) Amino acid transport and metabolism, (K) Transcription, (L) Replication, recombination and repair, and (J) Translation, ribosomal structure and biogenesis, followed by (C) Energy production and conversion, (P) Inorganic ion transport and metabolism, and (H) Coenzyme transport and metabolism. Comparison of relative abundances of functional COG classes between strains P4023^T and P7388^T is depicted in **Figure 3**. The genome of P7388^T was predicted to harbor significantly more CAZymes with 176 in total, compared to 76 from P4023^T. Overall, P7388^T harbors 41 families of GHs, nine families of CEs, 10 families of GTs, seven families of CBMs and two polysaccharide lyases (PLs). Prediction

TABLE 2 | Phenotypic characteristics that differentiate *F. flabelliforme* sp. nov and *F. geliluteum* sp. nov. from each other and from their closest phylogenetically related *Flavobacterium* spp. strains.

Tests	*F. flabelliforme* sp. nov.	*F. geliluteum* sp. nov.	1.	2.	3.	4.	5.	6.
Flexirubin-type of pigments	−	+	+	+	+	+	+	+
Growth at:								
pH 9	+	−	+	w	w	w	+	+
1°C	+	−	+	+	+	+	+	w
5°C	+	v	+	−	−	+	+	+
30°C	+	+	+	+	+	w	+	−
Growth on marine agar	+	−	+	−	−	−	−	−
Growth in presence of:								
1.0% NaCl	+	−	+	−	+	+	−	+
Oxidase	+	−	+	+	−	+	+	−
Nitrate reduction	−	−	+	+	−	+	−	+
Nitrite reduction	−	−	+	+	−	−	−	−
Hydrolysis of:								
Aesculin	−	+	+	+	+	+	+	+
ONPG	−	+	+	+	+	+	+	+
Starch	−	+	+	+	+	+	+	+
Tween 80	−	+	−	−	−	−	−	−
DNA	−	−	+	−	+	+	+	+
Tyrosine	+	+	−	+	+	+	+	−
CMC	−	+	+	−	+	+	−	+
Agar	+	−	−	−	−	−	+	−
Production of								
Acid from fructose (aerobically)	−	v	+	+	+	+	+	+
Acid from mannitol (aerobically)	v	−	+	−	−	−	−	+
Acid from xylose (aerobically)	−	+	+	+	+	+	+	+
Arginine dihydrolase	+	−	−	−	+	−	+	−
Utilization of								
Arabinose (API 20 NE)	−	+	−	+	+	−	+	+
Mannose (API 20 NE)	−	+	+	+	+	+	+	+
N-acetyl-glucosamine (API 20 NE)	−	+	+	−	w	−	+	−
API ZYM:								
Valine arylamidase	+	−	+	+	+	+	+	+
Naphtol-AS-BI-phosphohydrolase	−	+	−	−	−	−	−	w
β-galactosidase	−	−	−	−	−	−	+	w
α-glucosidase	−	−	+	−	+	−	+	+
β-glucosidase	−	+	+	+	+	+	+	+
N-acetyl-β-glucosaminidase	−	−	−	−	−	−	+	+

Strains: 1, F. hercynium CCM 9054^T; 2, F. branchiicola CCM 9061^T; 3, F. chilense CCM 7940^T; 4, F. araucananum CCM 7939^T; 5, F. saccharophilum CCM 8770^T; 6, F. psychroterrae CCM 8827^T. +, positive; w, weakly positive; −, negative; v, variable results observed between strains belonging to the same species.

of PULs found seven putative loci involved in degradation of carrageenan, dextran, hemicellulose, chitin, pectin, starch and xylan, which implies the ability of strain P7388^T to degrade various polysaccharide substrates.

Biosynthetic Potential of the Novel Strains

Antarctic bacteria have been recognized as an understudied source of secondary metabolites (Benaud et al., 2020; Waschulin et al., 2020). AntiSMASH annotation of secondary metabolite biosynthesis was performed to explore the biosynthetic potential of strains P4023^T and P7388^T. The analysis indicated that these strains harbor four (P4023^T) and seven (P7388^T) gene clusters potentially related to secondary metabolite production. Out of these, six clusters did not match any known BGCs deposited in the MIBiG database (Kautsar et al., 2020) and other three clusters were below 30% similarity to any known BGCs. Strain P4023^T harbors a BGC encoding a terpene with 28% similarity to a carotenoid BGC of *Algoriphagus* sp. KK10202C. Three additional BGCs were categorized as terpene, arylpolyene and betalactone, and none of these clusters showed significant similarity to known BGCs. Arylpolyenes are widely distributed bacterial natural products that are structurally and functionally similar to carotenoid pigments (Schöner et al., 2016), and are particularly important regarding protection from reactive oxidation, a severe threat in Antarctica. No prediction of flexirubin-associated genes suggests that the yellow pigment of

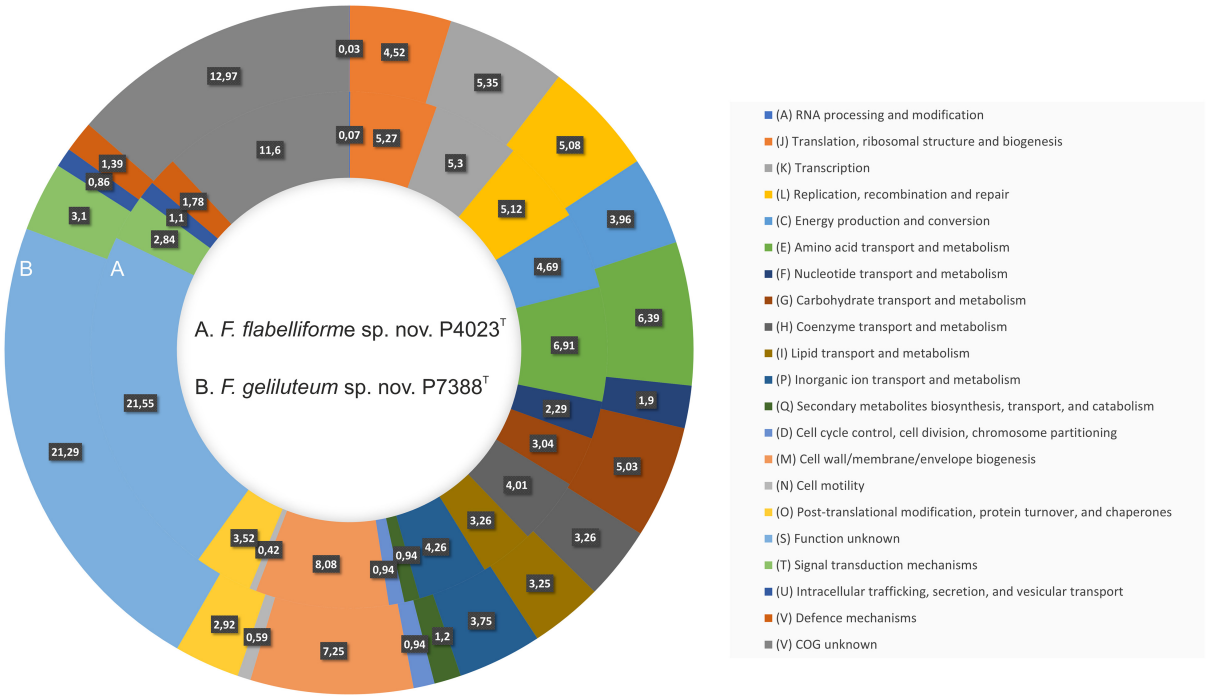


FIGURE 3 | Comparison of COG functional categories between *F. flabelliforme* sp. nov. P4023^T and *F. geliluteum* sp. nov. P7388^T. Each colored segment indicates the relative contribution of a functional category as a percentage of total COGs with the color of COG family indicated in the legend. Ring A, *F. flabelliforme* sp. nov. P4023^T; Ring B, *F. geliluteum* sp. nov. P7388^T.

P4023^T may be of arylpolyene or carotenoid nature. Interestingly, strain P7388^T was predicted to harbor more BGCs than strain P4023^T, and the number of predicted BGCs was similar to *F. chilense* LMG 26360^T and *F. hercynium* DSM 18292^T. Among these BGCs, a core structure was predicted from one hybrid polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) BGC for strain P7388T that did not match any known BGCs (Cluster 3.1, **Figure 4**). Only one putative BGC from strain P7388^T, Cluster 5.1, encoding an arylpolyene/resorcinol pathway showed >50% similarity to a known cluster. Annotation of Cluster 5.1 (**Figure 5**) revealed numerous PFAM hits with genes involved in the core biosynthetic process such as Beta-ketoacyl synthases, as well as transport-related genes, glycosyl transferases, and carbohydrate kinases. Comparison of Cluster 5.1 to the MiBIG database showed that this cluster was 88% similar to a polyketide cluster encoding a flexirubin-type pigment of *F. johnsoniae* UW101 (**Figure 5**) along with two terpenes sharing 20% and 28% similarity to carotenoid BGCs of *Algoriphagus* sp. KK10202C and *Streptomyces avermitilis*, respectively. Thus, yellow pigmentation of the strain P7388^T is likely a result of a pigment mixture of flexirubin-type and carotenoid pigments resulting in the dark yellow to orange pigmentation of colonies. Furthermore, the production of secondary metabolites, especially activities of PKSs and NRPSs, has been associated with bacteria expressing gliding motility (Nett and König, 2007). The Antarctic *Flavobacterium* strains studied not only were predicted to encode some PKS and NRPS clusters but were also found to express different degrees of gliding motility promoted by lower temperatures (<20°C) (**Supplementary Figures 2**, **3**). This phenotypic characteristic is genomically encoded by genes essential or related to gliding motility and type 9 secretion systems (T9SS) present in the genomes of both P4023^T and P7388^T and is listed in **Supplementary Table 4**.

Genomic Features Associated With Cold-Adaptation

A number of genes related to cold-adaptation have been identified in type strain genomes of the two proposed Antarctic species and were compared to three phylogenetically related mesophilic flavobacteria, *F. hercynium* DSM 18292^T, *F. saccharophilum* DSM 1811^T and *F. pectinovorum* DSM 6368^T. Genome-mining focused on cold-adaptation confirmed that *Flavobacterium* spp. are well adjusted to environmental stress and harbor a significant number of genes associated with cold-adaptation regardless of their thermotype (**Supplementary Table 5**).

Cold-shock-inducible proteins are widely distributed among various bacteria and include cold-shock proteins (), ribosome-binding factor A (*rbfA*), transcription termination/antitermination factor (*nusA*), translation initiation factors IF1 and IF2 (*infA*, *infB*), polynucleotide nucleotidyltransferase (*pnp*), heat-shock cognate proteins (*hsc*), recombinase A (*recA*) and many others (Gualerzi et al., 2003; Hesami et al., 2011; Barria et al., 2013). Comparison of psychrotrophic and mesophilic genomes showed that presence of these genes is highly conserved among *Flavobacterium* spp. and may be partially responsible for their ubiquitous nature. Most

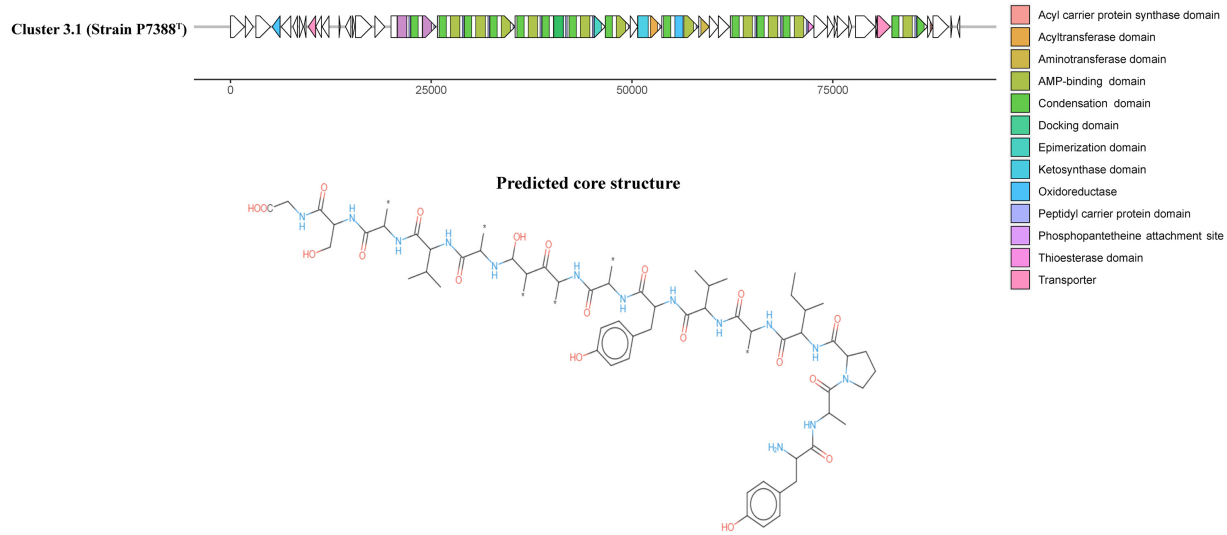


FIGURE 4 | A polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) hybrid BGC, Cluster 3.1, and the predicted core structure annotated by antiSMASH (Blin et al., 2021) from the genome of strain P7388^T. ORFs are depicted as white arrows and the core PKS-NRPS biosynthetic domains are color coded in each ORF.

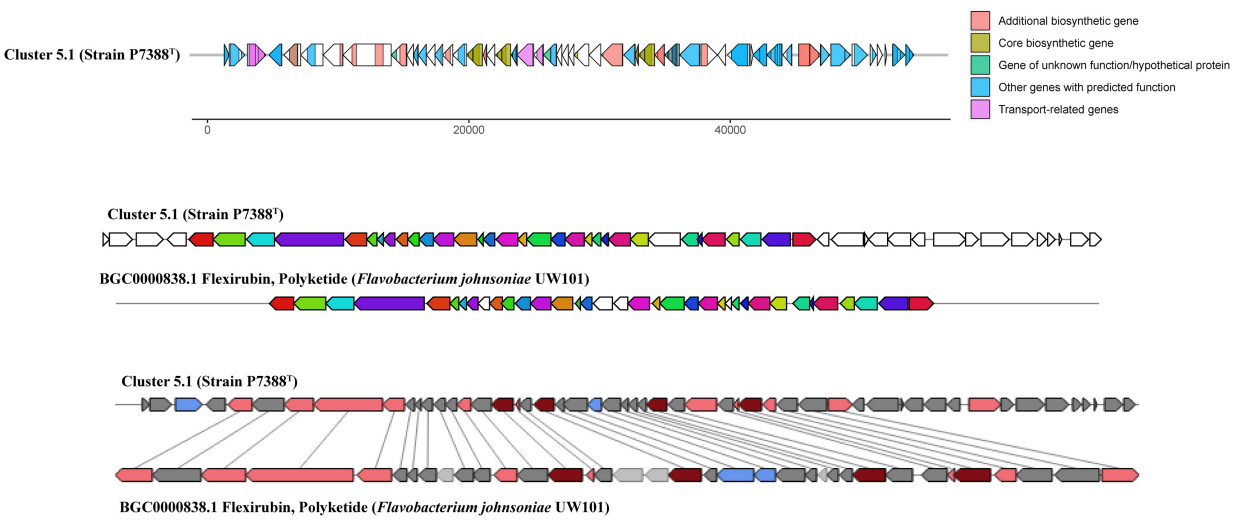


FIGURE 5 | Annotation of an arylpolyene/resorcinol BGC, Cluster 5.1, from strain P7388^T based on antiSMASH analysis and similarity of Cluster 5.1 to a PKS flexirubin biosynthetic gene cluster from *Flavobacterium johnsoniae* UW101 (BGC0000838.1).

of the cold-induced genes were harbored in all analyzed strains. One copy of a conserved *cspB* gene was found in all analyzed genomes and one copy of *cspC* gene was present in P7388^T, *F. saccharophilum* DSM 1811^T and *F. pectinovorum* DSM 6368^T. The *cspB* gene isolated from an arctic strain of *Polaribacter irgensii* (KOPRI 22228), a member of family *Flavobacteriaceae*, was found to substantially increase tolerance to freezing and thus considered a primary response to ensure freeze-tolerance of its hosts (Jung et al., 2018). Additionally, all strains encode a trigger factor Tgi, which has been suggested as a main cold-related chaperone of an Antarctic strain, *Pseudoalteromonas haloplanktis* TAC125 (Piette et al., 2010). Additional cold-inducible genes, such as putative cold-shock DEAD-box helicase A (*deaD*), chromosomal replication initiation ATPase (*dnaA*), DNA topoisomerases IV (*gyrA*, *gyrB*), transcription antitermination factor (*nusA*), translation factors IF-1, 2, 3 (*infA, infB, infC*) along with other cold-inducible genes were present in all genomes in one copy, regardless of the thermotype (**Supplementary Table 5**). Surprisingly, housekeeping genes associated with cold-adaptation, such as sigma factor 70 (*rpoD*) and sigma factor 24 (*rpoE*) were present in genomes only in one and two copies, respectively, although psychrotrophic *Flavobacterium* spp. were previously found to harbor multiple copies of these genes (Liu et al., 2019). Ice-binding (IBP) or antifreeze (ATF) proteins

inhibiting formation of ice crystals and their recrystallization were not found in either genome, although gene *ffIBP* encoding IBP was already detected in the genome of psychrophilic marine strain, *Flavobacterium frigoris* PS1 (Kim et al., 2019). As ice-nucleating activity was detected in flavobacteria strains found confined in ice cores (Christner et al., 2000; Wilson et al., 2006), it is likely that genes encoding IPBs and ATFs are present, however, these remain unassigned with hypothetical functions.

Cold-adaptation is inseparably associated with oxidative stress response as a result of higher oxygen solubility at lower temperatures leading to increased levels of reactive oxygen species (ROS). Notably, *Flavobacterium* spp. are well prepared to withstand oxidative stress by the production of proteins and enzymes involved in removal of ROS, such as catalases (KatA, KatG, KatE), superoxide-dismutase and peroxidase (SodA, SodC, Bcp), or thioredoxin and peroxiredoxin reductases (TrxA, TrxB, OsmC, OsmC-like proteins). Although all analyzed strains harbored peroxiredoxin genes in multiple copies (*osmC, osmC*-like), strain P4023^T harbored multiple copies of superoxide-dismutase (*sodA*) and thioredoxin (*trxA*), which may be explained by its isolation from surface rather than deeper soil layers and therefore exposed to higher oxidative stress and stronger UV radiation. Similarly versatile microorganisms to flavobacteria are pseudomonads that colonize diverse Antarctic areas including soils, fresh or marine waters (Reddy et al., 2004; Kosina et al., 2013; Jang et al., 2020). Considering they share the same environment, it is not surprising that pseudomonads also share certain anti-oxidative mechanisms with flavobacteria, particularly, various catalases, superoxide-dismutases, thioredoxins, or peroxiredoxins (Orellana-Saez et al., 2019). However, some mechanisms seem to be unique to pseudomonads, such as the expression of glutathione-related proteins and glyoxalase by *Pseudomonas* sp. MPC6 or the production of polyhydroxyalkanoates (PHAs) by *Pseudomonas extremaustralis* 14-3^T (Ayub et al., 2009; Tribelli et al., 2020). None of these anti-oxidative proteins have been found in the genome analysis of studied Antarctic strains, which suggests that cold-tolerant bacteria have developed several mechanisms to sustain oxidative stress, and these mechanisms can vary across bacterial taxa including those that co-inhabit the same extreme environment.

Genome analysis further revealed homologs to genes involved in carotenoid biosynthesis among all analyzed strains. Carotenoid pigments are associated with cold-adaptation as regulators of membrane fluidity, protectants against UV radiation and oxidative stress (Baraúna et al., 2017). Interestingly, only P4023^T harbored a *crtZ* gene required for zeaxanthin biosynthesis (Pasamontes et al., 1997), while the remaining *Flavobacterium* spp. lacked this gene, suggesting that their carotenoid biosynthetic pathway ends with β-carotene. Interestingly, while yellow coloration of strain P4023^T colonies seems to be a result of carotenoid pigments, strain P7388^T also harbors a biosynthetic gene cluster encoding flexirubin biosynthesis that may also contribute to its coloration. Phenotypic tests for the presence of flexirubin-type pigments correlated with predicted genomic features, resulting in a negative test result for P4023^T and positive result for P7388^T.

Proteorhodopsin (PR) is a membrane light-driven protein acting as an outward H^+ translocating pump which converts light energy to biochemical energy *via* translocation of protons through ATP-proton pumps (Yoshizawa et al., 2012). Both the analyzed psychrotrophic genomes in contrast to the analyzed mesophilic *Flavobacterium* spp. contained a PR-like protein whose activity has been found to enhance growth in some species along with a PR-associated *blh* gene, an essential cofactor of the PR-pump, that encodes 15,15′-β-carotene dioxygenase converting β-carotene to retinal (Peck et al., 2001). PRs were found widely spread among marine psychrophilic bacteria (McCarren and DeLong, 2007) and interestingly, studies on their growth stimulation effects are contradictory in their outcomes (Yoshizawa et al., 2012). Nevertheless, stimulation of growth by a PR-pump was proved in a marine strain belonging to the family *Flavobacteriaceae*, *Dokdonia* sp. MED134 (Kimura et al., 2011), which may suggest that a similar effect can be expected in Antarctic *Flavobacteria* spp., especially as none of these genes were found among the compared mesophilic species.

Presence of Prophages and CRISPRs

Both genomes of proposed type strains harbor a significant number of prophages, with half of them predicted as active by genomic prediction tools (**Supplementary Tables 6**, **7**). Interactions between phages and bacterial hosts lead to dynamic co-evolution of the microbial community rather than to the simplest outcome of predation, infection, and lysis of susceptible hosts. This interplay is far more complex and may result in higher adaptation of hosts to environmental factors in a specific niche (Casas and Maloy, 2018). Selective pressure provided by predator-prey relationships can increase overall fitness of a bacterial population by enhancing reproduction rate among infected bacteria (Shapiro et al., 2016), promoting exchange of chromosomal and plasmid DNA (Thierauf et al., 2009) or by lysogenic conversion (Casas and Maloy, 2018). Considering low temperatures in Antarctica negatively influence bacterial growth rate, a possible acceleration of reproduction rates caused by phage infection represents a notable advantage for the hosts. Genetic alterations caused by phages can directly change hosts' phenotype by acquisition of resistance genes (against antimicrobials or phages/protozoans) or induce metabolic changes (Arnold and Koudelka, 2014; Hurwitz et al., 2015; Gorter et al., 2016; Lee et al., 2017) and thus facilitate the ability of bacterial hosts to adapt and thrive in harsh Antarctic conditions. It is interesting that CRISPR/Cas-systems do not seem to be induced by phage infections in these psychrotrophic *Flavobacterium* strains, since the type II CRISPR/Cas-system was found to be an adaptive immune system in other cold-adapted flavobacteria (Liu et al., 2019). While only the P7388^T genome was predicted to contain one CRISPR array without *cas*-associated genes, the remaining analyzed genomes lack *cas*-associated genes as well as CRISPR arrays. This suggests that phage infection of these flavobacteria may be actually advantageous or if they developed defense mechanisms, these are likely associated with the surface receptors alterations, restriction/modification mechanisms or even immunity caused

by already present prophages rather than with the Cas-system system (Casas and Maloy, 2018).

Resistome

Although routine screening did not find any resistance genes, RGI prediction of the resistome from draft genomes of P4023^T and P7388^T suggested the presence of multidrug-resistance phenotypes. Ten putative antimicrobial resistance genes (ARGs) were predicted in the P4023^T genome likely to provide resistance against 16 different drug classes (**Supplementary Table 8**). Inactivation is the most abundant mechanism of action among predicted ARGs, followed by target alteration and efflux mechanisms. Although only seven putative ARGs were predicted from the P7388^T genome (**Supplementary Table 9**), they may be active against 20 different drug classes, with inactivation as the main mechanism of action. Furthermore, the predicted resistome did not fully correlate with *in vitro* antibiotic susceptibility testing. Correlation between resistome predictions and *in vitro* testing was found for resistance to aminoglycosides which is presumably encoded by *aadS* gene (MBP4138141.1) in the genome of P7388^T, a widely spread bacterial adenylyltransferase conferring resistance to aminoglycoside antibiotics (Cox et al., 2015). All Antarctic strains were further found to be resistant or borderline-susceptible to most β-lactams including cephalosporines (**Supplementary Table 10**). This wide resistance to β-lactam antibiotics may be induced by the gene *JOHN-1* present in P4023^T (MBP4142760.1) and P7388^T (MBP4137930.1) genomes. This gene provides resistance to a broad spectrum of β-lactam antibiotics in *Flavobacterium johnsoniae* (Naas et al., 2003), however, with a weak effect on susceptibility to carbapenems, which correlates with *in vitro* susceptibility results. Additive effects on resistance to β-lactam antibiotics may result from the presence of putative OXA beta-lactamase *OXA-29* (MBP4140625.1) in P4023^T, known to hydrolyze penams and cephalosporines but is not effective against carbapenems (Franceschini et al., 2001). Both strains were further predicted to be resistant against tetracyclines, fluoroquinolones, and phenicol; however, *in vitro* tests revealed susceptibility to tetracycline (30 μg), ciprofloxacin (5 μg) and chloramphenicol (30 μg). Disagreement between the phenotypic susceptibility pattern and ARGs computational prediction is a well-known controversy, especially in clinical bacteriology (Hughes and Andersson, 2017). There are several explanations for the phenomenon of phenotype-genotype dissociation which include environmental modulation of gene expression (*in vivo* vs. *in vitro*), testing of single strain culture vs. biofilm formation, presence of specific metabolites and growth factors, presence of specific phages or inadequate time/concentration of tested antibiotics that induce expression of resistance genes. Any of above-mentioned factors or even their combination may be also responsible for discrepancies observed among analyzed Antarctic *Flavobacterium* isolates, in particular considering extreme environment they colonize and presence of phages in their genomes.

Strain P4023^T additionally harbors a putative *rosA* gene (MBP4140897.1) encoding an efflux pump/potassium antiporter system. This efflux transporter has been detected in marine *Flavobacterium* spp. (Hao et al., 2018) and within a metagenomic study of Antarctic pristine soils which included *Bacteroidetes* (Van Goethem et al., 2018). The *rosA* gene is related to polymyxin B resistance and represents a part of a two component efflux antiporter system (RosAB) found in *Yersinia enterocolitica* and confers resistance to cationic antimicrobial peptides including polymyxin B (Bengoechea and Skurnik, 2000). So far, no *Flavobacterium* spp. has been detected that could encode a fully functional RosAB system and no data exists to support the effectiveness of this efflux pump without both components. However, all isolated Antarctic strains were resistant or borderline susceptible to polymyxin B suggesting that these strains harbor some unrevealed mechanisms conferring polymyxin B resistance. An environmental study focused on pristine Arctic soils and ancient ARGs revealed that ARGs of soil microbiota are predominantly associated with aminoglycosides resistance and multidrug defense systems including efflux pumps (McCann et al., 2019). Another study focused on Antarctic soils in pristine areas and presence of ancient ARGs clearly showed a high abundance of ARGs which reflects presence of natural antibiotics (Van Goethem et al., 2018). Out of these, efflux pumps were shown as a common trait in a majority of soil microbiota colonizing pristine Antarctic areas which they developed as an adaptive response to diverse chemical stressors (Van Goethem et al., 2018). These findings imply that the polymyxin B resistance of Antarctic *Flavobacterium* spp. may be related to an efflux system, but it can not be concluded at this time that this system is in fact associated with the predicted *rosA* gene.

Further comparison of antibiotic susceptibility profiles of the type strains to the closest related strains showed that they similarly express multidrug resistant phenotypes with susceptibility only to ciprofloxacin, imipenem, co-trimoxazole, and tetracycline. Such results are not surprising as these reference strains are associated with aquaculture (originally isolated from water or fish hosts), which is a well-known environment with resistant bacteria (Cabello et al., 2016). Although the genus *Flavobacterium* comprises numerous species isolated from pristine areas, or various abiotic sources, data on antibiotic susceptibility of environmental isolates are rather scarce. Furthermore, the methods used to determine these antibiotic resistance profiles significantly differs between studies making accurate comparisons challenging for species in this genus. For the moment, the question remains unanswered whether a majority of all *Flavobacterium* spp. express multidrug-resistance phenotypes or if this pattern is simply unique to certain species inhabiting specific niches.

Morphological, Physiological, and Biochemical Characteristic

Isolates from both groups represented by type strains P4023^T and P7388^T were aerobic, Gram-negative rods with rounded ends, with average cell size 0.4–0.6 μm × 1.2–2.4 μm and 0.3–0.4 μm × 1.5–3.0 μm, respectively (**Supplementary Figures 4**, **5**). All strains formed yellowish and dark yellow to orange-colored colonies often expressing gliding motility and opalescence on the R2A medium. All strains were catalase

positive and oxidase was congruently positive only for strains from the P4023^T group. Isolates from both groups can be characterized and distinguished by physiological and biochemical tests listed in **Table 2** along with tests useful for their separation from their closest phylogenetic neighbors. Characteristics specific for both groups are given in the description part as well as in a comprehensive formal format in **Supplementary Table 11**.

Chemotaxonomic Analyses

The analysis of fatty acid methyl esters showed that both groups of Antarctic isolates differ from each other in their major fatty acids (**Table 3**). The triad comprising strain P4023^T revealed five major fatty acids, specifically $C_{15:1}$ ω*6c* (18.1%), Summed Feature 3 ($C_{16:1}$ ω*7c*/$C_{16:1}$ ω*6c*) (9.9%), anteiso-$C_{15:0}$ (9.8%), iso-$C_{15:0}$ (9.3%) and iso-$C_{16:0}$ 3OH (9.3%). The major fatty acid of the group comprising strain P7388^T were iso-$C_{15:0}$ (23.4%), Summed Feature 3 ($C_{16:1}$ ω*7c*/$C_{16:1}$ ω*6c*) (9.9%), iso-$C_{17:0}$ 3OH (8.5%) and iso-$C_{15:0}$ 3OH (8.4%). The proposed species differ not only qualitatively and quantitatively in the profiles of major fatty acids, but also quantitatively in minor fatty acids, especially iso-$C_{14:0}$, iso-$C_{16:1}$ H and Summed Feature 9 ($C_{16:0}$ 10-methyl/iso-$C_{17:1}$). The overall fatty acids profile of both groups is in agreement with the genus description (Bernardet and Bowman, 2015) and it is similar to those of the reference strains with mostly quantitative differences. A significant difference is the higher presence of $C_{15:1}$ ω*6c* detected in the P4023^T triad that enables its clear separation from the compared strains.

The quinone system of strain P4023^T was composed of 94.3% menaquinone MK-6, 5.4% MK-5 and 0.2% MK-7, and that of strain P7388^T was 90.2% MK-6, 9.6% MK-5 and 0.2% MK-7. The polar lipid profile of strain P7388^T (**Figure 6A**) consisted of the major lipids, phosphatidylethanolamine, an ornithine lipid, two unidentified lipids lacking a functional group (L3, L4), moderate amounts of unidentified lipid L1, minor amounts of lipids L2 and L6 and an unidentified glycolipid (GL). Strain P4023^T (**Figure 6B**) displayed major amounts of phosphatidylethanolamine, an ornithine lipid, and two unidentified lipids (L3, L4) lacking a functional group, moderate amounts of unidentified lipid L1 and glycolipid GL and minor amounts of an unidentified aminophospholipid (APL), and two unidentified lipids (L2, L5). Additionally, a yellow pigment spot was visible (yPigm). The polyamine pattern of strain P4023^T consisted of the major polyamine *sym*-homospermidine [27.3 μmol (g dry weight^{-1})] and minor amounts of putrescine [0.7 μmol (g dry weight^{-1})] and spermidine [0.2 μmol (g dry weight^{-1})]. The polyamine pattern of strain P7388^T was similar with the major polyamine sym-homospermidine [20.5 μmol (g dry weight^{-1})] and minor amounts of putrescine [0.2 μmol (g dry weight^{-1})] and spermidine [0.3 μmol (g dry weight^{-1})]. Quinone systems, polar lipid profile and polyamine pattern are well in agreement with the genus description (Bernardet et al., 1996; Dong et al., 2013). Both strains differed in their polar lipid profiles from their close relatives *F. hercynium* KACC 14934^T and *F. resistens* KACC 14246^T (Kim et al., 2012). Specifically, the absence of a lipid labeled PDE (explanation not provided) was noted in the related strains, whereas in the novel strains we detected an aminolipid (OL) with the same chromatographic motility as PDE and corresponded to ornithine lipid 2 (Kawai et al., 1988). The presence of an ornithine lipid or an unidentified aminolipid with the same chromatographic motility as the ornithine lipid in **Figure 6** has been reported for numerous *Flavobacterium* species including the type species of the genus *Flavobacterium aquatile* (Kämpfer et al., 2012, 2015, 2020; Lee et al., 2012).

CONCLUSION

Eight *Flavobacterium* strains isolated from James Ross Island were analyzed in this study. Using a polyphasic taxonomic approach by combing phenotypic, chemotaxonomic and phylogenomic characteristics, these eight strains forming separated phylogenetic lines were classified as two novel species for which the names *Flavobacterium flabelliforme* sp. nov. and *Flavobacterium geliluteum* sp. nov. are proposed. Comparative genomics analysis showed that the genomes of these Antarctic strains encode numerous genes related to adaptation to cold temperature which is likely a major factor influencing growth in harsh Antarctic conditions. Adaptation mechanisms of Antarctic strains were related to production of pigments, proteorhodopsin, cold-shock proteins, oxidative and osmotic stress, gene expression and membrane fluidity. Although these novel Antarctic species were obtained from an isolated and pristine environment of James Ross Island, they were resistant against a broad spectrum of antimicrobials and predicted to harbor putative resistance genes for a wider spectrum of antimicrobial drugs. This is of particular importance as metagenomic studies have implied that Antarctic microbiota represent a source of ancient antibiotic resistance genes, but so far multidrug resistant isolates were only found in areas with high anthropogenic influence such as King George Island or seawater. The findings in this study thereby makes strains P4023^T and P7388^T a promising source for future studies on antibiotic resistance genes and resistance mechanisms from environmental microorganisms. Genomic analysis also showed a surprisingly high number of prophage sequences that may influence adaptive response to environmental stresses. In addition, genomes of type strains were predicted to harbor biosynthetic gene clusters with a majority not matching any known BGCs in current databases of biosynthetic genes. These findings suggest that these novel Antarctic psychrophiles are not only interesting taxonomically, but also a promising object of further studies in order to elucidate the effect of phages on their adaptation, the function of their biosynthetic genes, and their antibiotic resistance mechanisms.

TAXONOMY

Description of *Flavobacterium flabelliforme* sp. nov.

Flavobacterium flabelliforme (fla.bel.li.for'me. L. n. *flabellum*, fan, vane; L. suff. *-formis*, -e, -like, in the shape of; N.L. neut. adj. *flabelliforme*, fan-like shaped, referring to fan-like shaped colonies).

TABLE 3 | Cellular fatty acids composition (%) of *F. flabelliforme* sp. nov., *F. geliluteum* sp. nov. and their closest phylogenetically related *Flavobacterium* spp.

	***F. flabelliforme* sp. nov.**			***F. geliluteum* sp. nov.**										
Fatty acid	**P4023^T**	**CCM 9063**	**P4911**	**P7388^T**	**CCM 9065**	**P7381**	**P7475**	**P9670**	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**
iso-$C_{14:0}$	3.3	3.0	3.3	1.0	1.0	1.0	1.0	1.0	1.0	1.1	TR	TR	1.0	TR
iso-$C_{15:1}$ G	6.5	4.8	5.8	6.9	9.3	6.6	8.9	9.5	8.0	8.2	7.6	8.9	5.6	7.7
anteiso-$C_{15:1}$ A	1.2	1.7	1.2	TR	TR	TR	TR	TR	TR	TR	TR	TR	TR	TR
iso-$C_{15:0}$	10.1	7.9	8.8	24.8	22.8	24.1	22.5	22.6	24.7	16.4	26.0	28.4	11.8	20.2
anteiso-$C_{15:0}$	8.6	11.6	9.1	6.4	5.5	5.8	5.4	6.1	8.3	6.0	3.5	5.2	5.6	8.2
$C_{15:1}$ *ω6c*	20.5	16.8	17.1	7.4	6.9	7.8	7.1	7.0	4.4	6.5	3.7	4.1	5.5	3.6
iso-$C_{16:1}$ H	4.7	6.8	6.3	1.7	1.7	1.6	1.5	1.8	TR	1.5	TR	TR	2.0	1.0
iso-$C_{16:0}$	1.8	3.1	3.8	1.8	2.3	1.7	2.0	2.5	1.7	2.1	1.7	1.9	4.0	2.0
$C_{16:0}$	TR	TR	TR	TR	TR	TR	TR	TR	1.6	2.0	2.4	1.9	3.8	1.6
iso-$C_{15:0}$ 3OH	5.4	4.5	4.3	9.1	7.9	8.9	8.1	8.2	7.8	9.0	8.8	7.3	5.6	8.1
$C_{15:0}$ 2OH	TR	TR	TR	TR	TR	TR	TR	TR	TR	1.1	TR	TR	1.4	1.1
$C_{15:0}$ 3OH	3.0	3.0	2.8	ND	ND	2.1	2.3	ND	2.0	2.7	2.1	ND	ND	2.5
$C_{17:1}$ *ω8c*	TR	TR	TR	TR	TR	TR	TR	TR	TR	1.0	TR	TR	TR	TR
$C_{17:1}$ *ω6c*	7.0	5.0	6.3	4.5	4.8	4.8	4.4	4.2	2.4	5.3	3.6	4.5	4.6	4.0
iso-$C_{16:0}$ 3OH	9.1	9.6	9.3	9.1	6.2	5.4	5.6	5.8	3.7	5.1	4.0	3.1	6.1	5.3
$C_{16:0}$ 3OH	1.2	1.2	1.5	1.5	1.4	1.4	1.3	1.1	3.2	2.4	3.9	2.1	2.6	2.0
iso-$C_{17:0}$ 3OH	3.6	3.4	3.6	8.3	9.2	8.3	8.6	8.3	8.8	9.3	8.4	8.8	7.7	9.6
$C_{17:0}$ 2OH	TR	TR	TR	TR	1.0	TR	1.0	1.0	1.3	1.0	TR	TR	1.6	2.4
Summed Feature 3[a]	7.6	9.7	10.0	10.1	9.6	9.9	10.1	10.0	12.2	11.2	13.9	12.0	21.6	11.6
Summed Feature 9[b]	1.4	2.1	1.9	5.3	5.3	5.2	5.1	5.3	4.1	4.8	3.4	5.3	5.4	4.5

[a] $C_{16:1}$ *ω7c*/$C_{16:1}$ *ω6c*.
[b] $C_{16:0}$ *10-methyl/iso-*$C_{17:1}$ *ω9c*.
ND, not detected; TR, trace amounts (<1%); major fatty acids of proposed species are highlighted in blue.
Strains: 1, F. hercynium CCM 9054^T; 2, F. branchiicola CCM 9061^T; 3, F. chilense CCM 7940^T; 4, F. araucananum CCM 7939^T; 5, F. saccharophilum CCM 8770^T; 6, F. psychroterrae CCM 8827^T.

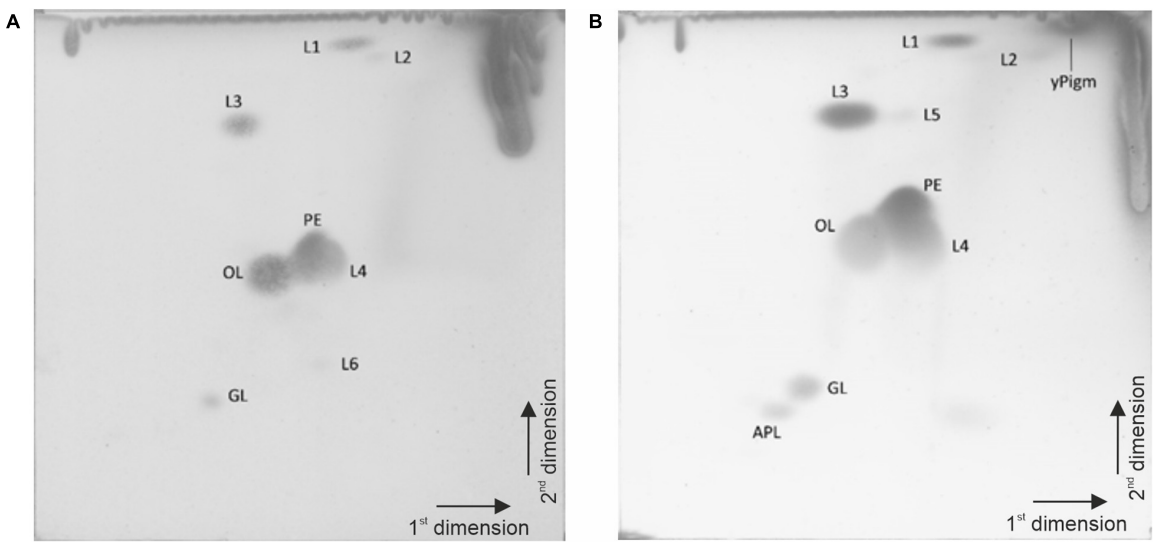


FIGURE 6 | Polar lipid profiles of strains P7388^T **(A)** and P4023^T **(B)** after two-dimensional thin layer chromatography at detection with 5% ethanolic molybdatophosphoric acid and development at 140°C. PE, phosphatidylethanolamine; OL, ornithine lipid; APL unidentified amino lipid; GL, unidentified glycolipid; L1-5, unidentified lipids lacking a functional group.

Cells are Gram-stain-negative rods with rounded ends, cell size in range 0.4–0.6 μm × 1.2–2.4 μm, cells occurring in irregular clusters, occasionally singly or in pairs. Endospores are not formed. Does not produce a capsule. Negative for presence of flexirubin-type pigments. Does not adhere to agar. Yellowish colonies. Motile with gliding-activity. Catalase and

oxidase positive. Growth occurs on R2A, PCA, TSA, marine agar NA, blood agar with 5% sheep blood, BHI, Mueller-Hinton and Endo agar. Does not grow on MacConkey agar. Grows in pH range 6–9 and temperature range 1–30°C with optimum growth at 20°C and pH around 7. Cells grow well in presence of 1% NaCl, and 2% NaCl inhibits growth. Does grow in microaerophilic conditions but growth in anoxic conditions is limited. Does not produce fluorescein on King B medium. No utilization of Simmon's citrate, malonate and acetamide. Negative for reduction of nitrates and nitrites. Negative for production of urease and indole. Positive for hydrolysis of gelatine, casein, and tyrosine. Does not produce brown diffusible pigment on L-tyrosine agar. Negative for hydrolysis of Tween 80, aesculin, ONPG, starch, DNA, CMC and agar. Does not produce lecithinase. Does not produce H_2S. Positive for arginine dihydrolase, and negative for ornithine and lysine decarboxylases. Does produce acid from glucose and maltose in aerobic conditions. Negative for production of acid from fructose and xylose in aerobic conditions. Positive for utilization of glucose and maltose by API 20 NE. Negative for utilization of arabinose, mannose, *N*-acetylglucosamine, gluconic acid, capric acid, adipic acid, malic acid, citric acid and phenylacetic acid by API 20 NE. Positive for alkaline phosphatase, leucine arylamidase, valine arylamidase and acid phosphatase by API ZYM. Negative for esterase (C 4), esterase lipase (C 8), lipase (C 14), cystine arylamidase, trypsin, α-chymotrypsin, naphtol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, *N*-acetyl-β-glucosaminidase, α-mannosidase, and α-fucosidase by API ZYM. Carbon source utilization ability *via* respiration determined by Biolog GEN III MicroPlate test panels is positive for D-maltose, α-D-glucose, D-glucose-6-PO_4, gelatine, L-arginine, L-aspartic acid, L-glutamic acid, acetoacetic acid and acetic acid. Negative for D-trehalose, D-cellobiose, gentiobiose, sucrose, stachyose, D-raffinose, α-D-lactose, D-melibiose, β-methyl-D-glucoside, D-salicin, *N*-acetyl-β-D-mannosamine, *N*-acetyl neuraminic acid, D-mannose, D-fructose, D-galactose, 3-methyl glucose, D-fucose, L-fucose, L-rhamnose, inosine, D-sorbitol, D-mannitol, D-arabitol, myo-inositol, glycerol, D-fructose-6-PO_4, D-aspartic acid, D-serine, L-alanine, L-histidine, L-pyroglutamic acid, L-serine, D-galacturonic acid, D-galactonic acid lactone, D-gluconic acid, D-glucuronic acid, glucuronamide, mucic acid, quinic acid, D-saccharic acid, p-hydroxy phenylacetic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, citric acid, α-keto glutaric acid, D-malic acid, L-malic acid, bromo-succinic acid, Tween 40, γ-amino-butyric acid, α-hydroxy-butyric acid, β-hydroxy-D,L-butyric acid, α-keto butyric acid, propionic acid and formic acid.

The major fatty acids are iso-$C_{15:1}$ ω*6c*, Summed Feature 3 ($C_{16:1}$ ω*7c*/ $C_{16:1}$ ω*6c*), anteiso-$C_{15:0}$, iso-$C_{15:0}$ and iso-$C_{16:0}$ 3OH. Major respiratory quinone is MK-6 and major polyamine is *sym*-homospermidine. Polar lipid profile contains major amounts of phosphatidylethanolamine, an ornithine lipid, and two unidentified lipids (L3, L4) lacking a functional group, moderate amounts of unidentified lipid L1, unidentified glycolipid GL, minor amounts of an unidentified aminophospholipid (APL), and two unidentified lipids (L2, L5). The DNA G + C content of the type strain is 31.2 mol%.

Type strain P4023^T (= CCM 9062^T = LMG 31963^T) was isolated in 2011 from the organic material of an abandoned bird nest located at the Lachman Cape (GPS: −63.778333 S, −57.781666 W). All characteristics listed in the species description are shared by all strains, except for the following strain-dependent test results: motility, production of acid from mannitol, and variable results observed by Biolog GEN III MicroPlate were utilization of dextrin, D-turanose, *N*-acetyl-D-glucosamine, glycyl-L-proline and pectin. A formal proposal of the species "*Flavobacterium flabelliforme* sp. nov." is given in **Supplementary Table 11**.

The GenBank/EMBL/DDBJ accession numbers for the near full length 16S rRNA gene sequences and whole genome sequences for *Flavobacterium flabelliforme* sp. nov. P4023^T are MW691162 and JAGFBU000000000, respectively.

Description of *Flavobacterium geliluteum* sp. nov.

Flavobacterium geliluteum (ge.li.lu'te.um. L. neut. n. *gelum*, cold, frost; L. adj. *luteus*, yellow; N.L. neut. n. *geliluteum*, forming yellow colonies in the cold).

Cells are Gram-stain-negative rods with rounded ends, cell size in range 0.3–0.4 μm × 1.5–3.0 μm, occurring singly and in irregular clusters, occasionally in pairs. Endospores are not formed. Does not produce a capsule. Dark yellow to orange pigmented colonies. Flexirubin-type pigment present. Does not adhere to agar. Motile with gliding-activity. Catalase positive. Oxidase negative. Growth occurs on R2A, PCA, TSA, NA, blood agar with 5% sheep blood, BHI and Mueller-Hinton agar. Does not grow on marine and MacConkey agar. Grows in pH range 6–8 and temperature range 15–30°C with optimum growth at 20°C and pH around 7. Cells grow in presence of 0.5% NaCl, and 1% NaCl inhibits growth. Does grow in microaerophilic conditions but growth in anoxic conditions is limited. Does not produce fluorescein on King B medium. No utilization of Simmon's citrate, malonate and acetamide. Negative for reduction of nitrates and nitrites. Negative for production of urease and indole. Positive for hydrolysis of gelatine, aesculin, ONPG, starch, casein, tyrosine and CMC. Does not produce brown diffusible pigment on L-tyrosine agar. Negative for hydrolysis of DNA and agar. Does not produce lecithinase. Does not produce H_2S. Positive for arginine dihydrolase and negative for ornithine and lysine decarboxylases. Does produce acid from glucose, maltose and xylose in aerobic conditions. Negative for production of acid from mannitol in aerobic conditions. Positive for utilization of glucose, arabinose, mannose, *N*-acetyl-glucosamine, maltose, and hydrolysis of aesculin by API 20 NE. Negative for utilization of gluconic acid, capric acid, adipic acid, malic acid, citric acid and phenylacetic acid by API 20 NE. Positive for alkaline phosphatase, leucine arylamidase, acid phosphatase, naphtol-AS-BI-phosphohydrolase and β-glucosidase by API ZYM. Negative for esterase (C 4), esterase lipase (C 8), lipase (C 14), valine arylamidase, cystine arylamidase, trypsin, α-chymotrypsin, α-galactosidase,

β-galactosidase, β-glucuronidase, *N*-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase by API ZYM. Carbon source utilization ability *via* respiration determined by Biolog GEN III MicroPlate test panels is positive for D-trehalose, D-cellobiose, gentiobiose, *N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine, α-D-glucose, D-mannose, D-glucose-6-PO_4, glycyl-L-proline, L-aspartic acid, L-glutamic acid, D-galacturonic acid, acetoacetic acid and acetic acid. Negative for sucrose, stachyose, D-raffinose, α-D-lactose, D-melibiose, *N*-acetyl-β-D-mannosamine, *N*-acetyl neuraminic acid, 3-methyl glucose, D-fucose, L-fucose, L-rhamnose, inosine, D-sorbitol, D-mannitol, D-arabitol, myo-inositol, glycerol, D-aspartic acid, D-serine, L-alanine, D-gluconic acid, D-glucuronic acid, glucuronamide, D-saccharic acid, quinic acid, p-hydroxy phenylacetic acid, L-lactic acid, α-keto glutaric acid, D-malic acid, L-malic acid, bromo-succinic acid, γ-amino-butyric acid, α-hydroxy-butyric acid, β-hydroxy-D,L-butyric acid, α-keto butyric acid, propionic acid, and formic acid.

The major fatty acids are iso-$C_{15:0}$, Summed Feature 3 ($C_{16:1}$ ω*7c*/$C_{16:1}$ ω*6c*), iso-$C_{17:0}$ 3OH and iso-$C_{15:0}$ 3OH. Major respiratory quinone is menaquinone MK-6 and major polyamine is *sym*-homospermidine. Polar lipid profile contains major lipids phosphatidylethanolamine, an ornithine lipid, two unidentified lipids lacking a functional group (L3, L4), moderate amounts of unidentified lipid L1, minor amounts of lipids L2 and L6 and an unidentified glycolipid (GL). The DNA G + C content of the type strain is 34.5 mol%.

Type strain P7388^T (= CCM 9064^T = LMG 31962^T) was isolated from water samples taken in 2016 from a small temporary lake (GPS: −63.795894 S, −57.809928 W). All characteristics listed in the species description are shared by all strains, except for the following strain-dependent test results: growth at 5 and 10°C, growth on Endo agar, hydrolysis of Tween 80, production of acid from fructose, presence of α-glucosidase in API ZYM and variable results observed by Biolog GEN III MicroPlate were utilization of dextrin, D-maltose, pectin, β-methyl-D-glucoside, D-salicin, D-fructose, D-galactose, D-fructose-6-PO_4, gelatine, L-arginine, D-galactonic acid lactone, mucic acid, methyl pyruvate, D-lactic acid methyl ester, citric acid and Tween 40. A formal proposal of the species "*Flavobacterium geliluteum* sp. nov." is given in **Supplementary Table 11**.

The GenBank/EMBL/DDBJ accession numbers for the near full length 16S rRNA gene sequences and whole genome sequences for *Flavobacterium geliluteum* sp. nov. P7388^T are MW691150 and JAGFBV000000000, respectively.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW691162; https://www.ncbi.nlm.nih.gov/genbank/, JAGFBU000000000; https://www.ncbi.nlm.nih.gov/genbank/, MW691150; https://www.ncbi.nlm.nih.gov/genbank/, JAGFBV000000000.

AUTHOR CONTRIBUTIONS

SK performed phylogenetic, phylogenomic, genomic analysis, analysis of fatty acid-methyl esters, and drafted and finalized the manuscript. H-JB performed chemotaxonomic analyses of polar lipids, menaquinones, and polyamines. MB was responsible for whole-genome sequencing. MS-P was involved in genomic analysis of biosynthetic potential. MN was involved in genome assembly and quality of WGS data. DK performed electron microscopy. ES and IS performed morphological, physiological, and biochemical characterization including antibiotic susceptibility. SK drafted the manuscript with inputs of H-JB, MB, MS-P, and IS. All authors edited the draft manuscript and agreed to the final manuscript version for submission.

FUNDING

This work was supported by the Ministry of Education, Youth and Sports of the Czechia (LM2015078) Czech-Polar 2 Czech Polar Research Infrastructure.

ACKNOWLEDGMENTS

The authors would like to thank the scientific infrastructure of the J. G. Mendel Czech Antarctic Station, part of the Czech Polar Research Infrastructure (CzechPolar2), and its crew for their assistance and the Czech Antarctic Foundation for their support. The authors would also like to thank Bernard Schink (University of Konstanz, Germany) for name corrections. The authors would also further like to thank for excellent technical assistance provided by Jana Bajerová and Renée Šubrtová.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.729977/full#supplementary-material

REFERENCES

Aislabie, J. M., Jordan, S., and Barker, G. M. (2008). Relation between soil classification and bacterial diversity in soils of the Ross Sea region. *Antarctica. Geoderma* 144, 9–20. doi: 10.1016/j.geoderma.2007.10.006

Ali, Z., Cousin, S., Frühling, A., Brambilla, E., Schumann, P., Yang, Y., et al. (2009). *Flavobacterium rivuli* sp. nov., *Flavobacterium subsaxonicum* sp. nov., *Flavobacterium swingsii* sp. nov. and *Flavobacterium reichenbachii* sp. nov., isolated from a hard water rivulet. *Int. J. Sys. Evol. Microbiol.* 59, 2610–2617. doi: 10.1099/ijs.0.008771-8770

Altenburger, P., Kämpfer, P., Makristathisc, A., Werner, L., and Busse, H.-J. (1996). Classification of bacteria isolated from a medieval wall painting. *J. Biotechnol.* 47, 39–52.

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

Arnold, J. W., and Koudelka, G. B. (2014). The trojan horse of the microbiological arms race: phage-encoded toxins as a defence against eukaryotic predators. *Environ. Microbiol.* 16, 454–466. doi: 10.1111/1462-2920.12232

Atlas, R. M. (2010). *Handbook of Microbiological Media*, 4th Edn. Washington, D.C: ASM Press.

Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., et al. (2021). dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res.* 49, D523–D528. doi: 10.1093/nar/gkaa742

Ayub, N. D., Tribelli, P. M., and López, N. I. (2009). Polyhydroxyalkanoates are essential for maintenance of redox state in the Antarctic bacterium *Pseudomonas sp*. 14-3 during low temperature adaptation. *Extremophiles* 13, 59–66. doi: 10.1007/s00792-008-0197-z

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

Baraúna, R. A., Freitas, D. Y., Pinheiro, J. C., Folador, A. R. C., and Silva, A. (2017). A proteomic perspective on the bacterial adaptation to cold: integrating OMICs data of the psychrotrophic bacterium *Exiguobacterium antarcticum* B7. *Proteomes* 5:9. doi: 10.3390/proteomes5010009

Barria, C., Malecki, M., and Arraiano, C. M. Y. (2013). Bacterial adaptation to cold. *Microbiology* 159, 2437–2443. doi: 10.1099/mic.0.052209-52200

Barrow, G. I., and Feltham, R. K. A. (eds) (1993). *Cowan and Steel's Manual for the Identification of Medical Bacteria*, 3rd Edn. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511527104

Benaud, N., Edwards, R. J., Amos, T. G., D'Agostino, P. M., Gutiérrez-Chávez, C., Montgomery, K., et al. (2020). Antarctic desert soil bacteria exhibit high novel natural product potential, evaluated through long-read genome sequencing and comparative genomics. *Environ. Microbiol.* 23, 3646–3664. doi: 10.1111/1462-2920.15300

Bengoechea, J. A., and Skurnik, M. (2000). Temperature-regulated efflux pump/potassium antiporter system mediates resistance to cationic antimicrobial peptides in Yersinia. *Mol. Microbiol.* 37, 67–80. doi: 10.1046/j.1365-2958.2000.01956.x

Bernardet, J.-F., and Bowman, J. P. (2015). "Flavobacterium," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds N. R. Krieg, J. T. Staley, D. R. Brown, B. P. Hedlund, B. J. Paster, N. L. Ward, et al. (Hoboken, NJ: John Wiley & Sons, Ltd), doi: 10.1002/9781118960608.gbm00312

Bernardet, J.-F., Nakagawa, Y., Holmes, B., and Subcommittee on the taxonomy of Flavobacterium and Cytophaga-like bacteria of the International Committee on Systematics of Prokaryotes (2002). Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int. J. Syst. Evol. Microbiol.* 52, 1049–1070. doi: 10.1099/00207713-52-3-1049

Bernardet, J.-F., Segers, P., Vancanneyt, M., Berthe, F., Kersters, K., and Vandamme, P. (1996). Cutting a gordian knot: emended classification and description of the genus flavobacterium, emended description of the family flavobacteriaceae, and proposal of *Flavobacterium hydatis* nom. nov. (Basonym, Cytophaga aquatilis Strohl and Tait 1978). *Int. J. Sys. Evol. Microbiol.* 46, 128–148. doi: 10.1099/00207713-46-1-128

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607

Biswas, A., Staals, R. H. J., Morales, S. E., Fineran, P. C., and Brown, C. M. (2016). CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics* 17:356. doi: 10.1186/s12864-016-2627-2620

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi: 10.1093/nar/gkab335

Boetius, A., Anesio, A. M., Deming, J. W., Mikucki, J. A., and Rapp, J. Z. (2015). Microbial ecology of the cryosphere: sea ice and glacial habitats. *Nat. Rev. Microbiol.* 13, 677–690. doi: 10.1038/nrmicro3522

Brooks, K., and Sodeman, T. (1974). A rapid method for determining decarboxylase and dihydrolase activity. *J. Clin. Pathol.* 27, 148–152.

Busse, H.-J., Bunka, S., Hensel, A., and Lubitz, W. (1997). Discrimination of members of the family pasteurellaceae based on polyamine patterns. *Int. J. Sys. Evol. Microbiol.* 47, 698–708. doi: 10.1099/00207713-47-3-698

Busse, J., and Auling, G. (1988). Polyamine pattern as a chemotaxonomic marker within the *Proteobacteria*. *Syst. Appl. Microbiol.* 11, 1–8. doi: 10.1016/S0723-2020(88)80040-80047

Cabello, F. C., Godfrey, H. P., Buschmann, A. H., and Dölz, H. J. (2016). Aquaculture as yet another environmental gateway to the development and globalisation of antimicrobial resistance. *Lancet Infect. Dis.* 16, e127–e133. doi: 10.1016/S1473-3099(16)00100-106

Casas, V., and Maloy, S. (2018). "The role of phage in the adaptation of bacteria to new environmental niches," in *Molecular Mechanisms of Microbial Evolution Grand Challenges in Biology and Biotechnology*, ed. P. H. Rampelotto (Cham: Springer International Publishing), 267–306. doi: 10.1007/978-3-319-69078-0_11

Cavicchioli, R., Charlton, T., Ertan, H., Mohd Omar, S., Siddiqui, K. S., and Williams, T. J. (2011). Biotechnological uses of enzymes from psychrophiles. *Microb. Biotechnol.* 4, 449–460. doi: 10.1111/j.1751-7915.2011.00258.x

Christensen, W. B. (1946). Urea decomposition as a means of differentiating proteus and paracolon cultures from each other and from *Salmonella* and *Shigella* Types 1. *J. Bacteriol.* 52, 461–466.

Christner, B. C., Mosley-Thompson, E., Thompson, L. G., Zagorodnov, V., Sandman, K., and Reeve, J. N. (2000). Recovery and identification of viable bacteria immured in glacial ice. *Icarus* 144, 479–485. doi: 10.1006/icar.1999.6288

Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., et al. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int. J. Syst. Evol. Microbiol.* 68, 2386–2392. doi: 10.1099/ijsem.0.002809

Clarridge, J. E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* 17, 840–862. doi: 10.1128/CMR.17.4.840-862.2004

CLSI (2015). *Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fifth Informational Supplement (M100-S25)*. Wayne, PA: Clinical and Laboratory Standards Institute.

Cox, G., Stogios, P. J., Savchenko, A., and Wright, G. D. (2015). Structural and molecular basis for resistance to aminoglycoside antibiotics by the adenylyltransferase ANT(2″)-Ia. *mBio* 6:e02180-14. doi: 10.1128/mBio.02180-2114

Dong, K., Chen, F., Du, Y., and Wang, G. (2013). Flavobacterium enshiense sp. nov., isolated from soil, and emended descriptions of the genus *Flavobacterium* and *Flavobacterium cauense*, *Flavobacterium saliperosum* and *Flavobacterium suncheonense*. *Int. J. Syst. Evol. Microbiol.* 63, 886–892. doi: 10.1099/ijs.0.039974-39970

Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., et al. (2020). MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* 48, D561–D569. doi: 10.1093/nar/gkz1010

Edwards, U., Rogall, T., Blöcker, H., Emde, M., and Böttger, E. C. (1989). Isolation and direct complete nucleotide determination of entire genes. characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* 17, 7843–7853.

EUCAST (2017). *Breakpoint Tables for Interpretation of MICs and Zone Diameters, Version 7.1*. Sweden: The European Committee on Antimicrobial Susceptibility Testing.

Ewing, W. H. (1960). *Enterobacteriaceae. Biochemical Methods for Group Differentiation. Public Health Service Publication No. 734*. Atlanta, ATL: CDC.

Fox, G. E., Wisotzkey, J. D., and Jurtshuk, P. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* 42, 166–170. doi: 10.1099/00207713-42-1-166

Franceschini, N., Boschi, L., Pollini, S., Herman, R., Perilli, M., Galleni, M., et al. (2001). Characterization of OXA-29 from Legionella (Fluoribacter) gormanii:

molecular class D β-lactamase with unusual properties. *Antimicrob. Agents Chemother* 45, 3509–3516. doi: 10.1128/AAC.45.12.3509-3516.2001

Gilardi, G. L. (ed.) (1985). *Nonfermentative Gram-negative Rods: Laboratory Identification and Clinical Aspects*. New York, NY: M. Dekker.

González-Aravena, M., Urtubia, R., Campo, K. D., Lavín, P., Wong, C. M. V. L., Cárdenas, C. A., et al. (2016). Antibiotic and metal resistance of cultivable bacteria in the Antarctic sea urchin. *Antarct. Sci.* 28, 261–268. doi: 10.1017/S0954102016000109

Gorter, F. A., Scanlan, P. D., and Buckling, A. (2016). Adaptation to abiotic conditions drives local adaptation in bacteria and viruses coevolving in heterogeneous environments. *Biol. Lett.* 12:20150879. doi: 10.1098/rsbl.2015.0879

Gualerzi, C. O., Maria Giuliodori, A., and Pon, C. L. (2003). Transcriptional and post-transcriptional control of cold-shock genes. *J. Mol. Biol.* 331, 527–539. doi: 10.1016/S0022-2836(03)00732-730

Guo, Y., Zheng, W., Rong, X., and Huang, Y. (2008). A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int. J. Syst. Evol. Microbiol.* 58, 149–159. doi: 10.1099/ijs.0.65224-65220

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.01310-1313

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hao, S., Yang, P., Han, M., Xu, J., Yu, S., Chen, C., et al. (2018). Data-mining of antibiotic resistance genes provides insight into the community structure of ocean microbiome. *bioRxiv [preprint]* doi: 10.1101/246033

Hesami, S., Metcalf, D. S., Lumsden, J. S., and MacInnes, J. I. (2011). Identification of cold-temperature-regulated genes in *Flavobacterium psychrophilum*. *Appl. Environ. Microbiol.* 77, 1593–1600. doi: 10.1128/AEM.01717-1710

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314.doi: 10.1093/nar/gky1085

Hugh, R., and Leifson, E. (1953). The taxonomic significance of fermentative versus oxidative metabolism of carbohydrates by various Gram negative bacteria. *J. Bacteriol.* 66, 24–26.

Hughes, D., and Andersson, D. I. (2017). Environmental and genetic modulation of the phenotypic expression of antibiotic reistance. *FEMS Microbiol. Rev.* 41, 374–391. doi: 10.1093/femsre/fux004

Hurwitz, B. L., Brum, J. R., and Sullivan, M. B. (2015). Depth-stratified functional and taxonomic niche specialization in the "core" and "flexible". Pacific Ocean virome. *ISME J.* 9, 472–484. doi: 10.1038/ismej.2014.143

Ibrahim, A. M., Hamouda, R. A., El-Naggar, N. E. A., et al. (2021). Bioprocess development for enhanced endoglucanase production by newly isolated bacteria, purification, characterization and in-vitro efficacy as anti-biofilm of *Pseudomonas aeruginosa*. *Sci. Rep.* 11:9754. doi: 10.1038/s41598-021-87901-87909

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-7649

Jang, G. I., Lee, I., Ha, T. T., Yoon, S. J., Hwang, Y. J., Yi, H., et al. (2020). *Pseudomonas neustonica* sp. nov., isolated from the sea surface microlayer of the Ross Sea (Antarctica). *Int. J. Syst. Evol. Microbiol.* 70, 3832–3838. doi: 10.1099/ijsem.0.004240

Jara, D., Bello-Toledo, H., Domínguez, M., Cigarroa, C., Fernández, P., Vergara, L., et al. (2020). Antibiotic resistance in bacterial isolates from freshwater samples in Fildes Peninsula, King George Island, Antarctica. *Sci. Rep.* 10:3145. doi: 10.1038/s41598-020-60035-60030

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

Jung, Y. H., Lee, Y. K., Lee, H. K., Lee, K., and Im, H. (2018). CspB of an arctic bacterium, *Polaribacter irgensii* KOPRI 22228, confers extraordinary freeze-tolerance. *Braz. J. Microbiol.* 49, 97–103. doi: 10.1016/j.bjm.2017.04.006

Kämpfer, P., Busse, H.-J., McInroy, J. A., Xu, J., and Glaeser, S. P. (2015). Flavobacterium nitrogenifigens sp. nov., isolated from switchgrass (*Panicum virgatum*). *Int. J. Syst. Evol. Microbiol.* 65, 2803–2809. doi: 10.1099/ijs.0.000330

Kämpfer, P., Glaeser, S. P., McInroy, J. A., Xu, J., Busse, H.-J., Clermont, D., et al. (2020). *Flavobacterium panici* sp. nov. isolated from the rhizosphere of the switchgrass *Panicum virgatum*. *Int. J. Syst. Evol. Microbiol.* 70, 5824–5831. doi: 10.1099/ijsem.0.004482

Kämpfer, P., Lodders, N., Martin, K., and Avendaño-Herrera, R. (2012). *Flavobacterium chilense* sp. nov. and *Flavobacterium araucananum* sp. nov., isolated from farmed salmonid fish. *Int. J. Syst. Evol. Microbiol.* 62, 1402–1408. doi: 10.1099/ijs.0.033431-33430

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., et al. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48, D454–D458. doi: 10.1093/nar/gkz882

Kawai, Y., Yano, I., and Kaneda, K. (1988). Various kinds of lipoamino acids including a novel serine-containing lipid in an opportunistic pathogen *Flavobacterium*. their structures and biological activities on erythrocytes. *Eur. J. Biochem.* 171, 73–80. doi: 10.1111/j.1432-1033.1988.tb13760.x

Kim, E. J., Kim, J. E., Hwang, J. S., Kim, I.-C., Lee, S. G., Kim, S., et al. (2019). Increased productivity and antifreeze activity of ice-binding protein from *Flavobacterium frigoris* PS1 produced using *Escherichia coli* as bioreactor. *Appl. Biochem. Microbiol.* 55, 489–494. doi: 10.1134/S0003683819050077

Kim, J.-J., Kanaya, E., Weon, H.-Y., Koga, Y., Takano, K., Dunfield, P. F., et al. (2012). Flavobacterium compostarboris sp. nov., isolated from leaf-and-branch compost, and emended descriptions of *Flavobacterium hercynium*, *Flavobacterium resistens* and *Flavobacterium johnsoniae*. *Int. J. Syst. Evol. Microbiol.* 62, 2018–2024. doi: 10.1099/ijs.0.032920-32920

Kimura, H., Young, C. R., Martinez, A., and Delong, E. F. (2011). Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine *Flavobacterium*. *ISME J.* 5, 1641–1651. doi: 10.1038/ismej.2011.36

Kopalová, K., Nedbalová, L., Nıvlt, D., Elster, J., Vijver, B. V., and de. (2013). Diversity, ecology and biogeography of the freshwater diatom communities from Ulu Peninsula (James Ross Island, NE Antarctic Peninsula). *Polar Biol.* 36, 933–948. doi: 10.1007/s00300-013-1317-1315

Kosina, M., Barták, M., Mašlaòová, I., Pascutti, A. V., Šedo, O., Lexa, M., et al. (2013). *Pseudomonas* prosekii sp. nov., a novel psychrotrophic bacterium from antarctica. *Curr. Microbiol.* 67, 637–646. doi: 10.1007/s00284-013-0406-406

Kumar, N., Grogan, P., Chu, H., Christiansen, C. T., and Walker, V. K. (2013). The effect of freeze-thaw conditions on arctic soil bacterial communities. *Biology (Basel)* 2, 356–377. doi: 10.3390/biology2010356

Kurup, V. P., and Babcock, J. B. (1979). Use of casein, tyrosine, and hypoxanthine in the identification of nonfermentative gram-negative bacilli. *Med. Microbiol. Immunol.* 167, 71–75. doi: 10.1007/BF02123556

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lapage, S. P., Shelton, J. E., and Mitchell, T. G. (1970). "Chapter I media for the maintenance and preservation of bacteria," in *Methods in Microbiology*, eds J. R. Norris and D. W. Ribbons (Cambridge, MA: Academic Press), 1–133. doi: 10.1016/S0580-9517(08)70539-7

Larsbrink, J., Zhu, Y., Kharade, S. S., Kwiatkowski, K. J., Eijsink, V. G. H., Koropatkin, N. M., et al. (2016). A polysaccharide utilization locus from *Flavobacterium johnsoniae* enables conversion of recalcitrant chitin. *Biotechnol. Biofuels* 9:260. doi: 10.1186/s13068-016-0674-z

Lee, J. Y., Li, Z., and Miller, E. S. (2017). Vibrio Phage KVP40 encodes a functional NAD+ Salvage pathway. *J. Bacteriol.* 199:e00855-16. doi: 10.1128/JB.00855-816

Lee, S., Weon, H.-Y., Han, K., and Ahn, T.-Y. (2012). *Flavobacterium dankookense* sp. nov., isolated from a freshwater reservoir, and emended descriptions of *Flavobacterium cheonanense*, *F. chungnamense*, *F. koreense* and *F. aquatile*. *Int. J. Syst. Evol. Microbiol.* 62, 2378–2382. doi: 10.1099/ijs.0.032698-32690

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, A.-Z., Han, X.-B., Zhang, M.-X., Zhou, Y., Chen, M., Yao, Q., et al. (2019). Culture-Dependent and -independent analyses reveal the diversity, structure,

and assembly mechanism of benthic bacterial community in the Ross Sea, Antarctica. *Front. Microbiol.* 10:2523. doi: 10.3389/fmicb.2019.02523

Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretdin, A., et al. (2021). RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* 49, D1020–D1028. doi: 10.1093/nar/gkaa1105

Liu, Q., Liu, H.-C., Zhou, Y.-G., and Xin, Y.-H. (2019). Microevolution and adaptive strategy of psychrophilic species *Flavobacterium bomense* sp. nov. isolated from glaciers. *Front. Microbiol.* 10:1069. doi: 10.3389/fmicb.2019.01069

Lowe, G. H. (1962). The rapid detection of lactose fermentation in paracolon organisms by the demonstration of beta-D-galactosidase. *J. Med. Lab. Technol.* 19, 21–25.

McBride, M. J., Xie, G., Martens, E. C., Lapidus, A., Henrissat, B., Rhodes, R. G., et al. (2009). Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl. Environ. Microbiol.* 75, 6864–6875. doi: 10.1128/AEM.01495-1499

McCann, C. M., Christgen, B., Roberts, J. A., Su, J.-Q., Arnold, K. E., Gray, N. D., et al. (2019). Understanding drivers of antibiotic resistance genes in High Arctic soil ecosystems. *Environ. Int.* 125, 497–504. doi: 10.1016/j.envint.2019.01.034

McCarren, J., and DeLong, E. F. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ. Microbiol.* 9, 846–858. doi: 10.1111/j.1462-2920.2006.01203.x

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., and Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 10:2182. doi: 10.1038/s41467-019-10210-10213

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, 280–285. doi: 10.1007/s12275-018-8014-8016

Naas, T., Bellais, S., and Nordmann, P. (2003). Molecular and biochemical characterization of a carbapenem-hydrolysing β-lactamase from *Flavobacterium johnsoniae*. *J. Antimicrob. Chemother.* 51, 267–273. doi: 10.1093/jac/dkg069

Nedbalová, L., Nɪvlt, D., Kopáèek, J., Šobr, M., and Elster, J. (2013). Freshwater lakes of Ulu Peninsula, James Ross Island, north-east Antarctic Peninsula: origin, geomorphology and physical and chemical limnology. *Antarct. Sci.* 25, 358–372. doi: 10.1017/S0954102012000934

Nett, M., and König, G. M. (2007). The chemistry of gliding bacteria. *Nat. Prod. Rep.* 24, 1245–1261. doi: 10.1039/B612668P

Oberhofer, T. R., and Rowen, J. W. (1974). Acetamide agar for differentiation of nonfermentative bacteria. *Appl. Microbiol.* 28, 720–721.

Orellana-Saez, M., Pacheco, N., Costa, J. I., Mendez, K. N., Miossec, M. J., Meneses, C., et al. (2019). In-Depth genomic and phenotypic characterization of the antarctic psychrotolerant strain *Pseudomonas* sp. MPC6 reveals unique metabolic features, plasticity, and biotechnological potential. *Front. Microbiol.* 10:1154. doi: 10.3389/fmicb.2019.01154

Owens, J. J. (1974). The egg yolk reaction produced by several species of bacteria. *J. Appl. Bacteriol.* 37, 137–148. doi: 10.1111/j.1365-2672.1974.tb00424.x

Pacova, Z., and Kocur. (1984). New medium for detection of esterase and gelatinase activity. *Zentralbl. Bakteriol. Mikrobiol. Hyg. Ser. A* 258, 69–73. doi: 10.1016/S0176-6724(84)80010-80013

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Pasamontes, L., Hug, D., Tessier, M., Hohmann, H.-P., Schierle, J., and van Loon, A. P. G. M. (1997). Isolation and characterization of the carotenoid biosynthesis genes of *Flavobacterium sp*. strain R1534. *Gene* 185, 35–41. doi: 10.1016/S0378-1119(96)00624-625

Peck, R. F., Echavarri-Erasun, C., Johnson, E. A., Ng, W. V., Kennedy, S. P., Hood, L., et al. (2001). brp and blh are required for synthesis of the retinal cofactor of bacteriorhodopsin in *Halobacterium salinarum*. *J. Biol. Chem.* 276, 5739–5744. doi: 10.1074/jbc.M009492200

Piette, F., D'Amico, S., Struvay, C., Mazzucchelli, G., Renaut, J., Tutino, M. L., et al. (2010). Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Mol. Microbiol.* 76, 120–132. doi: 10.1111/j.1365-2958.2010.07084.x

Reddy, G. S. N., Matsumoto, G. I., Schumann, P., Stackebrandt, E., and Shivaji, S. (2004). *Psychrophilic pseudomonads* from Antarctica: *Pseudomonas antarctica* sp. nov., *Pseudomonas meridiana* sp. nov. and *Pseudomonas proteolytica* sp. nov. *Int. J. Syst. Evol. Microbiol.* 54, 713–719. doi: 10.1099/ijs.0.02827-2820

Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., et al. (2018). The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* 46, W282–W288. doi: 10.1093/nar/gky467

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

Romanenko, L. A., Tanaka, N., Svetashev, V. I., Kurilenko, V. V., and Mikhailov, V. V. (2015). *Flavobacterium maris* sp. nov. isolated from shallow sediments of the Sea of Japan. *Arch. Microbiol.* 197, 941–947. doi: 10.1007/s00203-015-1128-x

Sasser, M. (1990). *Identification of Bacteria by Gas Chromatography of Cellular Fatty Acids, MIDI Technical Note 101*. Newark, NJ: MIDI Inc.

Schöner, T. A., Gassel, S., Osawa, A., Tobias, N. J., Okuno, Y., Sakakibara, Y., et al. (2016). Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *ChemBioChem.* 17, 247–253. doi: 10.1002/cbic.201500474

Shapiro, J. W., Williams, E. S. C. P., and Turner, P. E. (2016). Evolution of parasitism and mutualism between filamentous phage M13 and *Escherichia coli*. *PeerJ* 4:e2060. doi: 10.7717/peerj.2060

Song, W., Sun, H.-X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., et al. (2019). Prophage hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* 47, W74–W80. doi: 10.1093/nar/gkz380

Stolz, A., Busse, H.-J., and Kämpfer, P. (2007). *Pseudomonas knackmussii* sp. nov. *Int. J. Syst. Evol. Microbiol.* 57, 572–576. doi: 10.1099/ijs.0.64761-64760

Taboada, B., Estrada, K., Ciria, R., and Merino, E. (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* 34, 4118–4120. doi: 10.1093/bioinformatics/bty496

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

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Thierauf, A., Perez, G., and Maloy, S. (2009). "Generalized transduction," in *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions Methods in Molecular BiologyTM*, eds M. R. J. Clokie and A. M. Kropinski (Totowa, NJ: Humana Press), 267–286. doi: 10.1007/978-1-60327-164-6_23

Tindall, B. J. (1990b). Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiol. Lett.* 66, 199–202. doi: 10.1111/j.1574-6968.1990.tb03996.x

Tindall, B. J. (1990a). A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. *Sys. Appl. Microbiol.* 13, 128–130. doi: 10.1016/S0723-2020(11)80158-X

Tribelli, P. M., Pezzoni, M., Brito, M. G., Montesinos, N. V., Costa, C. S., and López, N. I. (2020). Response to lethal UVA radiation in the Antarctic bacterium *Pseudomonas extremaustralis*: polyhydroxybutyrate and cold adaptation as protective factors. *Extremophiles* 24, 265–275. doi: 10.1007/s00792-019-01152-1151

Van Goethem, M. W., Pierneef, R., Bezuidt, O. K. I., Van De Peer, Y., Cowan, D. A., and Makhalanyane, T. P. (2018). A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome* 6:40. doi: 10.1186/s40168-018-0424-425

Wahli, T., and Madsen, L. (2018). Flavobacteria, a never ending threat for fish: a review. *Curr. Clin. Micro. Rpt.* 5, 26–37. doi: 10.1007/s40588-018-0086-x

Waschulin, V., Borsetto, C., James, R., Newsham, K. K., Donadio, S., Corre, C., et al. (2020). Metabolic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *bioRxiv [preprint]* doi: 10.1101/2020.12.09.416412

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, 2nd Edn. Berlin: Springer International Publishing, doi: 10.1007/978-3-319-24277-4

Wilkins, D., and Kurtz, Z. (2019). *gggenes: Draw Gene Arrow Maps in "ggplot2". R package version 0.4. 0*.

Wilson, S. L., Kelley, D. L., and Walker, V. K. (2006). Ice-active characteristics of soil bacteria selected by ice-affinity. *Environ. Microbiol.* 8, 1816–1824. doi: 10.1111/j.1462-2920.2006.01066.x

Yi, H., and Chun, J. (2006). *Flavobacterium weaverense* sp. nov. and *Flavobacterium segetis* sp. nov., novel psychrophiles isolated from the Antarctic. *Int. J. Syst. Evol. Microbiol.* 56, 1239–1244. doi: 10.1099/ijs.0.64164-64160

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479

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

Yoshizawa, S., Kawanabe, A., Ito, H., Kandori, H., and Kogure, K. (2012). Diversity and functional analysis of proteorhodopsin in marine *Flavobacteria*. *Environ. Microbiol.* 14, 1240–1248. doi: 10.1111/j.1462-2920.2012.02702.x

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother* 67, 2640–2644. doi: 10.1093/jac/dks261

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Králová, Busse, Bezdíček, Sandoval-Powers, Nykrýnová, Staňková, Krsek and Sedláček. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 November 2021
doi: 10.3389/fmicb.2021.730231

The Piezo-Hyperthermophilic Archaeon *Thermococcus piezophilus* Regulates Its Energy Efficiency System to Cope With Large Hydrostatic Pressure Variations

*Yann Moalic[1,2], Jordan Hartunians[1,2], Cécile Dalmasso[1,2], Damien Courtine[1,2], Myriam Georges[1,2], Philippe Oger[3], Zongze Shao[2,4], Mohamed Jebbar[1,2] and Karine Alain[1,2]**

[1] Univ Brest, CNRS, Ifremer, Laboratoire de Microbiologie des Environnements Extrêmes LM2E, UMR 6197, IUEM, Plouzané, France, [2] IRP 1211 MicrobSea, Sino-French Laboratory of Deep-Sea Microbiology, LM2E, Plouzané, France, [3] Université de Lyon, INSA Lyon, CNRS UMR 5240, Villeurbanne, France, [4] Key Laboratory of Marine Biogenetic Resources, The Third Institute of Oceanography SOA, Xiamen, China

OPEN ACCESS

Edited by:
Rafael R. de la Haba,
University of Sevilla, Spain

Reviewed by:
Aude Picard,
University of Nevada, Las Vegas, United States
James F. Holden,
University of Massachusetts Amherst, United States

***Correspondence:**
Karine Alain
Karine.Alain@univ-brest.fr

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

Received: 24 June 2021
Accepted: 13 October 2021
Published: 03 November 2021

Citation:
Moalic Y, Hartunians J, Dalmasso C, Courtine D, Georges M, Oger P, Shao Z, Jebbar M and Alain K (2021) The Piezo-Hyperthermophilic Archaeon Thermococcus piezophilus Regulates Its Energy Efficiency System to Cope With Large Hydrostatic Pressure Variations. Front. Microbiol. 12:730231. doi: 10.3389/fmicb.2021.730231

Deep-sea ecosystems share a common physical parameter, namely high hydrostatic pressure (HHP). Some of the microorganisms isolated at great depths have a high physiological plasticity to face pressure variations. The adaptive strategies by which deep-sea microorganisms cope with HHP variations remain to be elucidated, especially considering the extent of their biotopes on Earth. Herein, we investigated the gene expression patterns of *Thermococcus piezophilus,* a piezohyperthermophilic archaeon isolated from the deepest hydrothermal vent known to date, under sub-optimal, optimal and supra-optimal pressures (0.1, 50, and 90 MPa, respectively). At stressful pressures [sub-optimal (0.1 MPa) and supra-optimal (90 MPa) conditions], no classical stress response was observed. Instead, we observed an unexpected transcriptional modulation of more than a hundred gene clusters, under the putative control of the master transcriptional regulator SurR, some of which are described as being involved in energy metabolism. This suggests a fine-tuning effect of HHP on the SurR regulon. Pressure could act on gene regulation, in addition to modulating their expression.

Keywords: pressure, piezophile, thermococci, hydrothermal, transcriptomics

INTRODUCTION

Seventy-five percent of the ocean volume are subjected to pressures above 10 MPa (Jannasch and Taylor, 1984), described as high hydrostatic pressure (HHP) environments (Oger and Jebbar, 2010). Some "pressure-adapted" microorganisms, called piezophiles, live in these deep marine biosphere ecosystems (Yayanos, 1995). By definition, piezophilic organisms are characterized by a higher growth rate under HHP than at atmospheric pressure. HHP have important impacts onto macromolecules, notably membrane and proteins, and onto physiology and metabolism (Oger and Jebbar, 2010). Piezophilic microorganisms have developed various strategies to cope with HHP. These strategies have mainly been described in psychro-piezophilic bacteria. They include changes

in gene expression, switch in metabolism to counteract the loss of biological activity, structural adaptations of macromolecules in order to face HHP and specific stress responses (Simonato et al., 2006; Oger and Jebbar, 2010; Picard and Daniel, 2013; Jebbar et al., 2015).

Although deep oceanic hydrothermal vents are characterized by high temperatures and hydrostatic pressures, very few (hyper)thermophilic piezophilic prokaryotes have been isolated from these ecosystems, and of the available isolates few have been tested for pressure tolerance, mostly due to the inherent technical constraints. Thus, the adaptation mechanisms of (hyper)thermophilic piezophiles remain unwell documented. Compared to psychrophiles, in which pressure and cold adaptation mechanisms may be overlapping, it is easier to independently sift through pressure adaptation mechanisms within (hyper)thermophilic models where high pressures and high temperatures have antagonistic effects (Siliakus et al., 2017). So far, only few (hyper)thermophilic piezophilic or piezosensitive microorganisms have been studied with respect to their response to HHP and most studies have focused on the order *Thermococcales*, a lineage of hyperthermophilic, heterotrophic, sulfur-reducing archaea, some of which are ubiquitous in the hot areas of deep-sea hydrothermal vents. Many studies focused particularly on the piezophiles *Thermococcus barophilus* and *Pyrococ*cus yayanosii (Cario et al., 2015b; Vannier et al., 2015; Michoud and Jebbar, 2016). The transcriptomic response to supra-optimal pressures for growth of *Thermococcus kodakarensis,* a piezosensitive archaeon belonging to *Thermococcales*, resembles a classical stress response. It involves an up-regulation of genes involved in replication, repair and defense mechanisms, and a down-regulation of genes involved in energy production/conversion and in amino acid transport and metabolism. In contrast, the piezophilic taxon *T. barophilus* responds to changes in hydrostatic pressure by some variations in the metabolic pathways used and notably by variations of expression levels of genes coding for transporters, hydrogenases and oxidoreductases (Vannier et al., 2015). At stressful sub-optimal pressures, it also accumulates mannosylglycerate, a compatible solute, probably to help maintain cell volume and homeostasis (Cario et al., 2016). Adaptations at the membrane level (variations of the ratio archaeol/caldarchaeol and of lycopene unsaturations) were also observed in *T. barophilus* to maintain membrane fluidity as a result of pressure changes (Cario et al., 2015a). In addition, proteome adaptations (flexibility linked to reduced hydration water dynamics) of piezophilic species were also reported (Martinez et al., 2016). The obligate piezophile *P. yayanosii* responds to supra- and sub-optimal pressure changes by a major metabolic shift, namely a repression of the H_2 metabolism and an increase in activity of sulfur-dependent hydrogenases, in addition to changes in the chemotaxis pathway, translation and CRISPR-*cas* gene expressions (Michoud and Jebbar, 2016). Overall, *Thermococcales* respond to pressure variations through structural flexibility and modulation of gene expression.

The hyperthermophilic piezophilic archaeon *Thermococcus piezophilus* strain CDGS^T was isolated from the world's deepest known hydrothermal vent field, at nearly 5,000 m water depth, at the Mid-Cayman rise (Dalmasso et al., 2016b). It has an optimal pressure for growth of 50 MPa and holds the current record of pressure range for growth, growing effectively from atmospheric pressure to at least 120 MPa, and, though with difficulty up to 130 MPa (Dalmasso et al., 2016b). Due to its wide range of growth pressure, it appeared as a good model organism to perform transcriptome-level studies of adaptation to various pressure conditions, as for *T. barophilus* and *P.* yayanosii (Vannier et al., 2015; Michoud and Jebbar, 2016). To address the question of the effects of HHP variations on the transcriptome of *T. piezophilus*, RNA-seq analyses were performed on cells grown at sub-optimal (0.1 MPa), optimal (50 MPa), and supra-optimal (90 MPa) pressures for growth. The selection of these pressures was experimentally determined (growth yields identical to 0.1and 90 MPa and sufficient biomass to perform transcriptomic analysis) (Dalmasso et al., 2016b). Here, by highlighting key gene clusters and the associated metabolic pathways engaged at optimal, sub-optimal and supra-optimal pressures for growth of this hyperthermophilic piezophilic archaeal model, we provide new insights into the molecular responses of piezophiles to pressure variations.

MATERIALS AND METHODS

Medium Culture and Growth Conditions

Thermococcus piezophilus strain CDGS^T (UBOCC 3296^T = ATCC TSD-33^T) was isolated from a hydrothermal chimney sample from the Mid Cayman rise Beebe vent field, at 4,964 m water depth (Dalmasso et al., 2016b). The strain was grown at 75°C under anaerobic conditions, in modified Ravot medium prepared without maltose, with polysulfur (1% of polysulfide from a 0.05 M stock solution), as described elsewhere (Dalmasso et al., 2016b). Cells were grown until the mid-exponential growth phase (11 h at 0.1 MPa, 9 h at 50 MPa, and 12 h at 90 MPa) into 20 mL sterile plastic syringes incubated at 0.1, 50, and 90 MPa of hydrostatic pressure (Thermostated HP/HT incubators Top Industrie, France), from independent startup cultures grown at 50 MPa. The 3 biological replicates per pressure were incubated in the same stainless-steel pressure vessel (pressurized by pumping water into them), at the same time, together with negative controls. Cells were then harvested by centrifugation.

Determination of Cell Numbers

Cell counts were performed on a Thoma chamber (Preciss, France; surface area: 0.0025 mm^2, depth: 0.100 mm) and by phase contrast microscopy (Olympus BX60), to verify that the cell density was similar to that obtained at the end of the exponential growth phase during growth kinetics performed previously under exactly the same conditions.

RNA Extraction and Purification

As soon as the cell pellets (from 20 mL cultures) were obtained, TRIzol® reagent (Ambion, Life Technologies) was added to them in order to extract total RNA. DNase I treatment (Promega®) was then performed at 37°C for 30 min to remove genomic

DNA from the isolated total RNA. Total RNA was further purified onto RNA spin columns (RNeasy Mini kit Qiagen®, RNA cleanup protocol), and stored at −80°C after verification of non-degradation with an Agilent 2100 Bioanalyzer (RNA 6000 Pico chip kit) and before sequencing. RNA concentrations were measured with a NanoDrop ND 1000® spectrophotometer (Thermo Fisher Scientific Inc.).

RNA-Seq

A Ribo-Zero rRNA Removal Bacteria kit (Epicenter, Madison, WI) used with manufacturer's protocol was applied to deplete a maximum of rRNAs and enrich total RNA in mRNA. As it is often the case with *Archaea*, ribodepletion attempts were poorly efficient and the RNA-seq dataset still contained 73–94% ribosomal sequences. Ion Total RNA-Seq Kit v2 was used to make cDNA, to add barcodes, and to amplify the library. All cleanup steps were performed using Agencourt Ampure XP beads (Thermo Fisher Scientific). Each step was validated by a bioanalyzer quality control on Agilent RNA chips and/or with the "Qubit™" fluorometer (Life Technologies) to determine the quality, size of fragments and quantity of produced material. RNA libraries were sequenced on P1v2 chips using the Ion S5™ System (Thermo Fisher Scientific). Sequencing was completed by the GeT-Genotoul platform in Toulouse (France). The RNA-seq data have been deposited into the NIH Sequence Read Archive (SRA) under the BioProject ID PRJNA739722 (accession numbers SRX11191114-11191122).

Bioinformatic Data Analysis

RNA-seq reads were mapped on all genes with Bowtie v2.3.1 (Langmead et al., 2009). Bowtie2 was run with default parameters and *–no-unal*. The result was converted from SAM to BAM format using Samtools v1.3.1 (Li et al., 2009). The reads counts were normalized to Transcripts Per Kilobase Million (TPM) (**Supplementary Table 1**). Then, TPM were mapped with Anvi'o (v7) (Eren et al., 2015) onto the *T. piezophilus* genome sequence deposited in the NCBI database under the accession number GCA_001647085.1 ASM164708v1. First, a contig database was generated (*anvi-gen-contigs-database*) with the genome sequence available on NCBI and a table containing all gene positions (*–external-gene-calls* and *–ignore-internal-stop-codons* arguments). Next, BAM files were processed individually with *anvi-init-bam* to get indexed alignment files. The latter were used to generate individual anvi'o profile to link the information available in the alignment file to the contigs databases (*anvi-profile* with required parameters and *–min-contig-length 70*). Last, all three profiles were merged with *anvi-merge* (default parameters). The command *anvi-interactive* was used to visualize the mapping results over the three conditions.

Differential Gene Expression Analysis

Data normalization and differential gene expression analysis were done with R/bioconductor packages edgeR (Robinson et al., 2010) and DEseq2 (Love et al., 2014). For each pressure condition, the data obtained from three independent biological replicate cDNA libraries were analyzed. Principal component analysis (PCA) was used to detect potential outlier on the samples. No outlier was detected. Regarding the filtration of transcripts, EdgeR's RLE ("Relative Log Expression") standardization was used, with filtering of weakly expressed genes, and then correction of tests with the BH method (Benjamini-Hochberg procedure for False Discovery Rate; alpha threshold = 0.05). The DEseq2 package was used to investigate the differential expression of genes between the different pressure conditions. Transcripts that were considered statistically significant were those with an adjusted p-value $< 1\%$.

Gene Ontology Analysis

Functional annotation clustering of genes differentially expressed was carried out with the online software DAVID (Huang da et al., 2009),[1] with default parameters.

Motifs Binding Sites Screening

The search for SurR and TrmB-like binding motifs within *T. piezophilus* genome was carried out using in-house Perl scripts. Gene promoters were defined as the 200 pb zones upstream of the start codon of the locus tag.

RESULTS AND DISCUSSION

General Response of *Thermococcus piezophilus* to Hydrostatic Pressure

Nine independent biological replicates, each inoculated with a different startup culture [generated at the optimal growth pressure (50 MPa)] were incubated in high-temperature stainless steel vessels at the three selected pressures (3 × 0.1 MPa; 3 × 50 MPa; 3 × 90 MPa). Their transcriptomes were generated after RNA extraction with TRIzol®, then ribodepletion. After sequence filtration, there were 87,940,548 non-ribosomal usable total reads that were mapped as transcripts per kilobase million (TPM) on the 2,126 concatenated coding DNA sequence (CDS) of *T. piezophilus* (**Figure 1A**). After the differential expression validation step, it appeared that 1,373 of the 2,126 genes were regulated by pressure changes (**Figure 1C** but see **Supplementary Table 2** for details) but, at most, the expression of 540 genes varied significantly when shifting from a tested pressure to another (**Figure 1B** and **Supplementary Figure 1**). At 0.1 MPa, 540 genes were overexpressed and 532 were underexpressed, while at 90 MPa, 220 genes were overexpressed and 224 genes were underexpressed compared to 50 MPa. There were more genes differentially regulated by pressure that were shared between high pressures (50 or 90 MPa) and atmospheric condition (0.1 MPa) (270 or 339 shared genes, respectively) than genes differentially regulated by pressure that were shared when comparing non-optimal pressures (0.1 or 90 MPa) to optimal growth condition (50 MPa) (109 or 125, respectively). The hierarchical clustering on z-scores highlighted a closer gene expression response between 0.1 and 90 MPa (**Figure 1C**).

It is therefore possible that the regulatory response to pressure variation compared to optimal growth condition (50 MPa) may

[1] https://david.ncifcrf.gov/summary.jsp

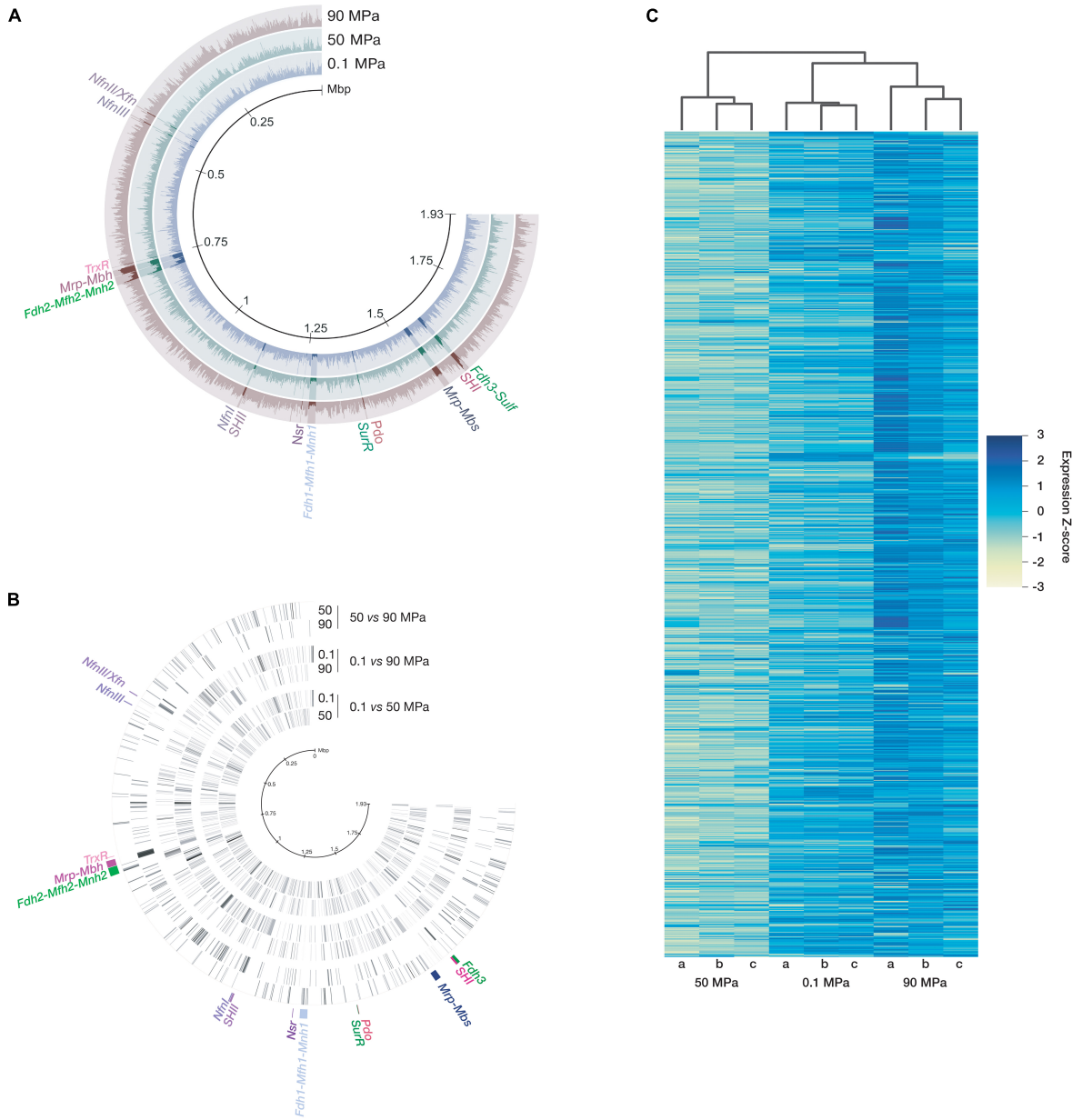


FIGURE 1 | Read mapping and differential gene expression as a function of pressure variation. After normalization, the TPM were mapped on concatenated CDSs of *T. piezophilus* **(A)**. Some of the pressure-regulated genes or gene clusters are highlighted. The concentric circles represent the 3 conditions. Each ray represents a gene expression level **(B)**, ordered according to the position of the gene on the genome (indicated by the scale in the center). The logFC values were used for signal intensity and the color black was given for the maximum. Key genes of the hydrogenogenic/sulfidogenic metabolisms are displayed. These representations were made with Anvi'o (Eren et al., 2015). **(C)** Z-score hierarchical clustering heat map for visualization of the 3 biological replicates (a, b, and c) per pressure.

be broader and more specific at atmospheric pressure than at 90 MPa, even though the growth rate is almost identical at 0.1 and 90 MPa (Dalmasso et al., 2016b).

In order to gain a better understanding of the molecular tuning that takes place into the cell, we explored more precisely (i) the expression level of the genes regulated by two-well characterized transcriptional regulators (SurR and regulators of the TrmB family), (ii) the categories of genes regulated by pressure and (iii) the function of gene clusters whose expression is impacted by pressure, with emphasis on those involved in energy conversion and conservation.

Is Pressure a Physical Parameter Able to Modify the Transcriptional Regulation Mechanism?

The adaptive response to fluctuating environmental conditions has been less studied in *Archaea* than in other life domains

TABLE 1 | Variations in the expression of genes or gene clusters, known to be involved in hydrogen and sulfur metabolism.

Product	Locus tag	Total CDS number	Number of overexpressed coding DNA sequence (CDS)					
			0.1 > 50 MPa	0.1 > 90 MPa	50 > 0.1 MPa	50 > 90 MPa	90 > 0.1 MPa	90 > 50 MPa
Mrp-Mbh	A7C91_RS04230-04165	14	8	nd	nd	nd	13	14
Fdh1-Mfh1-Mnh1	A7C91_RS06945-07020	16	nd	nd	6	7	nd	nd
Fdh2-Mfh2-Mnh2	A7C91_RS04320-04240	17	8	12	nd	7	nd	nd
F_{420}-reducing hydrogenase	A7C91_RS_04340-04330	3	3	3	nd	nd	nd	nd
SHI	A7C91_RS08740-08725	4	nd	nd	nd	nd	4	4
SHII	A7C91_RS06200-06185	4	4	3	nd	nd	nd	4
Mrp-Mbs	A7C91_RS08510-08450	13	12	13	nd	4	nd	nd
Nsr	A7C91_RS06865	1	nd	nd	nd	nd	nd	nd
SurR	A7C91_RS07565	1	1	1	nd	nd	nd	nd
Pdo	A7C91_RS07570	1	nd	nd	1	nd	nd	nd
Nfn I	A7C91_RS06205-6210	2	nd	nd	1	nd	2	2
Nfn II/Xfn	A7C91_RS02315-2310	2	nd	nd	2	1	1	nd
Nfn III	A7C91_RS02425-2430	2	nd	nd	1	nd	2	nd
Fdh3	A7C91_RS08745-08770	6	nd	nd	2	1	nd	nd
RNR	A7C91_RS02975	1	nd	nd	1	nd	1	nd
ATP synthase	A7C91_RS03355-03315	9	nd	nd	6	nd	7	nd
TrxR	A7C91_RS04130	1	nd	nd	1	nd	1	nd

(Karr, 2014). In *Thermococcales*, the regulators of the TrmB family and the SurR regulator are the two best-characterized transcriptional regulators. Regulators in the TrmB family control transcription according to sugar availability (Lee et al., 2003, 2005, 2008). These regulators bind to DNA in promoter regions to activate or repress the transcription of genes, thanks to TrmB-like binding motifs which have been published elsewhere (Gindner et al., 2014). SurR is the master regulator of primary electron flow pathways that responds directly to the presence or absence of sulfur in the environment (Lipscomb et al., 2009, 2017; Hidese et al., 2017). This ArsR-type master-regulator, which is unique to *Thermococcales* and conserved, is able to activate or repress the expression of the genes involved in the response to sulfur thanks to the presence of SurR binding motifs (GTTn_3AAC or GTTn_3AACn_5GTT) in their promoters (Lipscomb et al., 2009). Its sequence-specific DNA binding activity is driven by a redox-active displacement of cysteine residues within a CxxC motif. In the presence of sulfur, a disulfide bond is formed and this oxidized form of SurR inhibits its binding activity. As a result, it loses its transcriptional modulation activities which are strongly correlated with the position of its binding sites, facilitating the recruitment of transcriptional machinery or blocking its progression (Yang et al., 2010).

It is interesting to note that while this study focused on the effect of variation in hydrostatic pressure (with no change in medium composition between experiments), we observed an increase in transcripts for the three TrmB-like genes (A7C91_RS03095, A7C91_RS05650, and A7C91_RS07665) and SurR gene (A7C91_RS07565) at 0.1 MPa vs. 50 and 90 MPa (**Supplementary Table 2**). In order to assess the potential overall effect of this regulation, the presence of SurR and TrmB-like binding motifs was examined in *T. piezophilus* gene promoters.

TrmB-like binding motifs were detected in the promoter of only nine genes, of which only two were found to be regulated by pressure. This result seems to point to a limited role of TrmB and its targets in the response to pressure. However, due to evolutionary constraints, it is likely that the binding motifs used in our investigation in this archaeal species may be slightly different from those described in other species. In fact, these motifs were not detected in the promoter of the maltodextrin operon (MD) which is also overexpressed at 0.1 MPa vs. 50 and 90 MPa, whereas they were expected to be found there.

With regard to SurR, many genes under the control of this regulator have their expression modulated by pressure (**Table 1**; **Figure 2**, and **Supplementary Figure 2**; see next chapter). Thus, 170 promoters bearing at least one binding motif were detected among the 2,126 total CDS of *T. piezophilus*. In addition, among the 1373 CDS whose expression was regulated by pressure, 119 harbored a SurR binding motif. These values represent 8 and 8.7% of the total and overexpressed CDS, respectively, highlighting no significant increase in the expression of genes potentially under the control of SurR at different pressure conditions. However, given that 70% of genes potentially regulated by SurR are also pressure-regulated, this could be a clue of a high plasticity in adaptive response in *T. piezophilus,* but also within *Thermococcales* in general. Indeed, there is an effective variation in gene expression of SurR-regulated clusters. This could be the consequence of a pressure effect on the conformational state of SurR, modifying its binding capacity even in presence of sulfur, as has been shown for TrmB in *P. furiosus* (Krug et al., 2013). Moreover, a detailed analysis of the motif position in the promoters, revealed the presence of several potential binding sites, which, according to the BRE/TATA box, could provide some flexibility in terms of repression or activation, as

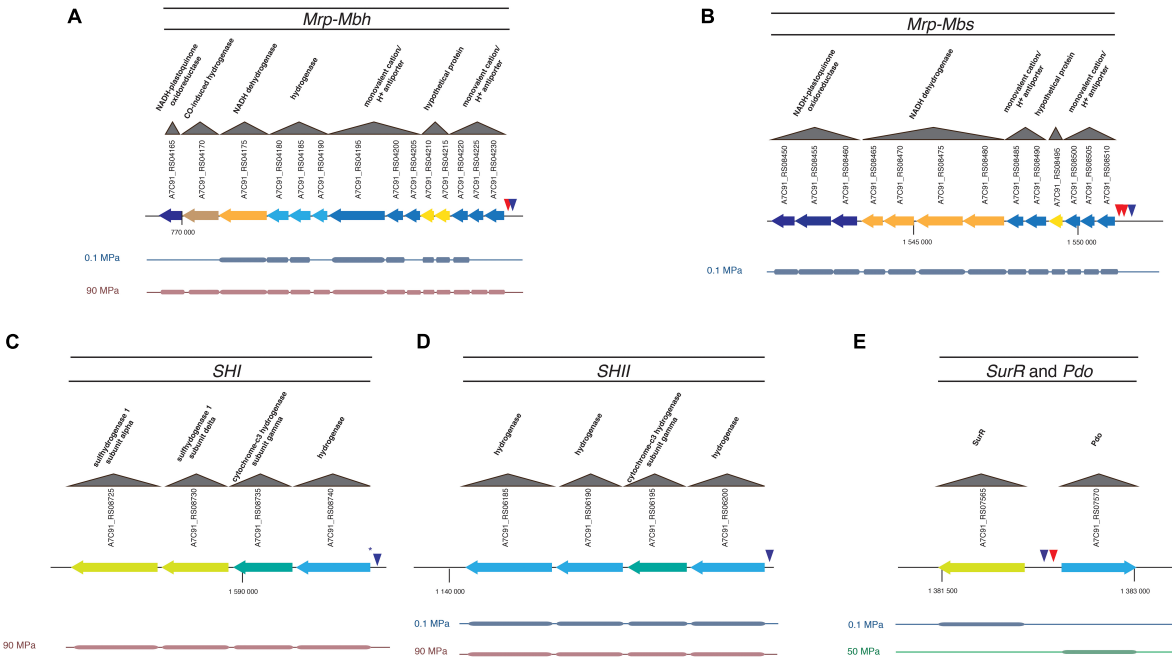


FIGURE 2 | MBH, MBS, SHI, SHII, and SurR gene clusters regulated in hydrogenogenic/sulfidogenic metabolisms and their SurR binding motifs. **(A)** *Mrp-Mbh*; **(B)** *Mrp-Mbs*; **(C)** *SHI*; **(D)** *SHII*; **(E)** *SurR* and *Pdo*. Each gene is colored according to its differential expression condition (0.1 MPa in blue, 50 MPa in green, or 90 MPa in red). Triangles represent the SurR binding motifs. Red triangle is for the short binding motif (GTTn_3AAC) and blue triangle is for the long one (GTTn_3AACn_5GTT). The asterisk means that the motif carries a mutation. The other gene clusters regulated in hydrogenogenic/sulfidogenic metabolisms are shown in **Supplementary Figure 2**.

for the *Mbs* cluster (**Table 2**). In addition, some of the genes described as being regulated by SurR in other *Thermococcales* bear mutated motifs, such as for the ones encoding SHI (**Table 2**), but this will need to be further investigated by an EMSA assay to see if this affects the binding affinity of SurR. One other hypothesis would be the involvement of another transcriptional regulator. This hypothesis is supported by the presence of 37 other regulators among the genes regulated by pressure (**Supplementary Table 2**). Their respective occurrences in terms of over-expression according to pressure, i.e., 0.1, 50, and 90 MPa, are 23, 14, and 11, respectively (**Supplementary Table 2**). In the light of these results, the sub-optimal condition (0.1 MPa) seems to require a higher number of regulator overexpressions than the supra-optimal growth condition (90 MPa), respectively 21 vs. 4. This correlates with the greater number of total genes regulated in these conditions (540 and 220, respectively, **Supplementary Figure 1**). Overall, these data support the hypothesis that atmospheric pressure requires a stronger adaptive response than higher hydrostatic pressure conditions in *T. piezophilus*.

Categories of Genes Regulated by Pressure

To uncover the molecular mechanisms behind the pressure adaptive responses, the differentially expressed genes (DEG) were analyzed with the functional annotation tool DAVID which allows identifying functional categories and ranking them based on gene-enrichments in annotation terms (EASE score) (Huang da et al., 2009; **Supplementary Table 3**). Gene clusters characterized by a gene-enrichment (in this instance having an EASE score superior to 1) were analyzed in details. Results are presented in **Supplementary Figure 1**. The broadest functional response was predicted for 90 MPa vs. 0.1 MPa with five different classes of genes that were enriched (iron, hydrogen ion transport, ligase, ribosomal protein and nucleotidyltransferase) vs. one or two for the other conditions. These results show that the number of regulated genes can be twice as high in one comparison of conditions as in another without necessarily being more functionally diverse (540 vs. 220 overexpressed genes at 0.1 and 90 MPa, respectively against 50 MPa). As observed in other *Thermococcus* species, the adaptive response to pressure variations in *Thermococcales* appears to be more balanced, global and integrative than a classical response to a physical stress (Vannier et al., 2015; Michoud and Jebbar, 2016).

Functions of the Gene Clusters Regulated by Pressure

Main Functions Impacted by Pressure

The vast majority of gene expression variation was related to a global metabolic response, as described for *T. barophilus* and *P. yayanosii* (Vannier et al., 2015; Michoud and Jebbar, 2016). Variations in the level of transcription occurred globally at the level of gene clusters (**Figure 1A**). *T. piezophilus* appeared to adapt to pressure changes by modulating amino

TABLE 2 | SurR binding motifs positions upstream from the start codon of CDS involved in the electron flow.

Product	Locus tag	First CDS of the product	Motif occurrence	Distance (size) of the binding motif from start codon of the first CDS		
Mrp-Mbh	A7C91_RS04230-04165	A7C91_RS04230	2	21 (short)	127 (long)	
Fdh1-Mfh1-Mnh1	A7C91_RS06945-07020	A7C91_RS06945	1	137 (short with mutation: gttaccaat)		
Fdh2-Mfh2-Mnh2	A7C91_RS04320-04240	A7C91_RS04320	1	123 (short)		
F_{420}-reducing hydrogenase	A7C91_RS_04340-04330	A7C91_RS04340	1	71 (long)		
SHI	A7C91_RS08740-08725	A7C91_RS_08740	1	105 (long with mutation: gttttaacctttggtt deletion)		
SHII	A7C91_RS06200-06185	A7C91_RS06200	1	69 (long)		
Mrp-Mbs	A7C91_RS08510-08450	A7C91_RS08510	3	29 (short)	62 (short)	98 (long)
Nsr	A7C91_RS06865	A7C91_RS06865	1	17 (short)		
SurR	A7C91_RS07565	A7C91_RS07565	2	49 (long)	103 (short)	
Pdo	A7C91_RS07570	A7C91_RS07570	2	23 (short)	69 (short)	
Nfn I	A7C91_RS06205-6210	A7C91_RS06205	1	72 (long)		
Nfn II/Xfn	A7C91_RS02315-2310	A7C91_RS02315	2	44 (short)	72 (short)	
Nfn III	A7C91_RS02425-2430	A7C91_RS02425	1	42 (short)		
Fdh3	A7C91_RS08745-08770	A7C91_RS04340	1	71 (long)		
RNR	A7C91_RS02975	A7C91_RS02975	0	/		
ATP synthase	A7C91_RS03355-03315	A7C91_RS03355	0	/		
TrxR	A7C91_RS04130	A7C91_RS04130	0	/		

Short: GTT*n*3AAC, long: GTT*n*3AAC*n*5GTT

acid and vitamin synthesis, production of potential compatibles solutes, and regulating the chemotaxis, S-layer and CRISPR-Cas systems. All these results are described in details in **Supplementary Data 1**. In summary, the strain does not possess the genetic potential to synthesize the three aromatic amino acids L-phenylalanine, L-tyrosine, and L-tryptophan, and underexpresses genes for the synthesis of a.a histidine while overexpressing genes encoding different families of transporters of amino acids, of oligopeptides and of dipeptides under non-optimal conditions. In the same fashion, few genes of thiamine and biotin pathways were underexpressed at 0.1 and 90 MPa while some vitamin transporters were overexpressed at 90 MPa. These observations suggest that the import of certain amino acids and vitamins may be preferred over *de novo* synthesis under non-optimal pressures. Furthermore, overexpression of glutamate biosynthesis genes was observed at 0.1 MPa, suggesting an overproduction of this metabolite at low pressures, possibly to help support macromolecule conformation and function. In addition, a large majority of the genes encoding the chemotaxis pathway were transcriptionally overexpressed at 50 MPa, and even more so at 90 MPa, compared to 0.1 MPa (**Supplementary Figure 3**), as described in **Supplementary Data 1**. Finally, two genes coding for protein subunits of the S-layer and the gene clusters of the CRISPR-*cas* type I-B and of the CRISPR-*cas* type III-A systems were overexpressed at 0.1 MPa, as detailed in **Supplementary Data 1**. These results are consistent with those reported for other piezophilic *Thermococales* or other piezophilic microorganisms (Amrani et al., 2014; Vannier et al., 2015; Michoud and Jebbar, 2016), hence they are mainly detailed in **Supplementary Data 1**.

Hydrogenases and energy conversion circuits that were transcriptionally modulated by pressure were explored in detail in this article. They are developed below.

Effect of Pressure on Hydrogenases and Energy Conversion Circuits

Effect of Hydrostatic Pressure on Membrane-Bound Hydrogenases MBH and MBS

Members of the *Thermococcales* order grow heterotrophically by fermentation of peptides or sugars used as carbon and energy sources. This leads to the production of CO_2, acetate, and gas, whose composition, respectively H_2S or H_2, depends in particular on the presence or absence of elemental sulfur in the medium (Verhees et al., 2003; Sakuraba et al., 2004). The metabolism generates reduced ferredoxins that are used by membrane-bound oxidoreductases to create an ion gradient with the extracellular environment (Kengen et al., 1996; Verhees et al., 2003; Sakuraba et al., 2004). This ion gradient can further be used by an ATP synthase to produce ATP (**Figure 3**; Pisa et al., 2007). In *Thermococcales* in general and *T. piezophilus* in particular, two main systems of membrane bound oxidoreductases exist (Schut et al., 2013). The first one is the simple respiratory system MBH (membrane-bound hydrogenase) that is generally activated in the absence of S° (Sapra et al., 2003) and links the oxidation of reduced ferredoxin generated during glycolysis to the formation of H_2 (Kanai et al., 2011; Schut et al., 2012). When sulfur is provided in the medium, a second type of membrane oxidoreductase, is active: the MBS (Membrane-bound sulfane reductase) system (Wu et al., 2018), highly homologous to MBH (13 vs. 14-genes cluster). Recent studies have shown the generation of H_2S from spontaneous breakdown of tri- and

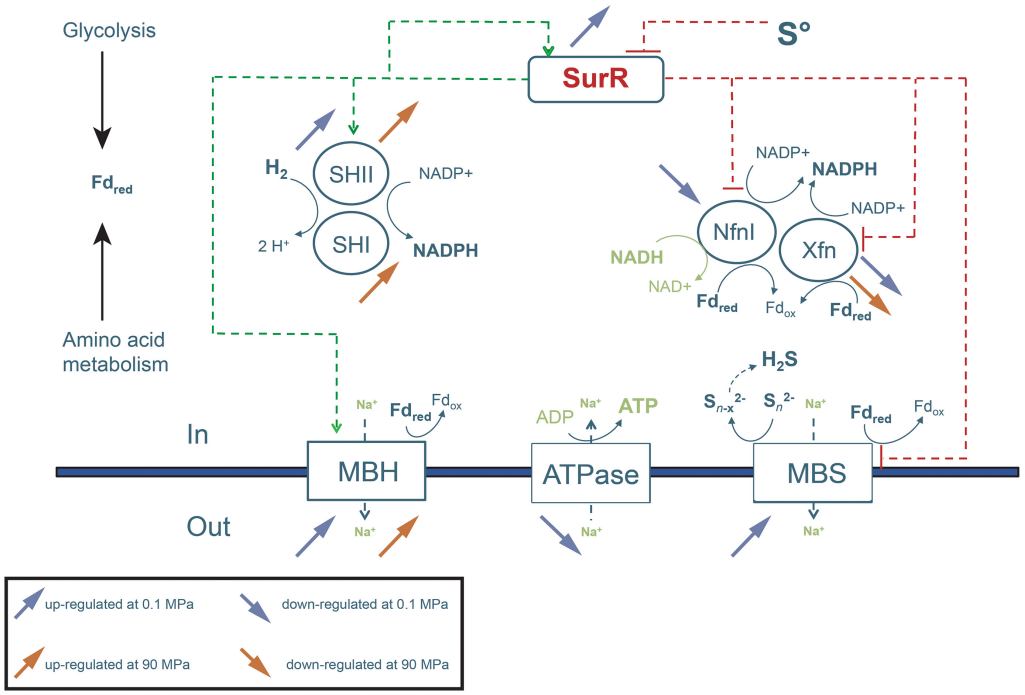


FIGURE 3 | Schematic representation of the main transcriptional metabolic regulations observed at sub-optimal (0.1 MPa) and supra-optimal (90 MPa) pressures for growth. This figure is centered on the general metabolism and focuses notably on hydrogenases and energy conversion processes and on their transcriptional regulator SurR, in the context of adaptive response to hydrostatic pressure. Blue and red arrows represent the overexpression (up direction) and subexpression (down direction) of genes at sub-optimal (0.1 MPa), and supra-optimal (90 MPa) pressures for growth, respectively, compared to the optimal pressure for growth (50 MPa).

disulfides after their release from MBS (Wu et al., 2018; Yu et al., 2020). This orientation toward the hydrogenogenic or/and the sulfidogenic pathway is orchestrated by the transcriptional sulfur-response regulator SurR (Lipscomb et al., 2009, 2017; Yang et al., 2010; Hidese et al., 2017).

Our transcriptomic study has shown that the level of expression of both MBH and MBS gene clusters is modulated by pressure (**Table 1**). This result is surprising, as *T. piezophilus* has been grown with polysulfur in all these experiments. Indeed, numerous studies on *Thermococcales* carried out at atmospheric pressure have shown that the genes coding for the MBS complex have their level of transcription increased when sulfur is present in the culture medium. On the contrary, it has been shown that the level of expression of the genes coding for the MBH complex decreases sharply in the presence of sulfur (still at atmospheric pressure) (Chou et al., 2007; Jager et al., 2014). Here, our results show that the expression level of the *Mbh* gene cluster varies despite the presence of sulfur. At the sub-optimal pressure of 0.1 MPa, the gene clusters of the MBS and MBH complexes are both overexpressed (**Table 1** and **Figures 2**, **3**), compared to 50 MPa. The genes of the *Mbh* cluster were considerably overexpressed at the supra-optimal pressure of 90 MPa compared to 50 and 0.1 MPa. It was also slightly more transcribed at 0.1 MPa compared to 50 MPa (**Table 1** and **Figures 2**, **3**). This suggests a pressure-effect overtaking a sulfur-effect on the regulation of gene expression.

Effect of Hydrostatic Pressure on Cytosolic Hydrogenases, Oxidoreductases, and Glycolytic Enzymes

Other enzymes involved in these H_2/sulfur metabolic pathways (**Figure 3**) include the cytosolic hydrogenases and oxidoreductases SHI/II, NfnI/II/III, Pdo, and TrxR, which have been shown to be under the regulation of SurR (Lipscomb et al., 2017). SHI and SHII are soluble hydrogenases, that could couple the synthesis of protons from an H_2 uptake to the regeneration of NADPH from NADP + (Schut et al., 2013). The ferredoxin:NADP + oxidoreductases (Nfn) of *Thermococcales* are involved in the electron transfer from ferredoxin to NADP +, under high concentrations of H_2 (Verhaart et al., 2010). The subsequent NADPH can be then reoxidized by a glutamate dehydrogenase and an alanine amino-transferase in the alanine biosynthetic pathway. Pdo is another glutaredoxin-like protein disulfide oxidoreductase whose roles are not totally understood. It is notably regulated by SurR and steps in the overall cellular redox balance and in the hydrogen metabolism. It could receive electrons from NADPH through a thioredoxin reductase (TrxR), and then serve as an electron carrier itself for ribonucleotide reductase (RNR), allowing links to deoxyribonucleotide synthesis (Schut et al., 2013; Demmer et al., 2015).

Here, an important overexpression of the gene cluster encoding the soluble hydrogenase SHII was observed at 0.1 and 90 MPa (**Figure 2**) compared to 50 MPa. This overexpression was more pronounced at 0.1 MPa (**Table 1**). Genes encoding the

cytoplasmic SHI hydrogenase subunits were overexpressed but only at 90 MPa compared to 50 and 0.1 MPa (**Figure 2**). These soluble hydrogenases could thus be mobilized under stressful pressure conditions to regenerate NADPH. An overexpression of the NADH-dependent ferredoxin:NADP + oxidoreductases (Nfn) was observed at 50 and 90 MPa, but not at atmospheric pressure. As with "*Thermococcus onnurineus*," *T. piezophilus* has 2 genes coding for homologs of the S and L subunits of NfnI (A7C91_RS06205-6210) of *P. furiosus* (Nguyen et al., 2017), and also possesses NfnII (A7C91_RS02315-2310) and NfnIII (A7C91_RS02425-2430). *NfnI* was overexpressed at 90 MPa vs. 50 MPa, while *NfnII* was overexpressed at 50 MPa (vs. 0.1 MPa). *NfnIII* was overexpressed at 90 MPa vs. 0.1 MPa. In addition, genes encoding a glutaredoxin-like protein (A7C91_RS07570) and the disulfide oxidoreductase Pdo (A7C91_RS07570) were overexpressed at 50 MPa. Genes coding for the thioredoxin reductase TrxR (A5C91_RS04130) and for the ribonucleotide reductase RNR (AC91_RS02975) were overexpressed at 50 and 90 MPa vs. 0.1 MPa. Altogether, these observations tend to indicate that under stressful conditions, H_2 production (with Mbh) is activated. In addition, at 0.1 MPa, the *Mbs,* and *Mbh* gene clusters are overexpressed, suggesting a cumulative effect to cope with pressure. Soundly, several genes coding for glycolytic enzymes, a pathway which regenerates reduced ferredoxin, were overexpressed at the reference pressure of 50 MPa [i.e., ADP-dependent glucokinase (A7C91_RS07115), phosphoglycerate kinase (A7C91_RS00540) and glyceraldehyde-3-phosphate dehydrogenase (A7C91_RS09215)], and also at 90 MPa [glyceraldehyde-3-phosphate dehydrogenase, glyceraldehyde-3-phosphate ferredoxin oxidoreductase (A7C91_RS04780) (GAPOR)].

Effect of Hydrostatic Pressure on Other Mrp-Mbh Oxidoreductases

In *T. piezophilus*, other types of *Mrp-Mbh* oxidoreductases clusters were also regulated by pressure. These clusters code for formate dehydrogenases (FDH) which are found only in some of the *Thermococcales*, such as *P. yayanosii* and "*T. onnurineus*" (Schut et al., 2013). Thus, three formate dehydrogenase gene clusters were found in *T. piezophilus* genome, similar to those of "*T. onnurineus*," namely the membrane-bound hydrogenase clusters *Fdh1-Mfh1-Mnh1* (A7C91_RS06945-07020) and *Fdh2-Mfh2-Mnh2* (A7C91_RS04320-04240), and the cytoplasmic hydrogenase cluster *Fdh3* (A7C91_RS08745-08770). The synteny of these clusters is conserved between both organisms. The presence of the *Fdh2-Mfh2-Mnh2* gene cluster together with the one encoding a formate transporter suggests that *T. piezophilus* could metabolize formate. This was not observed experimentally at atmospheric pressure (Dalmasso et al., 2016a), likely because growth on formate was tested at 75°C, under conditions where the reaction was probably not exergonic (Kim et al., 2010).

At 50 MPa, genes of the clusters *Fdh1-Mfh1-Mnh1* and *Fdh3* were overexpressed, as well as the associated ATP synthase clusters. Genes of the cluster *Fdh2-Mfh2-Mnh2* were also more transcribed at 50 MPa and even more again at 0.1 MPa. This might indicate that growth on formate could be stimulated under optimal and sub-optimal pressure conditions.

A cytoplasmic homolog of the F_{420}-reducing hydrogenase (A7C91_RS04340-04330) present in methanogenic archaea (*frhAGB*-encoding hydrogenase) whose expression is linked to various functions in *Thermococcus* sp., including carbon monoxide metabolism, chemotactic signal transduction and *Fdh3* regulation (Jeon et al., 2015; Lee et al., 2017), was also overexpressed at 0.1 MPa compared to 50 and to 90 MPa. This hydrogenase is involved in various cellular processes in "*T. onnurineus*," which could also be the case in *T. piezophilus*. This observation is also consistent with the fact that the cell undergoes a broad metabolic change to cope with pressure variations.

Effect of Pressure on SurR-Regulated Genes: A Synthesis

In summary, with the exception of genes of the *Mrp-Mbs* cluster, genes encoding enzymes which are known to be under SurR negative control, namely *Pdo* and *Nfn*, were all overexpressed at the reference pressure of 50 MPa vs. 0.1 MPa (in presence of sulfur). Interestingly, the genes and gene clusters that were under SurR positive control were overexpressed at supra- and sometimes sub-optimal pressures for growth (see above). This suggests that, in *T. piezophilus*, the expression patterns of genes known to be under the control of the SurR regulon appear to follow a classical response/behavior to sulfur only under optimal pressure conditions. Sub-optimal pressures and, to a lesser extent, supra-optimal pressures are likely to influence the expression of these genes in a new way of regulation (**Figure 3**). As an illustration, a cooperation between the membrane complexes MBH and MBS appears to occur at atmospheric pressure regardless of the presence of sulfur.

CONCLUSION

This study presents the transcriptional adaptive strategy implemented by *T. piezophilus*, the piezophilic microorganism with the widest range of tolerance to HHP (0.1–130 MPa) known to date. The overall differential gene expression revealed by RNA-seq confirms the results previously obtained with *T. barophilus* and *P. yayanosii* (Vannier et al., 2015; Michoud and Jebbar, 2016) that there is an overall metabolic change rather than a classical stress response. A thorough knowledge of the process of energy-conservation and its regulation in other *Thermococcales* models, has enabled us to highlight the involvement of these processes in the adaptive response, at least in the regulation of the expression of genes coding for the proteins involved in this process. These results are changing our knowledge of the mechanism of regulation of this energy-conservation process under conditions of non-optimal pressures and of the role of SurR under these conditions.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

KA and CD designed the experiment. CD and MG carried out the experiments. YM and DC performed the bioinformatics analyses. KA, YM, and JH interpreted the data and wrote the manuscript. PO, MJ, and ZS helped supervise the project. All authors reviewed the manuscript.

FUNDING

This work was supported by the program MERLIN Abyss to KA, the IRP 1211 MicrobSea to KA and ZS, and the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and the program "Investissements d'Avenir" to KA.

ACKNOWLEDGMENTS

We are grateful to the operation teams of DSV *Shinkai 6500* and the crews of RV *Yokosuka* for sample collection. We also thank two reviewers whose insightful comments have greatly improved this contribution.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.730231/full#supplementary-material

REFERENCES

Amrani, A., Bergon, A., Holota, H., Tamburini, C., Garel, M., Ollivier, B., et al. (2014). Transcriptomics reveal several gene expression patterns in the piezophile *Desulfovibrio hydrothermalis* in response to hydrostatic pressure. *PLoS One* 9:e106831. doi: 10.1371/journal.pone.0106831

Cario, A., Grossi, V., Schaeffer, P., and Oger, P. M. (2015a). Membrane homeoviscous adaptation in the piezo-hyperthermophilic archaeon Thermococcus barophilus. *Front. Microbiol.* 6:1152. doi: 10.3389/fmicb.2015.01152

Cario, A., Jebbar, M., Thiel, A., Kervarec, N., and Oger, P. M. (2016). Molecular chaperone accumulation as a function of stress evidences adaptation to high hydrostatic pressure in the piezophilic archaeon Thermococcus barophilus. *Sci. Rep.* 6:29483.

Cario, A., Lormieres, F., Xiang, X., and Oger, P. (2015b). High hydrostatic pressure increases amino acid requirements in the piezo-hyperthermophilic archaeon Thermococcus barophilus. *Res. Microbiol.* 166, 710–716. doi: 10.1016/j.resmic.2015.07.004

Chou, C. J., Shockley, K. R., Conners, S. B., Lewis, D. L., Comfort, D. A., Adams, M. W., et al. (2007). Impact of substrate glycoside linkage and elemental sulfur on bioenergetics of and hydrogen production by the hyperthermophilic archaeon Pyrococcus furiosus. *Appl. Environ. Microbiol.* 73, 6842–6853. doi: 10.1128/aem.00597-07

Dalmasso, C., Oger, P., Courtine, D., Georges, M., Takai, K., Maignien, L., et al. (2016a). Complete genome sequence of the hyperthermophilic and piezophilic archeon thermococcus piezophilus CDGST, able to grow under extreme hydrostatic pressures. *Genome Announc.* 4:e00610-16. doi: 10.1128/genomeA.00610-16

Dalmasso, C., Oger, P., Selva, G., Courtine, D., Lharidon, S., Garlaschelli, A., et al. (2016b). Thermococcus piezophilus sp. nov., a novel hyperthermophilic and piezophilic archaeon with a broad pressure range for growth, isolated from a deepest hydrothermal vent at the Mid-Cayman Rise. *Syst. Appl. Microbiol.* 39, 440–444. doi: 10.1016/j.syapm.2016.08.003

Demmer, J. K., Huang, H., Wang, S., Demmer, U., Thauer, R. K., and Ermler, U. (2015). Insights into flavin-based electron bifurcation via the NADH-dependent Reduced Ferredoxin:NADP oxidoreductase structure. *J. Biol. Chem.* 290, 21985–21995. doi: 10.1074/jbc.M115.656520

Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. doi: 10.7717/peerj.1319

Gindner, A., Hausner, W., and Thomm, M. (2014). The TrmB family: a versatile group of transcriptional regulators in Archaea. *Extremophiles* 18, 925–936. doi: 10.1007/s00792-014-0677-2

Hidese, R., Yamashita, K., Kawazuma, K., Kanai, T., Atomi, H., Imanaka, T., et al. (2017). Gene regulation of two ferredoxin:NADP(+) oxidoreductases by the redox-responsive regulator SurR in Thermococcus kodakarensis. *Extremophiles* 21, 903–917. doi: 10.1007/s00792-017-0952-0

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Jager, D., Forstner, K. U., Sharma, C. M., Santangelo, T. J., and Reeve, J. N. (2014). Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomic.* 15:684. doi: 10.1186/1471-2164-15-684

Jannasch, H. W., and Taylor, C. D. (1984). Deep-sea microbiology. *Ann. Rev. Microbiol.* 38, 487–514.

Jebbar, M., Franzetti, B., Girard, E., and Oger, P. (2015). Microbial diversity and adaptation to high hydrostatic pressure in deep-sea hydrothermal vents prokaryotes. *Extremophiles* 19, 721–740. doi: 10.1007/s00792-015-0760-3

Jeon, J. H., Lim, J. K., Kim, M.-S., Yang, T.-J., Lee, S.-H., Bae, S. S., et al. (2015). Characterization of the frhAGB-encoding hydrogenase from a non-methanogenic hyperthermophilic archaeon. *Extremophiles* 19, 109–118. doi: 10.1007/s00792-014-0689-y

Kanai, T., Matsuoka, R., Beppu, H., Nakajima, A., Okada, Y., Atomi, H., et al. (2011). Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon Thermococcus kodakarensis. *J. Bacteriol.* 193, 3109–3116. doi: 10.1128/JB.01072-10

Karr, E. A. (2014). "Chapter Three - Transcription Regulation in the Third Domain," in *Advances in Applied Microbiology*, eds S. Sariaslani and G. M. Gadd (Cambridge: Academic Press), 101–133. doi: 10.1016/b978-0-12-800259-9.00003-2

Kengen, S. M., Stams, A. J. M., and De Vos, W. M. (1996). Sugar metabolism of hyperthermophiles. *FEMS Microbiol. Rev.* 18, 119–137. doi: 10.1111/j.1574-6976.1996.tb00231.x

Kim, Y. J., Lee, H. S., Kim, E. S., Bae, S. S., Lim, J. K., Matsumi, R., et al. (2010). Formate-driven growth coupled with H_2 production. *Nature* 467:352. doi: 10.1038/nature09375

Krug, M., Lee, S.-J., Boos, W., Diederichs, K., and Welte, W. (2013). The three-dimensional structure of TrmB, a transcriptional regulator of dual function in the hyperthermophilic archaeon Pyrococcus furiosus in complex with sucrose. *Protein Sci.* 22, 800–808. doi: 10.1002/pro.2263

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Lee, S. H., Kim, M. S., Kim, Y. J., Kim, T. W., Kang, S. G., and Lee, H. S. (2017). Transcriptomic profiling and its implications for the H2 production of a non-methanogen deficient in the frhAGB-encoding hydrogenase. *Appl. Microbiol. Biotechnol.* 101, 5081–5088. doi: 10.1007/s00253-017-8234-4

Lee, S.-J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., et al. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis. *J. Biol. Chem.* 278, 983–990. doi: 10.1074/jbc.m210236200

Lee, S.-J., Moulakakis, C., Koning, S. M., Hausner, W., Thomm, M., and Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* 57, 1797–1807.

Lee, S. J., Surma, M., Hausner, W., Thomm, M., and Boos, W. (2008). The role of TrmB and TrmB-like transcriptional regulators for sugar transport and metabolism in the hyperthermophilic archaeon Pyrococcus furiosus. *Arch. Microbiol.* 190:247.

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

Lipscomb, G. L., Keese, A. M., Cowart, D. M., Schut, G. J., Thomm, M., Adams, M. W., et al. (2009). SurR: a transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in Pyrococcus furiosus. *Mol. Microbiol.* 71, 332–349. doi: 10.1111/j.1365-2958.2008.06525.x

Lipscomb, G. L., Schut, G. J., Scott, R. A., and Adams, M. W. W. (2017). SurR is a master regulator of the primary electron flow pathways in the order Thermococcales. *Mol. Microbiol.* 104, 869–881. doi: 10.1111/mmi.13668

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Martinez, N., Michoud, G., Cario, A., Ollivier, J., Franzetti, B., Jebbar, M., et al. (2016). High protein flexibility and reduced hydration water dynamics are key pressure adaptive strategies in prokaryotes. *Sci. Rep.* 6:32816. doi: 10.1038/srep32816

Michoud, G., and Jebbar, M. (2016). High hydrostatic pressure adaptive strategies in an obligate piezophile Pyrococcus yayanosii. *Sci. Rep*. 6:27289. doi: 10.1038/srep27289

Nguyen, D. M. N., Schut, G. J., Zadvornyy, O. A., Tokmina-Lukaszewska, M., Poudel, S., Lipscomb, G. L., et al. (2017). Two functionally distinct NADP(+)-dependent ferredoxin oxidoreductases maintain the primary redox balance of Pyrococcus furiosus. *J. Biol. Chem.* 292, 14603–14616. doi: 10.1074/jbc.M117.794172

Oger, P. M., and Jebbar, M. (2010). The many ways of coping with pressure. *Res. Microbiol.* 161, 799–809. doi: 10.1016/j.resmic.2010.09.017

Picard, A., and Daniel, I. (2013). Pressure as an environmental parameter for microbial life–a review. *Biophys. Chem.* 183, 30–41. doi: 10.1016/j.bpc.2013.06.019

Pisa, K. Y., Huber, H., Thomm, M., and Müller, V. (2007). A sodium ion-dependent A1AO ATP synthase from the hyperthermophilic archaeon Pyrococcus furiosus. *Febs J.* 274, 3928–3938. doi: 10.1111/j.1742-4658.2007.05925.x

Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Sakuraba, H., Goda, S., and Ohshima, T. (2004). Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. *Chem. Rec.* 3, 281–287. doi: 10.1002/tcr.10066

Sapra, R., Bagramyan, K., and Adams, M. W. (2003). A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. USA* 100, 7545–7550. doi: 10.1073/pnas.1331436100

Schut, G. J., Boyd, E. S., Peters, J. W., and Adams, M. W. (2013). The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol. Rev.* 37, 182–203. doi: 10.1111/j.1574-6976.2012.00346.x

Schut, G. J., Nixon, W. J., Lipscomb, G. L., Scott, R. A., and Adams, M. W. (2012). Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon pyrococcus furiosus. *Front. Microbiol.* 3:163. doi: 10.3389/fmicb.2012.00163

Siliakus, M. F., Van Der Oost, J., and Kengen, S. W. M. (2017). Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* 21, 651–670. doi: 10.1007/s00792-017-0939-x

Simonato, F., Campanaro, S., Lauro, F. M., Vezzi, A., Dangelo, M., Vitulo, N., et al. (2006). Piezophilic adaptation: a genomic point of view. *J. Biotechnol.* 126, 11–25. doi: 10.1016/j.jbiotec.2006.03.038

Vannier, P., Michoud, G., Oger, P., Marteinsson, V., and Jebbar, M. (2015). Genome expression of Thermococcus barophilus and Thermococcus kodakarensis in response to different hydrostatic pressure conditions. *Res. Microbiol.* 166, 717–725. doi: 10.1016/j.resmic.2015.07.006

Verhaart, M. R., Bielen, A. A., Van Der Oost, J., Stams, A. J., and Kengen, S. W. (2010). Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. *Environ. Technol.* 31, 993–1003. doi: 10.1080/09593331003710244

Verhees, C. H., Kengen, S. W., Tuininga, J. E., Schut, G. J., Adams, M. W., De Vos, W. M., et al. (2003). The unique features of glycolytic pathways in Archaea. *Biochem. J.* 375, 231–246. doi: 10.1042/bj20021472

Wu, C. H., Schut, G. J., Poole, F. L. II, Haja, D. K., and Adams, M. W. W. (2018). Characterization of membrane-bound sulfane reductase: A missing link in the evolution of modern day respiratory complexes. *J. Biol. Chem.* 293, 16687–16696. doi: 10.1074/jbc.ra118.005092

Yang, H., Lipscomb, G. L., Keese, A. M., Schut, G. J., Thomm, M., Adams, M. W., et al. (2010). SurR regulates hydrogen production in Pyrococcus furiosus by a sulfur-dependent redox switch. *Mol. Microbiol.* 77, 1111–1122.

Yayanos, A. A. (1995). Microbiology to 10,500 meters in the deep-sea. *Ann. Rev. Microbiol.* 49, 777–805. doi: 10.1146/annurev.mi.49.100195.004021

Yu, H., Haja, D. K., Schut, G. J., Wu, C.-H., Meng, X., Zhao, G., et al. (2020). Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat. Commun.* 11:5953.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Moalic, Hartunians, Dalmasso, Courtine, Georges, Oger, Shao, Jebbar and Alain. 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 November 2021
doi: 10.3389/fmicb.2021.751746

Genomic Insights Into New Species of the Genus *Halomicroarcula* Reveals Potential for New Osmoadaptative Strategies in Halophilic Archaea

Ana Durán-Viseras, Cristina Sánchez-Porro and Antonio Ventosa**

Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain

OPEN ACCESS

Edited by:
André Antunes, Macau University of Science and Technology, China

Reviewed by:
Heng-Lin Cui, Jiangsu University, China
Erhard Bremer, University of Marburg, Germany

***Correspondence:**
Ana Durán-Viseras
anaduran@us.es
Antonio Ventosa
ventosa@us.es

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

Received: *01 August 2021*
Accepted: *11 October 2021*
Published: *04 November 2021*

Citation:
Durán-Viseras A, Sánchez-Porro C and Ventosa A (2021) Genomic Insights Into New Species of the Genus Halomicroarcula Reveals Potential for New Osmoadaptative Strategies in Halophilic Archaea. Front. Microbiol. 12:751746. doi: 10.3389/fmicb.2021.751746

Metagenomic studies on prokaryotic diversity of hypersaline soils from the Odiel saltmarshes, South-west Spain, revealed a high proportion of genomic sequences not related to previously cultivated taxa, that might be related to haloarchaea with a high environmental and nutritional flexibility. In this study, we used a culturomics approach in order to isolate new haloarchaeal microorganisms from these hypersaline soils. Four haloarchaeal strains, designated strains F24A^T, F28, F27^T, and F13^T, phylogenetically related to the genus *Halomicroarcula*, were isolated and characterized in detail. The phylogenomic tree based on the 100 orthologous single-copy genes present in the genomes of these four strains as well as those of the type strains of the species *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula limicola* JCM 18640^T, that were determined in this study, revealed that these four new isolates clustered on three groups, with strains F24A^T and F28 within a single cluster, and altogether with the species of *Halomicroarcula*. Additionally, Orthologous Average Nucleotide Identity (OrthoANI), digital DNA-DNA hybridization (dDDH) and Average Amino-acid Identity (AAI) values, likewise phenotypic characteristics, including their polar lipids profiles, permitted to determine that they represent three new species, for which we propose the names *Halomicroarcula rubra* sp. nov. (type strain F13^T), *Halomicroarcula nitratireducens* sp. nov. (type strain F27^T) and *Halomicroarcula salinisoli* sp. nov. (type strain F24A^T). An in deep comparative genomic analysis of species of the genus *Halomicroarcula*, including their metabolism, their capability to biosynthesize secondary metabolites and their osmoregulatory adaptation mechanisms was carried out. Although they use a *salt-in* strategy, the identification of the complete pathways for the biosynthesis of the compatible solutes trehalose and glycine betaine, not identified before in any other haloarchaea, might suggest alternative osmoadaptation strategies for this group. This alternative osmoregulatory mechanism would allow this group of haloarchaea to be

versatile and eco-physiologically successful in hypersaline environments and would justify the capability of the species of this genus to grow not only on environments with high salt concentrations [up to 30% (w/v) salts], but also under intermediate to low salinities.

Keywords: *Halomicroarcula*, haloarchaea, comparative genomic analysis, compatible solutes, *Halomicroarcula rubra* sp. nov., *Halomicroarcula nitratireducens* sp. nov., *Halomicroarcula salinisoli* sp. nov.

INTRODUCTION

Hypersaline environments are extreme habitats that have permitted the isolation and characterization of many halophilic microorganisms that have been used as models for the study of basic and applied purposes (Ventosa, 2006; Ventosa et al., 2015). These habitats are characterized by high levels of salinity, frequently accompanied by other extreme conditions, such as high or low temperature or pH values, UV radiation, hydrostatic pressure or high concentrations of toxic compounds (Rodríguez-Valera, 1988; Ventosa, 2006). To withstand these high salt concentrations and the constant fluctuations in salinity levels, halophilic microorganisms have developed diverse physiological adaptations (Galinski and Trüper, 1994; Kempf and Bremer, 1998; Gunde-Cimerman et al., 2018). The best adapted organisms to these high salt concentrations are prokaryotes belonging to the extremely halophilic archaea, members of the class *Halobacteria* (also called haloarchaea) (Amoozegar et al., 2017; Oren and Ventosa, 2017a). Although haloarchaea have been traditionally considered as a coherent group limited to use a *salt-in* osmoadaptation strategy, i.e., the accumulation of K^+, Na^+, and Cl^- ions, recent studies have brought to light their potential to employ additional alternative mechanisms (Anderson et al., 2011; Youssef et al., 2014).

Hypersaline environments are represented by a wide range of habitats, from aquatic or terrestrial to deep-sea, salt mines, salt-cured food or plants (Ventosa, 2006). However, hypersaline aquatic systems and more recently saline soils constitute the most extensively studied hypersaline environments (Ventosa et al., 2015). Traditionally, aquatic environments have been thoroughly studied, specially saline lakes and salterns (Ventosa et al., 2014, 2015; Oren, 2015), but the number of studies in hypersaline soils is much more reduced (Ventosa et al., 2008; Oren, 2011). Recent metagenomic studies carried out on hypersaline soils from the Odiel saltmarshes (South-west Spain) revealed a high proportion of genomic sequences not related to any cultivated organisms as well as a high environmental and nutritional flexibility of microorganism inhabiting these systems (Vera-Gargallo and Ventosa, 2018; Vera-Gargallo et al., 2019). With the purpose of isolating such of these metabolically diverse groups not isolated to date, we focused on these habitats by using culture-dependent methods. For that purpose, we used the current "culturomics" approach (Durán-Viseras et al., 2019a, 2021), in order to isolate new haloarchaeal groups taking advantage of the previous metagenomic studies, that showed an unexpected large proportion of haloarchaea in these hypersaline soils (Vera-Gargallo and Ventosa, 2018). As a result, four haloarchaeal strains phylogenetically related to the genus *Halomicroarcula* were isolated in pure culture.

The genus *Halomicroarcula* belongs to the family *Haloarculaceae*, within the order *Halobacteriales*, class *Halobacteria* and phylum *Euryarchaeota* (Echigo, 2016; Oren and Ventosa, 2017b). At the time of writing the genus *Halomicroarcula* comprises four species: *Halomicroarcula pellucida* (Echigo et al., 2013), which is the type species of the genus and was isolated from French marine salt, designated "Sel marin de Guérande," *Halomicroarcula limicola* (Zhang and Cui, 2014) and *Halomicroarcula salina* (Zhang and Cui, 2015), both isolated from Yinggehai solar saltern, located in Hainan Province, China; and more recently, *Halomicroarcula amylolytica* (Chen et al., 2020), isolated from a salt mine in Yunnan Province, China. The genus *Halomicroarcula* includes Gram-stain-negative, motile and pleomorphic cells. They form non-pigmented and transparent or red-pigmented colonies. They are halophilic, neutrophilic and mesophilic. Other characteristics of this genus are their aerobic and heterotrophic metabolism (Echigo et al., 2013). Their major polar lipids are phosphatidylglycerol (PG), phosphatidylglycerolphosphate methyl ester (PGP-Me) and phosphatidylglycerol sulfate (PGS). Glycolipids, including sulfated mannosyl glucosyl diether (S-DGD-1) and mannosyl glucosyl diether (DGD-1) may be present in some species (Zhang and Cui, 2014; Echigo, 2016). The G + C content of strains of the genus *Halomicroarcula* range between 64.0 and 64.5 mol% (Echigo et al., 2013; Zhang and Cui, 2014; Echigo, 2016).

In this work, we have carried out an exhaustive taxogenomic study of the genus *Halomicroarcula*, and we have addressed the description of three new species of the genus *Halomicroarcula*. Besides, we have performed a comparative genomic analysis of species of this genus aimed at studying in depth their metabolism, their capacity to biosynthesize secondary metabolites and their osmoregulatory adaptation mechanisms that allow this group of microorganisms to be versatile and eco-physiologically successful in hypersaline environments.

MATERIALS AND METHODS

Strains Isolation, Reference Strains and Culture Conditions

Strains F24A^T, F28, F27^T, and F13^T were isolated from saline soil samples (conductivity 54.5 $CE_{1:5}$ mS/cm and pH 8.9) collected from the Odiel saltmarshes (Huelva, South-west Spain) (37°12′26″N 6°57′58″O). Samples were diluted, plated

under sterile conditions and incubated at 37°C up to two months. Strains F24A^T, F28, F27^T, and F13^T were isolated and routinely grown in R2A medium (Difco) (pH adjusted to 7.5) supplemented with 25% (w/v) seawater salt solution prepared by dilution of SW 30% (w/v) stock solution containing (g/L): NaCl, 195; $MgCl_2 \cdot 6H_2O$, 32.5; $MgSO_4 \cdot 7H_2O$, 50.8; $CaCl_2$, 0.83; KCl, 5.0; $NaHCO_3$, 0.17; NaBr, 0.58. Purified agar (2%; Oxoid) was used as solidifying agent, when needed. For long term maintenance, cultures were preserved at –80°C in this medium containing 20% (v/v) glycerol (Durán-Viseras et al., 2021).

For taxonomic comparative purposes strains *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula limicola* JCM 18640^T, obtained from culture collections, were used in this study. These strains were also grown in the same medium and culture conditions as described above.

DNA Extraction, Purification and Sequencing

The genomic DNA of strains F24A^T, F28, F27^T, F13^T, *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula limicola* JCM 18640^T was extracted and purified using the method described by Marmur (1961). DNA quality and concentration was checked by spectrophotometry (DeNovix DS-11 FX, DeNovix Technologies, Wilmington, Delaware, United States) and fluorometry (Qubit 3.0 Fluorometer, Thermofisher Scientific, United States). The 16S rRNA gene was amplified by PCR using the universal primers ArchF and ArchR (DeLong, 1992; Arahal et al., 1996), then cloned and sequenced as described previously (Durán-Viseras et al., 2019b). The *rpoB′* gene was amplified by PCR using the primers rpoBF and rpoBR (Fullmer et al., 2014). The genome of strains F13^T, F24A^T, F27^T, F28, *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula limicola* JCM 18640^T and *Halomicroarcula salina* JCM 18369^T were sequenced using the Illumina HiSeq 4000 platform at StabVida (Oeiras, Portugal) and Novogene (Cambridge United Kingdom).

Phylogenetic and Phylogenomic Analyses

Identification of phylogenetic neighbors and calculation of pairwaise 16S rRNA and *rpoB′* gene sequences similarities were conducted by using the Ez-BioCloud server (Yoon et al., 2017) and BLAST (Altschul et al., 1990), respectively. Additional 16S rRNA and *rpoB′* gene sequences, and genomes used for comparisons were retrieved from the GenBank/EMBL/DDBJ databases. Clustering of 16S rRNA and *rpoB′* gene sequences were determined using the neighbor-joining (Saitou and Nei, 1987) and maximum-likelihood (Felsenstein, 1981) algorithms implemented in the MEGA-X software (Kumar et al., 2018), using the Jukes-Cantor method (Jukes and Cantor, 1969) for evolutionary distances calculation. For phylogenetic tree branch support estimation, a bootstrap analysis based on 1000 replications was calculated (Felsenstein, 1985).

For phylogenomic analyses, core orthologous genes were identified from all analyzed genomes by an all-versus-all BLAST as implemented in the Enveomics collection toolbox (Rodriguez-R and Konstantinidis, 2016). As a result, a set of 100 conserved genes were retrieved and aligned using MUSCLE (Edgar, 2004). Phylogenomic tree of concatenated sequences was reconstructed by using the software MEGA-X (Kumar et al., 2018) with the neighbor-joining and maximum-likelihood methodology and Jukes-Cantor correction.

Genome Assembly, Annotation and Comparative Genomics

The *de novo* assembly of the reads was performed using Spades 3.9.1 (Bankevich et al., 2012). Draft genome assembly was quality checked using Quast v2.3 (Gurevich et al., 2013). Genome completeness, contamination and strain heterogeneity was estimated using CheckM v1.0.5 (Parks et al., 2015).

Genome sequences were annotated with Prokka (Seemann, 2014). BlastKOALA (Kanehisa et al., 2016) was used to assign KO identifiers (K numbers) to orthologous genes present in the genomes and mapped to the KEGG pathways and KEGG modules to perform the metabolic pathways reconstructions.

CRISPR/Cas systems, prophage sequences and secondary metabolites were identified by the tools CRISPRCasFinder (Couvin et al., 2018), PHASTER (Zhou et al., 2011; Arndt et al., 2016) and antiSMASH 5.0 (Blin et al., 2019), respectively. Isoelectric points of predicted proteins were computed using the iep program from EMBOSS package (Rice et al., 2000). The Orthologous Average Nucleotide Identity (ANI), Average Amino-acid Identity (AAI), and digital DNA-DNA hybridization (dDDH) were calculated using the tools OAT v0.93.1 (Lee et al., 2016), AAI-Matrix (Rodriguez-R and Konstantinidis, 2016) and Genome-to-Genome Distance Calculator (GGDC) (Meier-Kolthoff et al., 2013), respectively.

Phenotypic and Chemotaxonomic Characterization

Phenotypic features of strains F24A^T, F28, F27^T, and F13^T were performed according to the minimal standards established for the taxonomic description of novel taxa of the class *Halobacteria* (Oren et al., 1997) and following the methodology previously described by Durán-Viseras et al., 2020. *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula limicola* JCM 18640^T were used as reference strains for taxonomic comparisons.

Comparative polar lipids analysis of strains F24A^T, F28, F27^T, and F13^T were determined by High-Performance Thin Layer Chromatography (HPTLC) as described elsewhere (Corral et al., 2015, 2016), using as spray reagents 5% H_2SO_4 (in water) or molybdenum blue. In this case, strains *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T, *Halomicroarcula limicola* JCM 18640^T, *Halobacterium salinarum* DSM 3754^T and *Halorubrum saccharovorum* DSM 1137^T were used as reference species for polar lipids characterization. The polar lipids of strains F24A^T, F28, F27^T, F13^T, *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula limicola* JCM 18640^T were obtained from

biomass cultured in R2A medium supplemented with 25% (w/v) seawater salt solution and pH adjusted to 7.5.

RESULTS AND DISCUSSION

Phylogenetic and Phenotypic Analyses

Previous metagenomic studies on prokaryotic diversity of hypersaline soils showed a large proportion of haloarchaea, with a high percentage not closely related to any previously described taxa. For that reason, we isolated a large collection of strains from the Odiel saltmarsh hypersaline soils located in Huelva, South-west Spain, using different complex oligotrophic media and culture conditions, based on culturomics techniques as previously detailed (Durán-Viseras et al., 2021). Partial sequencing of the 16S rRNA gene permitted us to preliminary delineate the taxonomic position of the isolates. For this study we selected four new isolates, designated as strains F24A^T, F28, F27^T, and F13^T, that were identified as members of the genus *Halomicroarcula*.

According to current practice, we determined the almost complete 16S rRNA gene sequences of strains F24A^T, F28, F27^T, and F13^T. Similarly to other species of the genus *Halomicroarcula*, the four new strains showed two different copies of the 16S rRNA gene (designated as *rrnA* and *rrnB*), with sizes of 1441 bp and 1441 bp (for strain F13^T), 1441 bp and 1442 bp (for strain F27^T), 1446 bp and 1446 bp (for strain F24A^T), and 1447 bp and 1446 bp (for strain F28), respectively. Their percentages of similarity with the type strains of species of *Halomicroarcula* (detailed in **Supplementary Table 1**), were in all cases equal or lower than 99.2% with *Halomicroarcula limicola* YGHS32^T, 96.5% with *Halomicroarcula pellucida* BNERC31^T and 96.0% with *Halomicroarcula salina* YGHS18^T. The phylogenetic tree generated on the basis of the 16S rRNA gene showed that they clustered within the *Halomicroarcula* branch (**Supplementary Figure 1A**). However, since this gene has been proved to be not very useful as a phylogenetic marker for haloarchaea, we also sequenced the *rpoB′* gene, which has been recommended as an alternative for single-gene phylogenetic analyses (Enache et al., 2007; Minegishi et al., 2010). The phylogenetic tree obtained by the neighbor-joining method (**Supplementary Figure 1B**) showed a similar topology with the 16S rRNA gene tree, in which the four new strains were placed on the *Halomicroarcula* cluster, with strains F24A^T and F28 clustering together.

In order to determine in more detail the phylogenomic relationships of these four new isolates with respect to the species of *Halomicroarcula* and according to the Minimal Standards recommendations (Chun et al., 2018), we sequenced their genomes as well as those of the type strains of the species *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula limicola* JCM 18640^T. Additionally, the genome of *Halomicroarcula amylolytica* LR21^T, previously sequenced, was also included in this study. The main characteristics of these genomes are detailed in **Table 1**. All these genomic features are in accordance with the Minimal Standards established for the use of genomic data in prokaryotic taxonomy (Chun et al., 2018). The phylogenomic tree based on 100 orthologous single-copy genes present in all the genomes under study is shown in **Figure 1**. Bootstrap values of 100% in all branches related to *Halomicroarcula* supported this tree, generated by the neighbor-joining algorithm. This tree shows that the four new strains cluster within the *Halomicroarcula* branch, with strains F27^T and F13^T most closely related to *Halomicroarcula limicola* JCM 18640^T and *Halomicroarcula pellucida* CECT 7537^T, respectively, while strains F24A^T and F28 clustered very close each other, suggesting that they may be members of the same species. Besides, this tree shows that the type strain of the species *Halomicroarcula salina* JCM 18639^T clustered with the genus *Haloarcula*, suggesting its taxonomic position should be revised.

Currently, several Overal Genome Relatedness Indexes (OGRI) are used in order to measure similarities between genome sequences (Chun and Rainey, 2014). Many algorithms have been proposed for this purpose, but the three most widely used for taxonomic purposes are Orthologous Average Nucleotide Identity (OrthoANI) (Lee et al., 2016), and digital DNA-DNA hybridization (dDDH) (Meier-Kolthoff et al., 2013), especially useful for delineation at the species level, and Average Amino-acid Identity (AAI) (Rodriguez-R and Konstantinidis, 2016), for genus delineation. It is widely accepted that for delineation of prokaryotic species the threshold values for OrthoANI and dDDH are 95% and 70%, respectively (Konstantinidis and Tiedje, 2005; Auch et al., 2010; Chun and Rainey, 2014). However, for the delineation at the genus level there is no clear universal AAI cutoff value, with approximately 65% as a reference percentage (Konstantinidis et al., 2017). **Figure 2** shows the OrthoANI and dDDH percentages calculated for all pairs of strains F24A^T, F28, F27^T, F13^T, *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T, *Halomicroarcula limicola* JCM 18640^T and *Halomicroarcula amylolytica* LR21^T, as well as for the type strains of other related haloarchaea. OrthoANI and dDDH values between strains F24A^T and F28 were 99.2% and 94.5%, respectively, confirming unequivocally that these two strains are members of the same species. On the other hand, their percentages with respect to the strains F27^T and F13^T and the type strains of species of *Halomicroarcula* and to other haloarchaea are equal or lower than 80.8% and 29.5%, respectively (**Figure 2**), which clearly indicates that they constitute a separated species of the genus *Halomicroarcula*. With respect to strains F27^T and F13^T, their percentages of OrthoANI and dDDH with the most closely related species of *Halomicroarcula* are equal or lower than 92.3% and 52.6%, respectively (for strain F27^T) and 88.7% and 41.3%, respectively (for strain F13^T) (**Figure 2**), all of them values lower than those defined for the delineation of species, and thus supporting the new species status for both new isolates.

Concerning the AAI values between the four new strains and members of the genus *Halomicroarcula* and other related genera, they clearly confirm that these four new strains are members of the genus *Halomicroarcula*. Percentages of AAI of strains F24A^T, F28, F27^T, and F13^T among themselves and the species of *Halomicroarcula* are equal or higher than 71.3%, 71.5%, 72.7%, and 74.3%, respectively, values higher than the 65% generally

TABLE 1 | Main features of the sequenced genomes of strains F13^T, F24A^T, F28, F27^T, and the type strains of species of *Halomicroarcula* used in this study.

Genomic feature	Strain F13^T	Strain F24A^T	Strain F28	Strain F27^T	*Halomicroarcula limicola* JCM 18640^T	*Halomicroarcula pellucida* CECT 7537^T	*Halomicroarcula salina* JCM 18369^T	*Halomicroarcula amylolytica* LR21^T
Size (Mb)	4.7	4.0	4.1	5.2	3.9	3.9	3.8	5.0
Contigs	35	9	12	63	5	8	10	34
N50 (bp)	308161	1343505	742232	450140	1379950	1010699	594909	1435272
Completeness (%)	99.5	99.5	99.5	99.5	99.1	99.5	100	99.5
CDS	4781	4116	4189	5281	3888	4025	3827	5040
rRNA	6	5	3	3	4	3	2	4
tRNA	51	48	46	54	54	49	48	46
G + C (mol%)	64.4	64.1	63.9	63.2	65.9	65.5	66.7	62.0
Accession number	RKLR00000000	RKLQ00000000	RKLS00000000	RKLT00000000	JAHQXF000000000	RKLW00000000	JAHQXE000000000	SRIF00000000

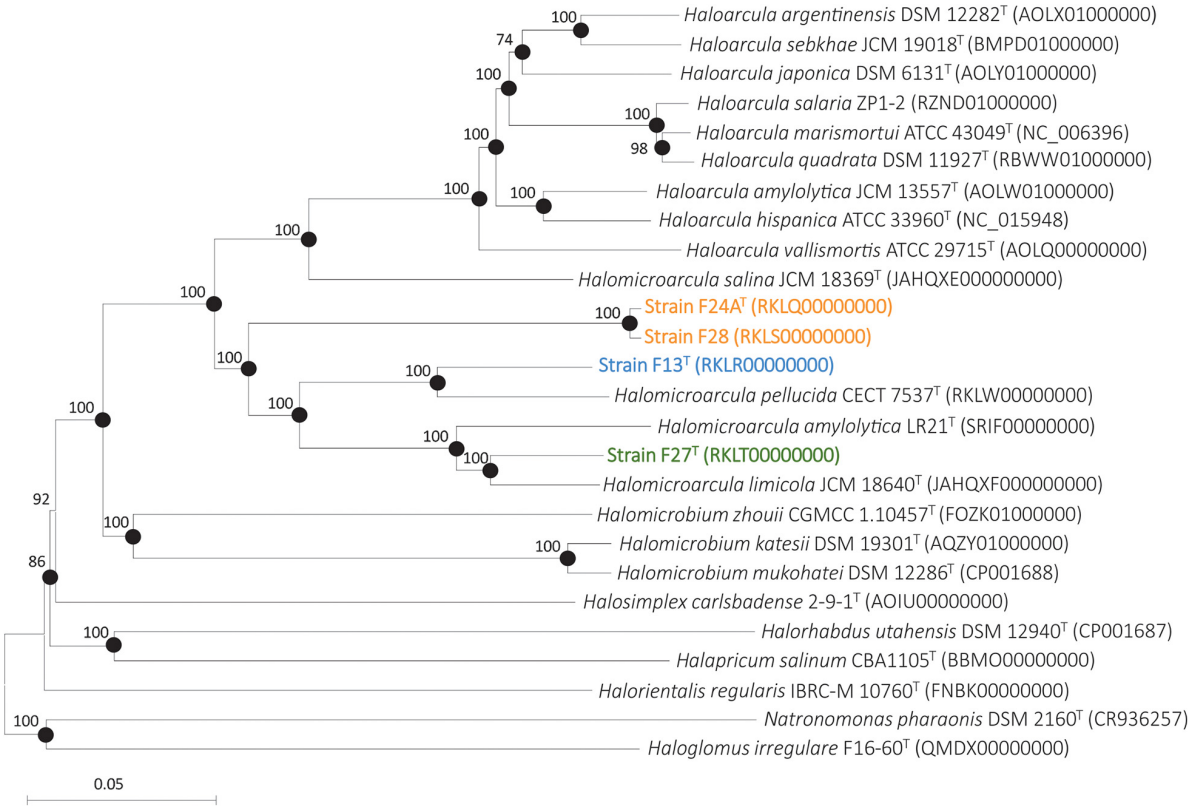


FIGURE 1 | Neighbor-joining phylogenomic tree based on the concatenation of 100 orthologous single-copy genes shared by strains F13^T, F24A^T, F27^T, F28, members of the genus *Halomicroarcula* and other related genera. Filled circles indicate branches that were also supported by the maximum-likelihood algorithm. Sequence accession number are shown in parentheses. Bootstrap values ≥70% are shown at branch points. Bar, 0.05 changes per nucleotide position.

accepted for the delineation at the genus level. It is noticeable that the AAI percentages for the new isolates and current species of the genus *Halomicroarcula* with *Haloarcula vallismortis* (type species of the genus *Haloarcula*) are 71.0%, 71.0%, 71.5%, and 73.1% for strains F24A^T, F28, F27^T, and F13^T, respectively, and 73.9%, 79.6%, 72.0%, and 71.6%, respectively, for the species *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T, *Halomicroarcula limicola* JCM 18640^T and *Halomicroarcula amylolytica* LR21^T. In fact, *Halomicroarcula salina* JCM 18369^T shows the higher percentage of AAI (79.6%) with *Haloarcula vallismortis* ATCC 29715^T, and lower values with the remaining members of the genus *Halomicroarcula*: 71.3%, 71.5%, 73.5%, 74.3%, 74.5%, 74.2%, and 72.8% with strains F24A^T, F28, F27^T, and F13^T and the species *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula limicola* JCM 18640^T and *Halomicroarcula amylolytica* LR21^T, respectively. Although

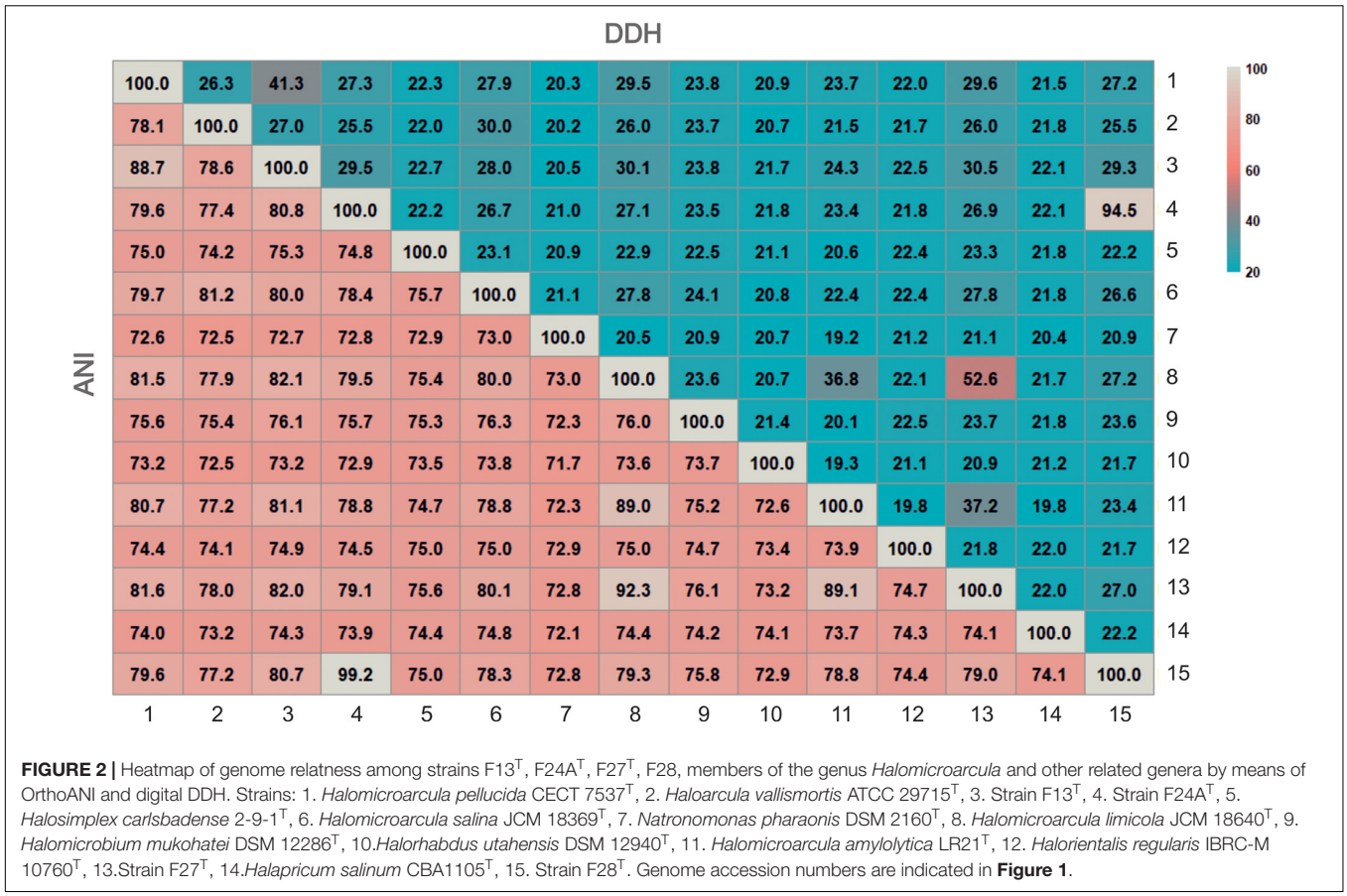


FIGURE 2 | Heatmap of genome relatness among strains F13^T, F24A^T, F27^T, F28, members of the genus *Halomicroarcula* and other related genera by means of OrthoANI and digital DDH. Strains: 1. *Halomicroarcula pellucida* CECT 7537^T, 2. *Haloarcula vallismortis* ATCC 29715^T, 3. Strain F13^T, 4. Strain F24A^T, 5. *Halosimplex carlsbadense* 2-9-1^T, 6. *Halomicroarcula salina* JCM 18369^T, 7. *Natronomonas pharaonis* DSM 2160^T, 8. *Halomicroarcula limicola* JCM 18640^T, 9. *Halomicrobium mukohatei* DSM 12286^T, 10.*Halorhabdus utahensis* DSM 12940^T, 11. *Halomicroarcula amylolytica* LR21^T, 12. *Halorientalis regularis* IBRC-M 10760^T, 13.Strain F27^T, 14.*Halapricum salinum* CBA1105^T, 15. Strain F28^T. Genome accession numbers are indicated in **Figure 1**.

the cutoff value of 65% for AAI delineation at the genus level is not universally accepted, considering these percentages of similarity, the species *Halomicroarcula salina* JCM 18369^T should be considered as a member of the genus *Haloarcula*. However, it must be noted that the AAI values for the remaining species of *Halomicroarcula* with respect to *Haloarcula vallismortis* ATCC 29715^T are also higher than 65% (71.6% for *Halomicroarcula amylolytica* LR21^T, 72.0% for *Halomicroarcula limicola* JCM 18640^T and 73.9% for the type species of the genus, *Halomicroarcula pellucida* CECT 7537^T). According to these results the species of the genera *Halomicroarcula* and *Haloarcula* should be merged into a single genus, as *Haloarcula*. A more exhaustive study, including a larger set of strains and based on additional genomic analysis reflecting their evolutionary relationships as well as their phenotypic features should be necessary in order to dilucidate the taxonomic status of the species of *Halomicroarcula* and *Haloarcula*.

We carried out a detailed phenotypic characterization of the four new isolates with respect to the type strains of species of *Halomicroarcula*, following the recommended minimal standards for describing new taxa of the class *Halobacteria* (Oren et al., 1997). These features included morphological, physiological, biochemical and nutritional characteristics, as well as the determination of the membrane polar lipids, which has been proved to be an important feature for the characterization of haloarchaeal genera (Oren et al., 1997, 2009). The results of the phenotypic features of the new isolates are shown in **Supplementary Table 2** and in the new species descriptions included at the conclusion section. These data confirm that there are differential features supporting the proposal of the three new species of the genus *Halomicroarcula* (**Supplementary Table 2**). Concerning the polar lipids composition, they were analyzed by High-Performance Thin Layer Chromatography (HPTLC) and the results are shown in **Supplementary Figure 2**. The four new isolates, strains F24A^T, F28, F27^T, and F13^T showed the typical polar lipid profile of species of the genus *Halomicroarcula*, composed of phosphatidylglycerol (PG), phosphatidylglycerol phosphate methyl ester (PGP-Me), phosphatidylglycerol sulfate (PGS) and sulfated diglycosil diether (S-DGD-1) (Echigo et al., 2013; Zhang and Cui, 2014). Overall, all these phenotypic features are in agreement with the previous genomic results and support the placement of the new isolates into three new species of the genus *Halomicroarcula*.

Genomic Features

In order to gain insights into the genome diversity of the genus *Halomicroarcula* we sequenced and analyzed the genome of the four strains isolated in this study and of the four already described species of the genus *Halomicroarcula*. The size of the genomes among the representatives of the genus *Halomicroarcula* ranged from 3.8 to 5.2 Mb, the DNA G + C content from 62.0 to 66.7 mol% and the total number of genes from 3827 to 5281.

TABLE 2 | CRISPR loci and mobile genetic elements determined in the genomes of strains F13^T, F24A^T, F27^T, F28 and the type strains of species of *Halomicroarcula*.

Strain	CRISPR loci	Prophage	Integrase	Transposase
Strain F13^T	4	1	16	28
Strain F24A^T	4	1	10	17
Strain F27^T	8	0	9	29
Strain F28	6	2	12	12
Halomicroarcula limicola JCM 18640^T	3	0	8	4
Halomicroarcula pellucida CECT 7537^T	3	5	11	7
Halomicroarcula salina JCM 18369^T	2	1	8	3
Halomicroarcula amylolytica LR21^T	4	1	4	11

Most genomes of these strains ranged from 3.8 to 4.0 Mb, with the exception of three genomes that showed higher values (**Table 1**), even considering that the quality of these genomes are within the standards already described for prokaryotes (Chun et al., 2018).

On the other hand, several genetic elements detailed in **Table 2** were also detected in those genomes. Members of the genus *Halomicroarcula* presented a large number of integrases and transposases, especially abundant in the isolated strains in comparison with the type strains of the previously described species of *Halomicroarcula*. Moreover, several CRISPR loci were also found, their number ranged from 2 in *Halomicroarcula salina* JCM 18369^T to 8 in strain F27^T (**Table 2**). While strains F24A^T, F13^T, F27^T and F28 presented *cas* cluster type IB and 94, 86, 84 and 74 spacers, respectively, not detectable *cas* genes were found in *Halomicroarcula limicola* JCM 18640^T, *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula amylolytica* LR21^T genomes, suggesting the lack of functionality of these systems in those strains. These data correlate with the high number of prophage sequences detected in *Halomicroarcula* genomes (**Table 2**), with a size between 5.4 and 33.1 kb. However, most of these prophage sequences were not complete.

The presence of these elements in high copy numbers in the genomes of members of the genus *Halomicroarcula* reflect a high genomic plasticity of these strains (including genetic rearrangements or horizontal gene transfer events), particularly in those isolated in this study from hypersaline soils. This fact could suggest a great adaptation of these taxa to different ecological niches and their success in nature.

Metabolism

Based on the genome annotation of members of the genus *Halomicroarcula* the major metabolic pathways of those strains could be reconstructed. For the carbohydrates uptake a large number of transporters were enconded in *Halomicroarcula* genomes reflecting their heterotrophic capabilities. Complete pathways involved in central carbohydrate metabolism (tricarboxylic acid cycle, oxidative pentose phosphate, Entner-Doudoroff or gluconeogenesis pathways) were present. However, in accordance with previous metabolic studies in haloarchaea (Falb et al., 2008; Anderson et al., 2011; Durán-Viseras et al., 2019a, 2021) the glycolysis pathway (Embden-Meyerhoff-Parnas) was truncated, therefore suggesting that alternative pathways like Entner-Doudoroff or the oxidative pentose phosphate could be used instead (Verhees et al., 2003). For the oxidation of the generated pyruvate to acetyl-CoA, both aerobic and anaerobic routes via pyruvate dehydrogenase and pyruvate ferredoxin oxidoreductase, respectively, were identified. Other carbohydrate metabolic pathways like the methylaspartate cycle, an anaplerotic acetate assimilation pathway, was also dilucidated in all *Halomicroarcula* genomes. This cycle has been previously identified in many other haloarchaea, such as the phylogenetically closest neighbor of the genus *Halomicroarcula*, the genus *Haloarcula* (Borjian et al., 2016). Its presence in haloarchaea has been frequently associated with the possession of genes for polyhydroxyalkanoate biosynthesis (Han et al., 2010; Borjian et al., 2016), which were also identified in *Halomicroarcula* genomes. Therefore, the presence of both pathways in *Halomicroarcula* genomes suggest the capability of members of this genus to biosynthetize polyhydroxialcanoates, as well as their adaptation for the assimilation of acetyl-CoA, produced from the internal carbon storage, during carbon starvation periods. This is a crucial factor for haloarchaea living under frequent starvation periods, such as on hypersaline soils from which our four strains were isolated. The genome of some members of the genus *Halomicroarcula* also encode for additional pathways related to carbohydrate metabolism (D-glucuronate and D-galacturonate degradation or glycogen biosynthesis), and with the hydrolysis of complex polysaccharides (α-amylase, chitinase, and endoglucanase enzymes), thus reflecting the metabolic diversity of representatives of this genus.

On the other hand, *Halomicroarcula* genomes encode the whole set of genes responsible for ammonia assimilation as a part of their nitrogen metabolism, such as enzymes for nitrate and nitrite reduction, the high-affinity ammonium transporter and glutamine synthetase and glutamate synthase. Moreover, complete biosynthesis pathways of several amino acids (i.e., arginine, cysteine, histidine, isoleucine, lysine, proline, serine, threonine, tryptophan, and valine) and few polyamine biosynthesis were identified along the genomes. Other determined sources of potential nitrogen rich compounds are different kinds of transporters for amino acids, urea and spermidine/putrescine uptake or a complete urease gene cluster.

Furthermore, *Halomicroarcula* genomes also encode genes for the ABC transporters for phosphate (pstSCAB) uptake, and in the case of *H. salina* JCM 18369^T and *H. pellucida* CECT 7537^T for phosphonate (phnCDE), as well as several mfs transporters related with multridrug efflux pump systems. Genes enconding

archaella were also found in all *Halomicroarcula* genomes, confirming their motility as phenotypic feature.

In addition, rhodopsin-like sequences were also identified in members of the genus *Halomicroarcula*, suggesting the phototrophic capabilities of this group of prokaryotic microorganisms. Strains F13^T, F24A^T, F28 and *Halomicroarcula salina* JCM 18369^T presented sensory rhodopsins, haloarchaeal proton pumps and halorhodopsins, while *Halomicroarcula pellucida* CECT 7537^T exhibited sensory rhodopsin and haloarchaeal proton pump, and strain F27^T and *Halomicroarcula amylolytica* LR21^T only sensory rhodopsins. No rhodopsin-like sequences were identified in *Halomicroarcula limicola* JCM 18640^T. The presence of rhodopsin like-sequences in most *Halomicroarcula* genomes suggest a versatile metabolic flexibility in illuminated conditions for members of this genus. These results are in accordance with recent comparative genomic studies in other haloarchaeal genera (Durán-Viseras et al., 2019a, 2020, 2021) in which rhodopsins were also identified, and also with previous metagenomic analyses on hypersaline systems (Fernández et al., 2014; Vera-Gargallo and Ventosa, 2018) that showed the existence of a large number of rhodopsin coding genes, clearly indicating the wide use of light by haloarchaea in these extreme habitats.

Regarding vitamins and cofactors metabolism the complete pathway for the vitamin B_{12} biosynthesis was determined in all members of the genus *Halomicroarcula*, suggesting the capability of this genus for its *de novo* synthesis. Remarkably, this pathway was recently identified by genomic studies in other haloarchaeal members of the genus *Halonotius* (Durán-Viseras et al., 2019a). An ABC transporter for biotin uptake was also encountered in *Halomicroarcula* genomes.

Besides, as a part of the metabolic analyses we also searched for metal resistance genes in *Halomicroarcula* genomes. While *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula amylolytica* LR21^T encode the complete arsenite resistance gene cluster (*ars*), this cluster was truncated in the other *Halomicroarcula* members. However, strains isolated in this study (F13^T, F24A^T, F27^T, and F28) and *Halomicroarcula amylolytica* LR21^T encode the CzcD transporter, a member of the cation diffusion facilitator (CDF) protein family, which was absent in the other *Halomicroarcula* reference strains. This transporter not only reflects a heavy metal resistance against cobalt, zinc or cadmium, but has also been suggested as a biomarker of nickel and vanadium pollution in some *Bacteria* (Anton et al., 1999; Fierros-Romero et al., 2020). On the contrary, none of the genes involved in copper or mercury resistance were identified in any of the studied genomes. These results could be related to the different stress conditions affecting the diverse ecological niches that they inhabit. In addition, several ABC metal transporters for zinc, cobalt and nickel were identified, which display a less critical role in maintaining metal homeostasis than CDF family transporters (Kaur et al., 2006).

Osmoadaptative Capabilities

To cope with the high salt concentrations and salinity fluctuations of hypersaline environments, halophilic microorganisms have developed diverse mechanisms of adaptation (Gunde-Cimerman et al., 2018). In order to delucidate the osmorregulatory strategy employed by *Halomicroarcula* members and their closest phylogenetic neighbors (the genera *Haloarcula* and *Halomicrobium*), their proteome was analyzed and compared with those of *salt-in* microorganisms [i.e., *Haloquadratum walsbyi* C23^T (Dyall-Smith et al., 2011) and *Salinibacter ruber* DSM 13855^T (Mongodin et al., 2005)] and with that of a *salt-out* bacterium [*Spiribacter salinus* M19-40^T (López-Pérez et al., 2013)] (**Supplementary Figure 3**). In accordance to *Haloquadratum walsbyi* C23^T and *Salinibacter ruber* DSM 13855^T and by contrast with *Spiribacter salinus* M19-40^T, *Halomicroarcula* representatives exhibited an acidic proteome with a low isoelectric point peak around 4.0 (**Supplementary Figure 3A**). Hence, highlighting a typical *salt-in* strategy for all members of this genus. Similar results were obtained for *Haloarcula* and *Halomicrobium* representatives (**Supplementary Figure 3B**). Besides, during the indeep genomic analysis of members of the genus *Halomicroarcula* we investigated the presence of genes putatively involved in osmorregulation. In conformity with results mentioned above, several transporters for Na^+ extrusion, K^+ uptake and Cl^- homeostasis were also identified, reinforcing its *salt-in* strategy.

According to previous suggestions (Youssef et al., 2014) and considering the frequent periods of low or fluctuenting salinity levels of the saline soils from which strains F13^T, F24A^T, F27^T, and F28 were isolated, it is not rare that they could not exclusively employ a *salt-in* strategy. Noteworthy, genes encoding *de novo* synthesis of trehalose via trehalose-6-phosphate synthase (OtsA) and trehalose-6-phosphatase (OtsB) were encountered in the genomes of strains F24A^T, F27^T, F28, *Halomicroarcula amylolytica* LR21^T and *Halomicroarcula limicola* JCM 18640^T suggesting their capability for trehalose biosynthesis. Trehalose is a dissacharide which has been reported to be used as an osmolyte by different organisms (Shivanand and Mugeraya, 2011). On the other side, the *otsAB* pathway was not found complete for strains F13^T, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula pellucida* CECT 7537^T, which lack the OtsB enzyme. However, the genome of strain F13^T was the only one exhibiting the enzyme trehalose synthase (TreT) suggesting this is the pathway used by this strain to produce trehalose. Although the synthesis of the compatible solute trehalose has been previously reported for other members of the haloarchaea (Youssef et al., 2014), no evidence has been demonstrated for any *Haloarcula* representatives (Youssef et al., 2014), the phylogenetically closest neighbors of the genus *Halomicroarcula*. Additionally, the presence of genes encoding the different trehalose biosynthesis pathways were analyzed during the course of this study in all members of the genera *Haloarcula* and *Halomicrobium* with available genomes (**Supplementary Table 3**). The enzyme trehalose synthase (TreT) was not found in any *Haloarcula* or *Halomicrobium* genomes, whereas the OtsAB pathway was only found complete in *Halomicrobium zhouii* CGMCC 1.10457^T (only OtsA was present in *Haloarcula sebkhae* JCM 19018^T, *H. salaria* ZP1-2, *H. quadrata* DSM 11927^T, *H. marismortui* ATCC 43049^T, *H. hispanica* ATCC 33960^T and *H. argentinensis* DSM 12282^T genomes; while *Halomicrobium katesii* DSM 19301^T only possessed OtsB). These results suggest

TABLE 3 | Number of genetic clusters involved in the biosynthesis of secondary metabolites determined in the genomes of strains F13^T, F24A^T, F28, F27^T, and the type strains of species of *Halomicroarcula*.

Cluster type	F13^T	F24A^T	F28	F27^T	*Halomicroarcula limicola* JCM 18640^T	*Halomicroarcula pellucida* CECT 7537^T	*Halomicroarcula salina* JCM 18369^T	*Halomicroarcula amylolytica* LR21^T
Terpene	2	2	2	2	2	2	2	2
Siderophore	1	1	1	1	1	2	1	1
Lanthipeptide	0	0	0	0	0	0	1	0
Thiopeptide	0	1	0	0	0	1	0	0

that the ability to synthethize trehalose is not extended neither in the genera *Haloarcula* and *Halomicrobium*, in accordance with previous presumptions of Youssef et al. (2014) for the genus *Haloarcula*. The ability to synthethize trehalose has been previously described for members of the class *Halobacteria* dwelling environments with lower salinity or salinity fluctuations (Youssef et al., 2014), such as the hypersaline soil from which strains F13^T, F24A^T, F27^T and F28 were isolated.

Surprinsingly, the complete pathway for the biosynthesis of the compatible solute glycine betaine from choline, was identified in the genomes of strain F13^T and *Halomicroarcula limicola* JCM 18640^T. No evidencies were identified in any of the genomes from species of the genera *Haloarcula* or *Halomicrobium*. Accordingly, the BCCT family transporters: OpuD and BetT, possibly for glycine betaine uptake or for other types of compatible solutes (Ziegler et al., 2010), were also present in the genomes of strains F13^T, F27^T, *Halomicroarcula pellucida* CECT 7537^T, *Halomicroarcula amylolytica* LR21^T and *Halomicroarcula salina* JCM 18369^T (in the case of OpuD) and strains F13^T, F27^T, *Halomicroarcula limicola* JCM 18640^T and *Halomicroarcula pellucida* CECT 7537^T (in the case of BetT). In the same way, OpuD was also present in *Haloarcula sebkhae* JCM 19018^T, *H. argentinensis* DSM 12282^T and *Halomicrobium zhouii* CGMCC 1.10457^T genomes, and BetT in the genomes of *Haloarcula sebkhae* JCM 19018^T and *H. argentinensis* DSM 12282^T. Moreover, *Halomicroarcula salina* JCM 18369^T and *Halomicroarcula amylolytica* LR21^T also showed the ABC transporter OpuA for betaine incorporation (Kempf and Bremer, 1995). Small conductance mechanosensitive channels from the MscS family were identified as well in *Halomicroarcula* genomes, and in the genomes of few members of the genera *Haloarcula* and *Halomicrobium*. These systems are ubiquitously used by microorganisms to manage the rapid transition from high salinity surroundings to environments with moderate salinities (Booth and Blount, 2012). The functionality of such safety valves has been demonstrated in some *Archaea* such as the marine thaumarchaeon *Nitrosopumilus maritimus* (Widderich et al., 2016). While BCCT family transporters are quite common in members of the class *Halobacteria* (Anderson et al., 2011; Youssef et al., 2014; Gunde-Cimerman et al., 2018), to the best of the authors knowledge, this is the first time that genes encoding glycine betaine synthesis are reported for any haloarchaea by far. This fact, could indicate an additional osmoadaptative mechanism for strains of the genus *Halomicroarcula*, which would corroborate the versatility of this group of microorganisms to adapt to environments with different salinities. The large amount of genes related with glycine betaine in the genomes of the genus *Halomicroarcula* raises the possibility that this organic solute could play an important role as osmoprotectant in this archaeal group that needs to be further investigated.

Biosynthesis of Secondary Metabolites

Secondary metabolites are a range of bioactive compounds of high interest for biotechnological, pharmaceutical or industrial applications (Charlesworth and Burns, 2015). The production of secondary metabolites give some environmental advantages to the microorganisms producing them (such as tolerance against environmental stress or interspecies defenses) (Osbourn, 2010; Piasecka et al., 2015; Wang and Lu, 2017; Wang et al., 2019). While secondary metabolites in the domains *Bacteria* and *Eukarya* have been deeply studied, analysis of these compounds in *Archaea* are more scarce (Charlesworth and Burns, 2015; Corral et al., 2020). As a part of our genomic analysis we also investigated the presence of genes involved on the biosynthesis of secondary metabolites in the genomes of members of the genus *Halomicroarcula*. The clusters that we have identified on these genomes are detailed in **Table 3** and **Figures 3–6**. Two different terpene clusters were detected in all *Halomicroarcula* genomes, which shared similarities within the different analyzed strains (**Figure 3**). To the best of the authors knowledge the presence of terpenes in *Archaea* has only been reported in a previous study (Wang et al., 2019), which also indicated the presence of two terpene clusters in the *Halobacteria* genomes. Terpenes are metabolites widely distributed in plants, fungi and bacteria in which they play a protective role; these compounds have also been suggested to be a source for the discovery of natural products (Pichersky et al., 2006; Yamada et al., 2015). On the other hand, siderophores, iron-chelating molecules produced during stress conditions or iron deficiency (Srivastava and Kowshik, 2013), were also distributed in the seven analyzed genomes of members of the genus *Halomicroarcula* (**Figure 4**). Despite the ubiquity of iron in the environment its solubility is very low, leading to the siderophore strategy to overcome this scarcity. The production of siderophores has also been detected before in some other haloarchaea (Dave et al., 2006). Finally, we detected a thiopeptide cluster in the genomes of strain F24A^T and *Halomicroarcula pellucida* CECT 7537^T (**Figure 5**), and a lanthipeptide in the genome of *Halomicroarcula salina* JCM 18369^T (**Figure 6**), both of them members of the ribosomally synthesized and post-translationally modified peptides (RiPP) family of natural products. No similarity was observed between the identified thiopeptide clusters (**Figure 5**).

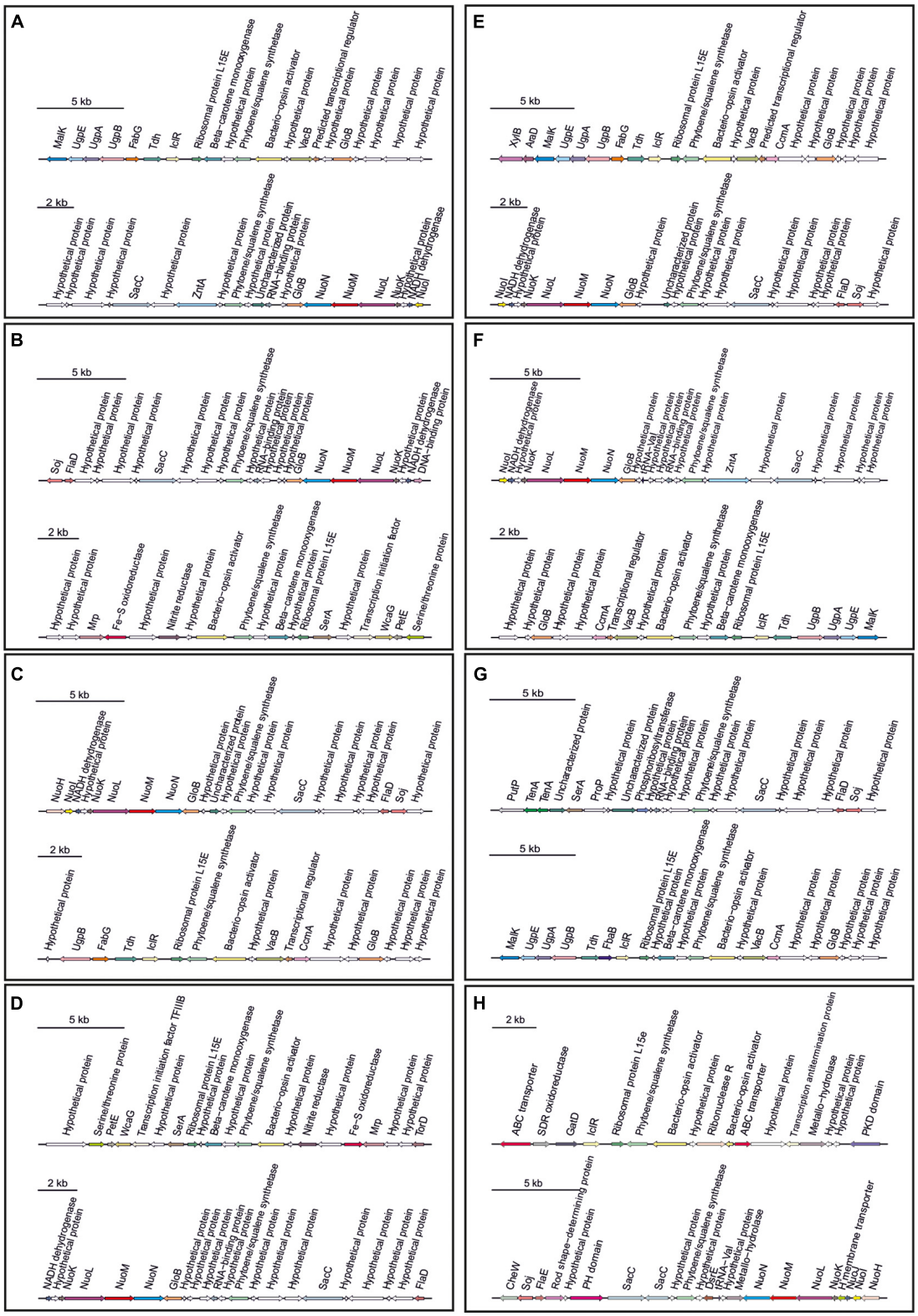


FIGURE 3 | Terpene biosynthetic gene clusters identified in *Halomicroarcula* genomes: **(A)** Strain F13^T, **(B)** Strain F24A^T, **(C)** Strain F27^T, **(D)** Strain F28, **(E)** *Halomicroarcula limicola* JCM 18640^T, **(F)** *Halomicroarcula pellucida* CECT 7537^T, **(G)** *Halomicroarcula salina* JCM 18369^T, and **(H)** *Halomicroarcula amylolytica* LR21^T.

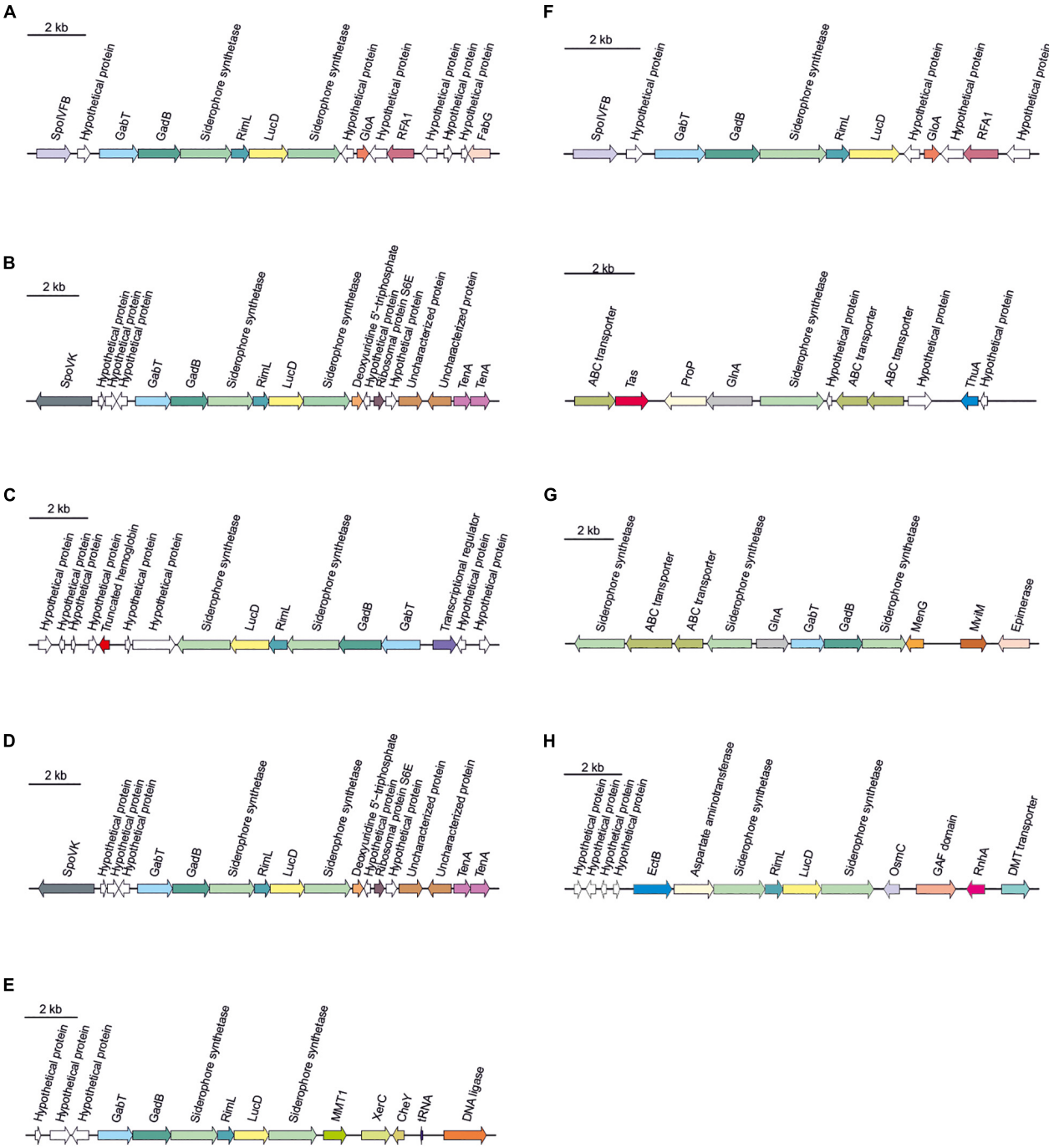


FIGURE 4 | Siderophore biosynthetic gene clusters identified in *Halomicroarcula* genomes: **(A)** Strain F13^T, **(B)** Strain F24A^T, **(C)** Strain F27^T, **(D)** Strain F28, **(E)** *Halomicroarcula limicola* JCM 18640^T, **(F)** *Halomicroarcula pellucida* CECT 7537^T, **(G)** *Halomicroarcula salina* JCM 18369^T, and **(H)** *Halomicroarcula amylolytica* LR21^T.

While thiopeptides constitute important components of the defense system in archaea, lanthipeptides display different biological activities, such as antimicrobial or antiallodynic (Repka et al., 2017; Wang and Lu, 2017). Thus, the biosynthesis of this wide variety of secondary metabolites by strains of the genus *Halomicroarcula* suggests the versatility of members of this genus, which have developed several ecological advantages such as a great adaptability to extreme conditions or defense mechanisms. In addition, our results brought to light their possible applications on the biotechnological field as a source of new natural compounds, also recently proposed for other novel haloarchaeal taxa (Durán-Viseras et al., 2021).



FIGURE 5 | Thiopeptide biosynthetic gene clusters identified in *Halomicroarcula* genomes: **(A)** Strain F24A^T and **(B)** *Halomicroarcula pellucida* CECT 7537^T.

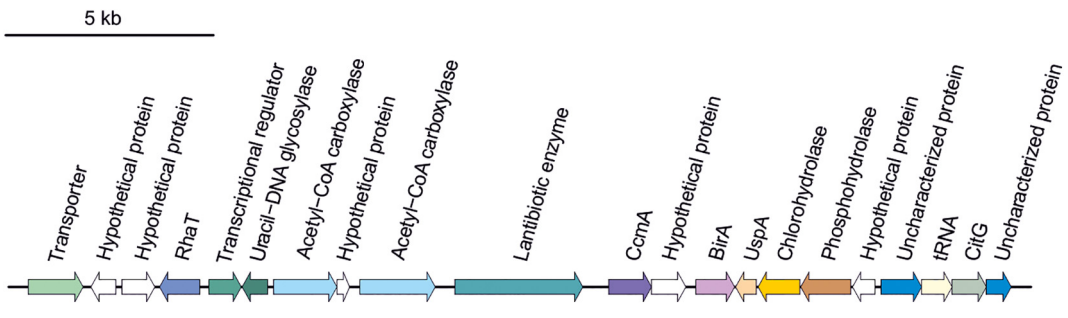


FIGURE 6 | Lanthipeptide biosynthetic gene cluster identified in *Halomicroarcula salina* JCM 18369^T.

CONCLUSION

The indeep comparative genomic analysis of the genus *Halomicroarcula* brought to light the presence of integrases, transposases and other gene transfer systems in high copy in the studied genomes, suggesting a vast plasticity for genes adquisition in members of the genus *Halomicroarcula*. This genomic plasticity leads to the versatile metabolism observed in the studied strains, such as the presence of metal resistance genes and genes coding for diverse carbohydrates pathways, some of them (i.e., methylaspartate cycle and polyhydroxyalkanoate biosynthesis) advantageous during carbon starvation periods. Besides, the capacity to biosynthethize diverse secondary metabolites (i.e., terpenes, siderophores, lanthipeptides or thiopeptides), could provide an ecological benefit for these microorganisms such as defense systems or adjustability to limiting conditions, and could also be source of novel compounds for biotechnological applications. The analysis of the proteome of members of the genus *Halomicroarcula* indicate they use a *salt-in* strategy. Nevertheless, complete pathways for the biosynthesis of compatible solutes (i.e., trehalose and glycine betaine), identified for the first time in haloarchaea during the detailed genomic analysis carried out in this study, suggests that alternative osmoadaptation strategies could be additionally employed. All these facts could give an ecological advantage for these microorganisms, which provide the genus *Halomicroarcula* the basis for the adaptation to a wide range of ecological niches and hence, playing a crucial role in the ecophysiological success of this taxa in the nature. Besides, this fact might justify the ability of these haloarchaea to grow at intermediate to low salinity environments.

On the other side, the exhaustive taxogenomic and phenotypic study carried out in this work, has permitted the characterization and description of three new species within the genus *Halomicroarcula*, for which we propose the new names *Halomicroarcula rubra* sp. nov., *Halomicroarcula nitratireducens*

sp. nov. and *Halomicroarcula salinisoli* sp. nov.; and whose descriptions are detailed below.

Halomicroarcula rubra sp. nov.

Halomicroarcula rubra (ru′bra. L. fem. adj. *rubra*, red).

Cells are Gram-stain-negative, motile rods with 1 × 1.2-2.5 μm. Does not grow anaerobically with L-arginine, dimethyl sulfoxide (DMSO) or potassium nitrate. Colonies are circular, entire, red pigmented with 0.2–0.3 mm in diameter on R2A25 medium after 14 days of incubation at 37°C. Extremely halophilic, able to grow in media with 10–30% (w/v) salts, with optimal growth at 25–30% (w/v) salts. No growth occurs in the absence of NaCl. Mg^{2+} is not required for growth. Able to grow in the pH range of 6.0–9.0 and from 25 to 50°C, with optimal growth at pH 7.5–8.0 and at 37°C. Chemoorganotrophic and aerobic. Catalase positive and oxidase negative. Gelatin is hydrolyzed but starch, Tween 80 and aesculin are not. Nitrate and nitrite are reduced, without gas production. H_2S is produced but indole and urease are not. Methyl red test is positive. Voges-Proskauer is negative. Acid is produced from D-arabinose, arbutin, L-citrulline, D-fructose, glycerol, D-glucose, D-ribose, L-xylitol and D-xylose but not from D-amygdalin, D-cellobiose, dulcitol, D-galactose, lactose, D-maltose, D-mannitol, D-mannose, D-melezitose, D-raffinose, D-sucrose, sorbitol or D-trehalose. The following compounds are used as carbon and energy source: D-melibiose, L-arginine, and L-methionine. The following compounds are not used as sole carbon and energy source: D-arabinose, D-cellobiose, fructose, D-galactose, D-glucose, lactose, maltose, D-mannose, L-raffinose, ribose, sucrose, D-trehalose, D-xylose, D-melezitose, salicin, butanol, dulcitol, ethanol, glycerol, D-mannitol, D-sorbitol, xylitol, methanol, benzoate, citrate, formate, fumarate, propionate, valerate, hippurate, malate, pyruvate, tartrate, L-alanine, L-cysteine, glutamine, L-glycine, L-lysine, isoleucine or valine. The major polar lipids are phosphatidylglycerol (PG), phosphatidylglycerol phosphate methyl ester (PGP-Me), phosphatidylglycerol sulfate (PGS) and sulfated mannosyl glucosyl diether (S-DGD-1). The DNA G + C content is 64.4 mol% (genome).

The type strain is $F13^T$ (= CCM 8888^T = CECT 9686^T = IBRC-M 11249^T = JCM 33313^T), isolated from a hypersaline soil located in Odiel saltmarshes, Huelva, Spain.

The GenBank/EMBL/DDBJ accession number for the 16S rRNA and *rpoB′* gene sequences of *Halomicroarcula rubra* $F13^T$ are MH447277 (*rrnA* gene), MH447279 (*rrnB* gene) and MH454085, respectively, and that of the complete genome is RKLR00000000.

Halomicroarcula nitratireducens sp. nov.

Halomicroarcula nitratireducens (ni.tra.ti.re.du'cens. N.L. masc. n. *nitras* (gen. *nitratis*), nitrate; L. pres. part. *reducens*, converting to a different state, reducing; N.L. part. adj. *nitratireducens*, reducing nitrate).

Cells are Gram-stain-negative, motile rods with 1 × 1.2–2.5 μm. Does not grow anaerobically with L-arginine, DMSO or potassium nitrate. Colonies are circular, entire, red to orange pigmented with 0.2–0.3 mm in diameter on R2A25 medium after 14 days of incubation at 37°C. Extremely halophilic, able to grow in media with 10–30% (w/v) salts, with optimal growth at 25–30% (w/v) salts. No growth occurs in the absence of NaCl. Mg^{2+} is not required for growth. Able to grow in the pH range of 6.0–9.0 and from 25 to 50°C, with optimal growth at pH 7.5 and at 37°C. Chemoorganotrophic and aerobic. Catalase and oxidase negative. Starch and aesculin are hydrolyzed but gelatin and Tween 80 are not. Nitrate and nitrite are reduced, without gas production. H_2S, indole and urease are not produced. Methyl red test is positive. Voges-Proskauer is negative. Acid is produced from D-arabinose, arbutin, L-citrulline, D-fructose, D-glucose, D-ribose and D-xylose but not from D-amygdalin, D-cellobiose, dulcitol, D-galactose, glycerol, lactose, D-maltose, D-mannitol, D-mannose, D-melezitose, D-raffinose, D-sucrose, sorbitol, D-trehalose or L-xylitol. The following compounds are used as carbon and energy source: fructose, D-galactose, D-glucose, maltose, ribose, sucrose, salicin, glycerol, D-mannitol, D-sorbitol, methanol, fumarate, or L-lysine. The following compounds are not used as sole carbon and energy source: D-arabinose, D-cellobiose, lactose, D-xylose, butanol, dulcitol, ethanol, xylitol, benzoate, citrate, formate, propionate, valerate, hippurate, malate, pyruvate, tartrate, L-alanine, L-arginine, L-cysteine, glutamine, L-methionine, L-glycine, isoleucine or valine. The major polar lipids are phosphatidylglycerol (PG), phosphatidylglycerol phosphate methyl ester (PGP-Me), phosphatidylglycerol sulfate (PGS) and sulfated mannosyl glucosyl diether (S-DGD-1). The DNA G + C content is 63.2 mol% (genome).

The type strain is $F27^T$ (= CCM 8887^T = CECT 9636^T = IBRC-M 11233^T = JCM 33314^T), isolated from a hypersaline soil located in Odiel saltmarshes, Huelva, Spain.

The GenBank/EMBL/DDBJ accession number for the 16S rRNA and *rpoB′* gene sequences of *Halomicroarcula nitratireducens* $F27^T$ are MH447286 (*rrnA* gene), MH447284 (*rrnB* gene) and MH454093, respectively, and that of the complete genome is RKLT00000000.

Halomicroarcula salinisoli sp. nov.

Halomicroarcula salinisoli (sa.li.ni.so'li. N.L. masc. adj. *salinus*, salty; L. neut. n. *solum*, soil; N.L. gen. n. *salinisoli*, of salty soil).

Cells are Gram-stain-negative, motile, pleomorphic rods with 1 × 1.2–2.5 μm. Does not grow anaerobically with L-arginine, DMSO or potassium nitrate. Colonies are circular, entire, pink pigmented with 0.2–0.3 mm in diameter on R2A25 medium after 14 days of incubation at 37°C. Extremely halophilic, able to grow in media with 15–30% (w/v) salts, with optimal growth at 25% (w/v) salts. No growth occurs in the absence of NaCl. Mg^{2+} is not required for growth. Able to grow in the pH range of 6.0–8.5 and from 25 to 50°C, with optimal growth at pH 7–7.5 and at 37°C. Chemoorganotrophic and aerobic. Catalase positive and oxidase negative. Gelatin, aesculin and Tween 80 are hydrolyzed but starch is not. Nitrate and nitrite are reduced, without gas production. H_2S production is variable, indole and urease are not produced. Methyl red test is positive. Voges-Proskauer is negative. Acid is produced from D-arabinose, arbutin, D-cellobiose, L-citrulline, D-fructose, D-glucose,

D-ribose and D-xylose but not from dulcitol, D-galactose, lactose, D-maltose, D-mannitol, D-mannose, D-melezitose, D-raffinose, D-sucrose, sorbitol, D-trehalose or L-xylitol. The following compounds are used as carbon and energy source: D-cellobiose, D-glucose, maltose, sucrose, D-sorbitol, citrate, fumarate or tartrate. The following compounds are not used as sole carbon and energy source: D-arabinose, D-galactose, lactose, ribose, D-xylose, salicin, glycerol, xylitol, benzoate, propionate, valerate, hippurate, pyruvate, L-arginine, L-cysteine, L-methionine, isoleucine or valine. The major polar lipids are phosphatidylglycerol (PG), phosphatidylglycerol phosphate methyl ester (PGP-Me), phosphatidylglycerol sulfate (PGS) and sulfated mannosyl glucosyl diether (S-DGD-1). The DNA G + C content is 63.9–64.1 mol% (genome).

The type strain is F24A^T (= CCM 8955^T = CECT 9687^T), isolated from a hypersaline soil located in Odiel saltmarshes, Huelva, Spain. The DNA G + C content of the type strain is 64.1 mol% (genome).

The GenBank/EMBL/DDBJ accession number for the 16S rRNA and *rpoB′* gene sequences of *Halomicroarcula salinisoli* F24A^T are MH447282 (*rrnA* gene), MH447281 (*rrnB* gene) and MH454092, respectively, and that of the complete genome is RKLQ00000000.

An additional strain of this species is strain F28. The DNA G + C content of this strain is 63.9 mol% (genome). The GenBank/EMBL/DDBJ accession number for the 16S rRNA and *rpoB′* gene sequences of this strain are MH450228 (*rrnA* gene), MH447330 (*rrnB* gene) and MH454094, respectively, and that of the complete genome is RKLS00000000.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, the 16S rRNA and *rpoB′* genes and the genome sequences generated for this study can be found in the GenBank/EMBL/DDBJ database under the accession numbers MH447277, MH447279, MH447282, MH447281, MH450228, MH447330, MH447286, MH447284, MH454085, MH454092, MH454094, MH454093, RKLR00000000, RKLQ00000000, RKLS00000000, RKLT00000000, JAHQXF000000000, RKLW00000000, and JAHQXE000000000.

AUTHOR CONTRIBUTIONS

AD-V, CS-P, and AV did the conceptualization. AD-V did the strains isolation, the taxogenomic characterization, and the comparative genomic analyses. AD-V and AV wrote the manuscript. AD-V prepared the tables and figures. AV and CS-P did the funding acquisition. All authors read and approved the final version of the manuscript.

FUNDING

This research was funded by Junta de Andalucía, Spain (grants US-1263771 [US/JUNTA/FEDER/UE], P20_01066 and BIO-213, which included FEDER funds), and FEDER/Spanish Ministry of Science and Innovation-State Research Agency (projects CGL2017-83385-P and PID2020-118136GB-I00).

ACKNOWLEDGMENTS

We thank A. Oren for his help on the nomenclature of the three new species.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.751746/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

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

Anderson, I., Scheuner, C., Göker, M., Mavromatis, K., Hooper, S. D., Porat, I., et al. (2011). Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes. *PLoS One* 6:e20237. doi: 10.1371/journal.pone.0020237

Anton, A., Große, C., Reißmann, J., Pribyl, T., and Nies, D. H. (1999). CzcD is a heavy metal ion transporter involved in regulation of heavy metal resistance in *Ralstonia* sp. strain CH34. *J. Bacteriol.* 181, 6876–6881. doi: 10.1128/jb.181.22.6876-6881.1999

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.

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

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

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

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: 10.1093/nar/gkz310

Booth, I. R., and Blount, P. (2012). The MscS and MscL families of mechanosensitive channels act as microbial emergency release valves. *J. Bacteriol.* 194, 4802–4809. doi: 10.1128/JB.00576-12

Borjian, F., Han, J., Hou, J., Xiang, H., and Berg, I. A. (2016). The methylaspartate cycle in haloarchaea and its possible role in carbon metabolism. *ISME J.* 10, 546–557. doi: 10.1038/ismej.2015.132

Charlesworth, J. C., and Burns, B. P. (2015). Untapped resources: biotechnological potential of peptides and secondary metabolites in *Archaea*. *Archaea* 2015:282035. doi: 10.1155/2015/282035

Chen, F., Xu, Y., Sun, S., Shi, X., Liu, A., and Chen, S. (2020). *Halomicroarcula amylolytica* sp. nov., a novel halophilic archaeon isolated from a salt mine. *Int. J. Syst. Evol. Microbiol.* 70, 4978–4985. doi: 10.1099/ijsem.0.004368

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

Corral, P., Amoozegar, M. A., and Ventosa, A. (2020). Halophiles and their biomolecules: recent advances and future applications in biomedicine. *Mar. Drugs* 18:33. doi: 10.3390/md18010033

Corral, P., Corcelli, A., and Ventosa, A. (2015). Halostagnicola bangensis sp. nov., an alkaliphilic haloarchaeon from a soda lake. *Int. J. Syst. Evol. Microbiol.* 65, 754–759. doi: 10.1099/ijs.0.000006

Corral, P., de la Haba, R. R., Sánchez-Porro, C., Amoozegar, M. A., Papke, R. T., and Ventosa, A. (2016). *Halorubrum halodurans* sp. nov., an extremely halophilicarchaeon isolated from a hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 66, 435–444. doi: 10.1099/ijsem.0.000738

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

Dave, B. P., Anshuman, K., and Hajela, P. (2006). Siderophores of halophilic archaea and their chemical characterization. *Indian J. Exp. Biol.* 44, 340–344.

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5685–5689. doi: 10.1073/PNAS.89.12.5685

Durán-Viseras, A., Andrei, A. Ş, Ghai, R., Sánchez-Porro, C., and Ventosa, A. (2019a). New *Halonotius* species provide genomics-based insights into cobalamin synthesis in haloarchaea. *Front. Microbiol.* 10:1928. doi: 10.3389/fmicb.2019.01928

Durán-Viseras, A., Sánchez-Porro, C., and Ventosa, A. (2019b). *Halorientalis pallida* sp. nov., an extremely halophilic archaeon isolated from a marine saltern. *Int. J. Syst. Evol. Microbiol.* 69, 3636–3643. doi: 10.1099/ijsem.0.003675

Durán-Viseras, A., Andrei, A. Ş, Vera-Gargallo, B., Ghai, R., Sánchez-Porro, C., and Ventosa, A. (2021). Culturomics-based genomics sheds light on the ecology of the new haloarchaeal genus *Halosegnis*. *Environ. Microbiol.* 23, 3418–3434. doi: 10.1111/1462-2920.15082

Durán-Viseras, A., Sánchez-Porro, C., and Ventosa, A. (2020). *Natronomonas salsuginis* sp. nov., a new inhabitant of a marine solar saltern. *Microorganisms* 8:605. doi: 10.3390/microorganisms8040605

Dyall-Smith, M. L., Pfeiffer, F., Klee, K., Palm, P., Gross, K., Schuster, S. C., et al. (2011). *Haloquadratum walsbyi*: limited diversity in a global pond. *PLoS One* 6:e20968. doi: 10.1371/journal.pone.0020968

Echigo, A. (2016). "Halomicroarcula," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. D. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. de Vos, et al. (Hoboken: John Wiley & Sons, Inc.), 1–9.

Echigo, A., Minegishi, H., Shimane, Y., Kamekura, M., Itoh, T., and Usami, R. (2013). *Halomicroarcula pellucida* gen. nov., sp. nov., a non-pigmented, transparent-colony-forming, halophilic archaeon isolated from solar salt. *Int. J. Syst. Evol. Microbiol.* 63, 3556–3562. doi: 10.1099/ijs.0.049965-0

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

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

Falb, M., Müller, K., Königsmaier, L., Oberwinkler, T., Horn, P., von Gronau, S., et al. (2008). Metabolism of halophilic archaea. *Extremophiles* 12, 177–196. doi: 10.1007/s00792-008-0138-x

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). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N. Y)* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Fernández, A. B., Vera-Gargallo, B., Sánchez-Porro, C., Ghai, R., Papke, R. T., Rodríguez-Valera, F., et al. (2014). 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

Fierros-Romero, G., Gómez-Ramírez, M., Sharma, A., Pless, R. C., and Rojas-Avelizapa, N. G. (2020). *czcD* gene from *Bacillus megaterium* and *Microbacterium liquefaciens* as a potential nickel–vanadium soil pollution biomarker. *J. Basic Microbiol.* 60, 22–26. doi: 10.1002/jobm.201900323

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

Galinski, E. A., and Trüper, H. G. (1994). Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiol. Rev.* 15, 95–108. doi: 10.1111/j.1574-6976.1994.tb00128.x

Gunde-Cimerman, N., Oren, A., and Plemenitaš, A. (2018). Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol. Rev.* 42, 353–375. doi: 10.1093/femsre/fuy009/4909803

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Han, J., Hou, J., Liu, H., Cai, S., Feng, B., Zhou, J., et al. (2010). Wide distribution among halophilic archaea of a novel polyhydroxyalkanoate synthase subtype with homology to bacterial type III synthases. *Appl. Environ. Microbiol.* 76, 7811–7819. doi: 10.1128/AEM.01117-10

Jukes, T. H., and Cantor, C. R. (1969). "Evolution of protein molecules," in *Mammalian Protein Metabolism*, ed. H. N. Munro (London: Elsevier), 21–132.

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

Kaur, A., Pan, M., Meislin, M., Facciotti, M. T., El-Gewely, R., and Baliga, N. S. (2006). A systems view of haloarchaeal strategies to withstand stress from transition metals. *Genome Res.* 16, 841–854. doi: 10.1101/gr.5189606

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

Kempf, B., and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch. Microbiol.* 170, 319–330. doi: 10.1007/s002030050649

Konstantinidis, K. T., Rosselló-Móra, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J.* 11, 2399–2406. doi: 10.1038/ismej.2017.113

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

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lee, I., Ouk Kim, Y., 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

López-Pérez, M., Ghai, R., León, M. J., Rodríguez-Olmos, A., Copa-Patiño, J. L., Soliveri, J., et al. (2013). Genomes of "*Spiribacter*", a streamlined, successful halophilic bacterium. *BMC Genomics* 14:787. doi: 10.1186/1471-2164-14-787

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J. Mol. Biol.* 3, 208–218. doi: 10.1016/S0022-2836(61)80047-8

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

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

Mongodin, E. F., Nelson, K. E., Daugherty, S., DeBoy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and *archaea*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152. doi: 10.1073/pnas.0509073102

Oren, A. (2011). "Diversity of halophiles," in *Extremophiles Handbook*, ed. K. Horikoshi (Tokyo: Springer), 309–325.

Oren, A. (2015). Halophilic microbial communities and their environments. *Curr. Opin. Biotechnol.* 33, 119–124. doi: 10.1016/j.copbio.2015.02.005

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. (2017a). "Halobacteria," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. D. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. de Vos, et al. (Hoboken: John Wiley & Sons, Inc.), 1–5.

Oren, A., and Ventosa, A. (2017b). "Haloarculaceae," in *Bergey's Manual of Systematics of Archaea and Bacteria*, eds W. D. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. de Vos, et al. (Hoboken: John Wiley & Sons, Inc.), 1–5.

Oren, A., Ventosa, A., and Grant, W. D. (1997). Proposed minimal standards for description of new taxa in the order *Halobacteriales*. *Int. J. Syst. Bacteriol.* 47, 233–238. doi: 10.1099/00207713-47-1-233

Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* 26, 449–457. doi: 10.1016/j.tig.2010.07.001

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

Piasecka, A., Jedrzejczak-Rey, N., and Bednarek, P. (2015). Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol.* 206, 948–964. doi: 10.1111/nph.13325

Pichersky, E., Noel, J. P., and Dudareva, N. (2006). Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* 311, 808–811. doi: 10.1126/science.1118510

Repka, L. M., Chekan, J. R., Nair, S. K., and Van Der Donk, W. A. (2017). Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev.* 117, 5457–5520. doi: 10.1021/acs.chemrev.6b00591

Rice, P., Longden, L., and Bleasby, A. (2000). EMBOSS: the european molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

Rodríguez-Valera, F. (1988). "Characteristics and microbial ecology of hypersaline environments," in *Halophilic Bacteria*, ed. F. Rodríguez-Valera (Boca Raton: CRC Press), 3–30.

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

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shivanand, P., and Mugeraya, G. (2011). Halophilic bacteria and their compatible solutes -osmoregulation and potential applications. *Curr. Sci.* 100, 1516–1521.

Srivastava, P., and Kowshik, M. (2013). Mechanisms of metal resistance and homeostasis in haloarchaea. *Archaea* 2013:732864. doi: 10.1155/2013/732864

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

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., Fernández, A. B., León, M. J., Sánchez-Porro, C., and Rodriguez-Valera, F. (2014). The Santa Pola saltern as a model for studying the microbiota of hypersaline environments. *Extremophiles* 18, 811–824. doi: 10.1007/s00792-014-0681-6

Ventosa, A., Mellado, E., Sánchez-Porro, C., and Márquez, M. C. (2008). "Halophilic and halotolerant micro-organisms from soils," in *Microbiology of Extreme Soils*, eds P. Dion and C. S. Nautiyal (Berlin: Springer), 87–115. doi: 10.1007/978-3-540-74231-9_5

Vera-Gargallo, B., Chowdhury, T. R., Brown, J., Fansler, S. J., Durán-Viseras, A., Sánchez-Porro, C., et al. (2019). Spatial distribution of prokaryotic communities in hypersaline soils. *Sci. Rep.* 9:1769. doi: 10.1038/s41598-018-38339-z

Vera-Gargallo, B., and Ventosa, A. (2018). Metagenomic insights into the phylogenetic and metabolic diversity of the prokaryotic community dwelling in hypersaline soils from the Odiel Saltmarshes (SW Spain). *Genes* 9:152. doi: 10.3390/genes9030152

Verhees, C. H., Kengen, S. W. M., Tuininga, J. E., Schut, G. J., Adams, M. W. W., de Vos, W. M., et al. (2003). The unique features of glycolytic pathways in *Archaea*. *Biochem. J.* 375, 231–246. doi: 10.1042/bj20021472

Wang, S., and Lu, Z. (2017). "Secondary metabolites in archaea and extreme environments," in *Biocommunication of Archaea*, ed. G. Witzany (Berlin: Springer), 235–239.

Wang, S., Zheng, Z., Zou, H., Li, N., and Wu, M. (2019). Characterization of the secondary metabolite biosynthetic gene clusters in archaea. *Comput. Biol. Chem.* 78, 165–169. doi: 10.1016/j.compbiolchem.2018.11.019

Widderich, N., Czech, L., Elling, F. J., Könneke, M., Stöveken, N., Pittelkow, M., et al. (2016). Strangers in the archaeal world: osmostress-responsive biosynthesis of ectoine and hydroxyectoine by the marine thaumarchaeon *Nitrosopumilus maritimus*. *Environ. Microbiol.* 18, 1227–1248. doi: 10.1111/1462-2920.13156

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., et al. (2015). Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 112, 857–862. doi: 10.1073/pnas.1422108112

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

Youssef, N. H., Savage-Ashlock, K. N., McCully, A. L., Luedtke, B., Shaw, E. I., Hoff, W. D., et al. (2014). Trehalose/2-sulfotrehalose biosynthesis and glycine-betaine uptake are widely spread mechanisms for osmoadaptation in the *Halobacteriales*. *ISME J.* 8, 636–649. doi: 10.1038/ismej.2013.165

Zhang, W., and Cui, H. (2014). *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

Zhang, W., and Cui, H. (2015). *Halomicroarcula salina* sp. nov., isolated from a marine solar saltern. *Int. J. Syst. Evol. Microbiol.* 65, 1628–1633. doi: 10.1099/ijs.0.000150

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485

Ziegler, C., Bremer, E., and Krämer, R. (2010). The BCCT family of carriers: from physiology to crystal structure. *Mol. Microbiol.* 78, 13–34. doi: 10.1111/j.1365-2958.2010.07332.x

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Durán-Viseras, Sánchez-Porro 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: 25 January 2022
doi: 10.3389/fmicb.2021.739234

Diversity and Distribution of Anaerobic Ammonium Oxidation Bacteria in Hot Springs of Conghua, China

Lan Liu[1], Ai-Ping Lv[1], Manik Prabhu Narsing Rao[1], Yu-Zhen Ming[1], Nimaichand Salam[1], Meng-Meng Li[1], Ze-Tao Liu[1], Xiao-Tong Zhang[1], Jing-Yi Zhang[1], Wen-Dong Xian[1], Jian-Yu Jiao[1] and Wen-Jun Li[1,2]**

[1] State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Life Sciences, Sun Yat-sen University, Guangzhou, China, [2] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Ürümqi, China

Anaerobic ammonium oxidation (anammox) is an important process of the nitrogen cycle, and the anammox bacteria have been studied in a wide variety of environments. However, the distribution, diversity, and abundance of anammox bacteria in hot springs remain enigmatic. In this study, the anammox process was firstly investigated in hot springs of Conghua, China. Anammox-like bacterial sequences that closely affiliated to "*Candidatus* Brocadia," "*Candidatus* Kuenenia," "*Candidatus* Scalindua," "*Candidatus* Anammoxoglobus," and "*Candidatus* Jettenia" were detected. Several operational taxonomic units (OTUs) from this study shared low sequence identities to the 16S rRNA gene of the known anammox bacteria, suggesting that they might be representing putative novel anammox bacteria. A quantitative PCR analysis of anammox-specific 16S rRNA gene confirmed that the abundance of anammox bacteria ranged from 1.60×10^4 to 1.20×10^7 copies L^{-1}. Nitrate was a key environmental factor defining the geographical distribution of the anammox bacterial community in the hot spring ecosystem. Dissolved inorganic carbon had a significant influence on anammox bacterial biodiversity. Our findings for the first time revealed that the diverse anammox bacteria, including putative novel anammox bacterial candidates, were present in Conghua hot spring, which extended the existence of anammox bacteria to the hot springs in China and expands our knowledge of the biogeography of anammox bacteria. This work filled up the research lacuna of anammox bacteria in Chinese hot spring habitat and would guide for enrichment strategies of anammox bacteria of Conghua hot springs.

Keywords: hot springs, anammox bacteria, diversity, physicochemical analysis, putative novel taxa

OPEN ACCESS

Edited by:
André Antunes, Macau University of Science and Technology, China

Reviewed by:
Yiguo Hong, Guangzhou University, China
Satya P. Singh, Saurashtra University, India
Isao Yumoto, National Institute of Advanced Industrial Science and Technology (AIST), Japan
Jinjun Kan, Stroud Water Research Center, United States

****Correspondence:***
Wen-Jun Li
liwenjun3@mail.sysu.edu.cn
Jian-Yu Jiao
jiaojy5@mail.sysu.edu.cn

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

Received: *10 July 2021*
Accepted: *07 December 2021*
Published: *25 January 2022*

Citation:
Liu L, Lv A-P, Narsing Rao MP, Ming Y-Z, Salam N, Li M-M, Liu Z-T, Zhang X-T, Zhang J-Y, Xian W-D, Jiao J-Y and Li W-J (2022) Diversity and Distribution of Anaerobic Ammonium Oxidation Bacteria in Hot Springs of Conghua, China. Front. Microbiol. 12:739234. doi: 10.3389/fmicb.2021.739234

INTRODUCTION

Anaerobic ammonium oxidization (anammox) is a process that can convert ammonium to dinitrogen gas (N_2) coupled with nitrite reduction under anoxic conditions (Thamdrup, 2012). Anammox is a microbe-mediated process that was predicted in 1977 (Broda, 1977) but was first described in 1995 by Mulder (Mulder et al., 1995) from the bioreactors of wastewater treatment plants. The discovery of anammox challenged the view that heterotrophic denitrification was the only known pathway for nitrogen loss to the atmosphere. The anammox process is mediated by bacteria affiliated to the same monophyletic branch named "*Candidatus* Brocadiales" within the phylum *Planctomyces*. At present, there are 23 *Candidatus* species from 6 different genera of anammox bacteria that have been described, including the *Brocadia anammoxidans*, *Brocadia fulgida*, *Brocadia sinica*, *Brocadia brasiliensis*, *Brocadia caroliniensis*, and *Brocadia sapporoensis* in the genus "*Candidatus* Brocadia" (Strous et al., 1999; Kartal et al., 2008; Hu et al., 2010; Araujo et al., 2011; Narita et al., 2017), *Kuenenia stuttgartiensis* in "*Candidatus* Kuenenia" (Schmid et al., 2000), *Jettenia asiatica*, *Jettenia ecosi*, *Jettenia moscovienalis*, and *Jettenia caeni* in "*Candidatus* Jettenia" (Quan et al., 2008; Ali et al., 2015; Nikolaev et al., 2015; Botchkova et al., 2018), *Scalindua sorokinii*, *Scalindua brodae*, *Scalindua wagneri*, *Scalindua arabica*, *Scalindua sinooilfield*, *Scalindua zhenghei*, *Scalindua profunda*, *Scalindua marina*, and *Scalindua richardsii* in "*Candidatus* Scalindua" (Kuypers et al., 2003; Schmid et al., 2003; Van De Vossenberg et al., 2008; Woebken et al., 2008; Li et al., 2010; Brandsma et al., 2011; Fuchsman et al., 2012; Oshiki et al., 2017; Speth et al., 2017), *Anammoxoglobus propionicus* in "*Candidatus* Anammoxoglobus" (Kartal et al., 2007), and *Brasilis concordiensis* in "*Candidatus* Brasilis" (Viancelli et al., 2011).

Anammox bacteria have a wide distribution in various natural habitats, such as marine (Rich et al., 2008; Dang et al., 2010; Brandsma et al., 2011; Hong et al., 2011), freshwater (Zhang et al., 2007; Yoshinaga et al., 2009; Moore et al., 2011) and terrestrial ecosystems (Humbert et al., 2010; Hu et al., 2011; Zhu et al., 2011), and the animal host (Hoffmann et al., 2009; Mohamed et al., 2010). They were also detected in some special environments, such as high-temperature environments. Using the 16S rRNA gene and *hzo* molecular biomarker, the group of Li (Li et al., 2010) indicated the occurrence of anammox bacteria in nine of seventeen oil reservoirs. The role of anammox in deep-sea hydrothermal vents was investigated by concurrent surveys including the amplification of 16S rRNA gene sequences, ladderanes lipids analysis, and isotope-pairing experiments (Byrne et al., 2009), which strongly indicated the presence and activity of new anammox bacteria in different hydrothermal areas. The existence of anammox bacteria was also detected in the sediments of Bulukey River in Turpan where the surface temperature can reach 75°C in summer (Zhu et al., 2015). As one typical high-temperature environment, there has been limited evidence for the existence and role of anammox bacteria in hot springs (Jaeschke et al., 2009). Therefore, whether anammox bacteria are prevalent in hot springs remains unknown, which impacts our understanding of anammox in the terrestrial nitrogen cycle.

Hot springs, remarkably similar to ancient environments, are the model ecosystem for study on the origin and evolution of life, biogeochemistry, and biogeography (Pace, 1997; Whitaker et al., 2003). The terrestrial hot springs are characterized by extreme geochemical conditions and diverse mineralogical compositions. These geothermal features usually harbor tremendous diversity of uncultivated or unexplored "microbial dark matter" (Jiao et al., 2021b), with novel metabolic capacities and special adaption strategies (Hua et al., 2019; Tan et al., 2019; Luo et al., 2020; Jiao et al., 2021a). Conghua, in the Guangdong province, is a county that encompasses many hot springs. These hot springs are characterized as sodium bicarbonate type containing fluoride, weak radioactive plutonium, and sodium, calcium, potassium, magnesium, silicon dioxide, and other beneficial elements; however, no attention was paid to the microorganisms in it. In this study, six representative hot spring samples with the temperature higher than 45°C were collected from Conghua to assess the diversity and structure of anammox bacteria by Illumina-based 16S rRNA gene amplicon sequencing, measure the abundance of anammox bacteria by quantitative PCR (qPCR) analysis of 16S rRNA gene, and analyze the relations between the physicochemical characteristics of hot spring and the abundance and community structure of anammox bacteria. This study provides a new understanding of the community composition of anammox bacteria in hot springs.

MATERIALS AND METHODS

Sample Collection

Six hot spring samples were collected from three stations: LiuXi River Left (LXRL, 23.649000°N, 113.652000°E), LiuXi River Right (LXRR, 23.649500°N 113.654000°E), and GuanWeiHui (GWH, 23.689363°N, 113.706496°E) of Conghua, Guangdong province in China (**Supplementary Figure 1**). Origins of five of these samples (GWH1, GWH2, GWH3, LXRR1, and LXRR2) are a subsurface with more than 60 m depth, which limit us to obtain only hot spring water *via* a pump. The source of sample LXRL1 is surface and covered by stones, with the hot spring water constantly gushing out, which made it easier for us to collect hot spring water *via* 10 L white plastic bottles (As One, China). These hot spring water samples were collected during July 2018. Following collection, 20 L of each sample was filtered through 0.22 μm membrane filters (Pall Gelman, Port Washington, NY, United States) with a 150 mm diameter. Biomass on the filter was then transferred to a sterilized 50 mL round-bottom centrifuge tube (Corning, Tewksbury, MA, United States) and stored with dry ice for further DNA extraction. The filtrates were collected in sterilized 250 mL white plastic bottles (As One, China) with 58 mm in diameter and 103 mm in height, transported to the laboratory, and stored at −20°C for subsequent analysis of physicochemical properties.

Environmental Parameter Analysis

Temperature and pH were measured *in situ* using a portable multimeter (HACH58258; Loveland, CO, United States). Nutrients in the water samples were analyzed according to standard procedures (Lu et al., 2015). The concentrations of total organic carbon, total carbon, and dissolved inorganic carbon were determined using a total organic carbon analyzer (TOC-VCPN; Shimadzu Corp., Tokyo, Japan) based on combustion–oxidation and combustion detection methods. The concentrations of soluble reactive phosphorus, nitrite, and nitrate were measured spectrophotometrically based on molybdenum blue and hydrazine sulfate-NEDD (N-(1-napthyl)ethylenediamine dihydrochloride) methods, respectively. Concentrations of total nitrogen and ammonium were performed using flow injection protocols (WESTCO, SmartChem 200 series flow injection analysis system, Scientific Instruments, Milan, Italy).

DNA Extraction, PCR Amplification, and Sequencing

The total community DNA from the samples was extracted using the FastDNA Spin Kit (MP Biomedicals, Santa Ana, CA, United States) according to the manufacturer's protocols. DNA quantity and quality were checked using a NanoDrop® ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States) and 1.0% agarose-gel electrophoresis. The 16S rRNA gene of anammox bacteria was amplified using a nested PCR approach with first-step primer pair PLA46F (5′-GGATTAGGCATGCAAGTC-3′)/1390R (5′-GACGGGCGGTGTGTACAA-3′) (Schmid et al., 2000) and second-step primer pair Amx368F (5′-TTCGCAATGCCCGAAAGG-3′)/Amx820R (5′-AAAACCCCTCTACTTAGTGCCC-3′) (Wang et al., 2017). PCR reaction conditions were as follows: 94°C for 4 min; 32 cycles of 94°C for 30 s, 56°C (first step) or 58°C (second step) for 30 s, followed by 72°C for 1 min; and a final 10 min extension at 72°C. High-throughput paired-end Illumina MiSeq (Illumina, San Diego, CA, United States) sequencing (2 × 300 bp) was performed at Genewiz, Suzhou, China.

Quantitative PCR Analysis

The abundance of the 16S rRNA gene of anammox bacteria was determined in triplicate on a Bio-Rad iQ5 thermal cycler (Bio-Rad Laboratories) using primer sets AMX-808F (5′-ARCYGTAAACGATGGGCACTAA-3′)/AMX1040R (5′-CAGCCATGCAACACCTGTRATA-3′) (Hamersley et al., 2007) with a thermal profile of 3 min at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at 55°C, and 30 s at 72°C. SYBR Green I-based real-time PCR assays were carried out in a volume of 20 μL, containing 10 μL of SYBR Premix Ex Taq (TAKARA, Guangzhou, China), 0.2 μL of each primer, and 1 μL DNA template, and was topped up with ddH_2O to a total volume of 20 μL. A standard plasmid carrying 16S rRNA gene fragment was generated by amplifying them from the extracted DNA sample and cloning into pEASY-T1 Vector (TransGen, Beijing, China). The concentration of plasmid DNA was determined on a NanoDrop® ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA), and the copy number of the target gene was calculated directly from the concentration of the extracted plasmid DNA. A significant linear relationship of 16S rRNA gene amplification (R^2 = 0.9966) was obtained between the log 10 values of the standard plasmid DNA concentration (2.00 × 10–2.00 × 10^7 copies L^{-1}) and the associated threshold cycles (Ct). Melting curves and gel electrophoresis were performed to confirm the specificity of the qPCR amplification. In all experiments, blank controls containing no template DNA were performed under the same qPCR procedure to ensure no contamination in our PCR system and the reliability of the results.

Data Analysis

Analysis of raw Illumina fastq files was performed using Quantitative Insights Into Microbial Ecology 2 (QIIME 2, ver. 2019.1) (Bolyen et al., 2019). DADA 2 (Callahan et al., 2016) was used to filter, trim, denoise, and merge sequences. Chimeric sequences were identified and removed. Taxonomy was assigned to chimeric-free sequences using a q2-feature classifier (Bokulich et al., 2018) classify-sklearn naïve Bayes taxonomy classifier against the Silva 132 97% OTU reference sequences (Quast et al., 2013). Phylogenetic analysis of the representative 16S rRNA gene sequences of anammox bacteria was conducted with the MEGA 7.0 software, and the phylogenetic tree was visualized and annotated using iTOL (Letunic and Bork, 2019). Optimal growth temperatures of top 30 dominant OTUs and the reference targets from the NCBI database were calculated by using the GC contents (Kimura et al., 2006). To analyze the alpha diversity, Shannon (Shannon–Weaver), Simpson (Gini–Simpson), and Pielou indices were calculated by using the R program (v.3.5.3)[1] with the vegan package. UpSetR (Conway et al., 2017) was used to investigate whether exclusively shared OTUs existed among the samples. The relationships of anammox bacterial diversity, anammox bacterial abundance, and physicochemical parameters of samples were calculated by Spearman's correlation analysis using R (corrplot package). The contribution of the chemical characteristics of the six samples to the variances of anammox bacterial communities was assessed with variance partitioning analysis and canonical correspondence analysis (CCA) using the package vegan[2] of the R program (v.3.5.3; see text footnote 1).

RESULTS

Physicochemical Analysis of the Sampling Sites

The basic physicochemical properties of hot spring samples collected from Conghua are shown in **Table 1**. There was little change in pH (6.5–7.5) and temperature (49–58°C) among all of the sites. A high degree of heterogeneity in concentrations of total organic carbon, total nitrogen, and ammonium was observed in the examined samples, which varied within the

[1]http://www.r-project.org/

[2]https://CRAN.R-project.org/package=vegan

TABLE 1 | Physicochemical characteristics of the samples collected from the hot springs of Conghua, China.

Samples	Temperature (°C)	pH	TOC (mg/L)	TC (mg/L)	DIC (mg/L)	TN (mg/L)	NH_4^+ (mg/L)	P (mg/L)	NO_3^- (mg/L)	NO_2^- (mg/L)
GWH1	49	7.20	0.054	21.250	21.190	25.220	0.261	0.000	0.020	0.002
GWH2	54	6.50	0.576	28.140	27.560	17.460	0.271	0.002	0.036	0.004
GWH3	52	7.00	1.294	29.030	27.740	4.560	0.046	0.000	0.036	0.005
LXRL1	58	6.50	14.23	31.170	16.940	23.990	0.121	0.004	0.027	0.030
LXRR1	55	7.50	1.269	28.680	27.410	22.190	0.100	0.035	0.113	0.010
LXRR2	54.5	7.50	7.488	36.050	28.560	21.290	0.005	0.004	0.103	0.003

TOC, total organic carbon; TC, total carbon; DIC, dissolved inorganic carbon; TN, total nitrogen; P, soluble reactive phosphorus; NH_4^+, ammonia; NO_3^-, nitrate; NO_2^-, nitrite.

range of 0.054–14.230, 4.560–25.220, and 0.005–0.271 mg/L, respectively. All samples had high concentrations of total carbon and dissolved inorganic carbon with values ranging from 21.250 to 36.050 mg/L and from 16.940 to 28.560 mg/L. However, concentrations of nitrate (0.020–0.113 mg/L), nitrite (0.002–0.030 mg/L), and soluble reactive phosphorus (0–0.350 mg/L) were relatively low at all sites.

Diversity of Anammox Bacteria

After quality control, a total of 181,802 high-quality anammox bacterial 16S rRNA gene sequences were obtained ranging between 13,884 and 57,231 sequences per sample (**Table 2**), which were clustered into 109 operational taxonomic units (OTUs) based on the 97% similarity level. The numbers of anammox bacterial OTUs in GWH1, GWH2, GWH3, LXRL1, LXRR1, and LXRR2 were 6, 41, 37, 15, 23, and 46, respectively. Each sample had its own unique anammox bacterial OTUs, with the proportion of independent anammox bacterial OTUs in samples GWH1, GWH2, GWH3, LXRL1, LXRR1, and LXRR2 of 50.0, 41.5, 40.5, 40.0, 47.8, and 60.9%, respectively, however, no OTUs were shared by all the samples (**Figure 1**). The alpha diversity of anammox bacteria was estimated by calculating the Shannon, Simpson, and Pielou indices. The Shannon and Simpson indices showed that the highest diversity of anammox bacteria were detected in sample GWH3 (4.50 and 0.94, respectively), whereas sample GWH1 (1.13) and sample LXRR1 (0.40) had the lowest diversity based on the Shannon index and Simpson index, respectively. The Pielou of anammox bacteria ranged from 0.35 to 0.86. Rarefaction curves for the anammox bacterial 16S rRNA gene at 97% similarity showed that the high-throughput sequencing could supply enough bio-information to investigate the community composition and diversity of anammox bacteria in the current study.

TABLE 2 | Diversity characteristics and abundance of anammox bacterial 16S rRNA gene.

Sample	High-quality sequences	OTUs	Shannon	Simpson	Pielou	Abundance (copies L^{-1})
GWH1	31,266	6	1.13	0.45	0.44	1.20×10^7
GWH2	23,742	41	2.84	0.69	0.53	2.30×10^5
GWH3	13,884	37	4.5	0.94	0.86	8.51×10^5
LXRL1	28,754	15	2.34	0.74	0.6	1.60×10^4
LXRR1	57,231	23	1.6	0.4	0.35	5.55×10^4
LXRR2	26,652	46	4.48	0.91	0.81	2.26×10^5

Community Composition of Anammox Bacteria

Based on the phylogenetic analysis, anammox-like bacteria that were closely affiliated to five known anammox bacterial genera were detected in the hot springs of Conghua. Among the 109 anammox bacterial OTUs in the collected samples, 65 OTUs were affiliated with "*Candidatus* Brocadia" (47.53%), "*Candidatus* Kuenenia" (0.134%), "*Candidatus* Anammoxoglobus" (0.896%), "*Candidatus* Jettenia" (0.133%), and "*Candidatus* Scalindua" (0.005%). The remaining unknown anammox bacterial OTUs accounted for 51.302%, which indicated that many putative novel anammox bacteria existed in the hot spring of Conghua. "*Candidatus* Brocadia" was the dominant genus in samples GWH2, GWH3, LXRL1, and LXRR2, while the putative novel anammox bacteria were dominant in samples GWH1 and LXRR1. "*Candidatus* Jettenia" was detected in GWH2, GWH3, and LXRR2; however, "*Candidatus* Kuenenia" and "*Candidatus* Scalindua" were only observed in LXRR2 and GWH3, respectively. The community structure of anammox bacteria in hot springs at the genus level is depicted in **Figure 2**.

The anammox bacterial distribution pattern was verified at the OTU level. A phylogenetic tree (**Figure 3**) was constructed to exhibit relationships between the representative sequences of the top 30 dominant OTUs and anammox bacterial reference sequences from the NCBI database. The result showed that most OTUs were widely distributed in the six samples. Seven phylogenetic clades were shown in the phylogenetic tree with five clades containing dominant OTUs. In *Brocadia* clade, eighteen OTUs were retrieved from the samples. OTU 17 and OTU 25 were assigned to the "*Candidatus* Kuenenia" and "*Candidatus* Anammoxoglobus" clades, respectively, and ten OTUs belonged to the putative novel anammox genera, which means that genes of "*Candidatus* Brocadia" had the highest diversity, and "*Candidatus* Kuenenia" and "*Candidatus* Anammoxoglobus" had a lower diversity. BLASTN analyses revealed that the retrieved anammox bacterial 16S rRNA gene sequences in the hot spring of Conghua were also closely affiliated with sequences detected in paddy field soil, aquatic ecosystem, reactor, and sediment from the rhizosphere, wetland, riparian, estuarine, freshwater lake, and sea with a 92.65–100% sequence similarity. **Supplementary Table 1** shows that the optimal growth temperatures of most targets including the sequences from petroleum reservoirs and hydrothermal vents were below 35°C.

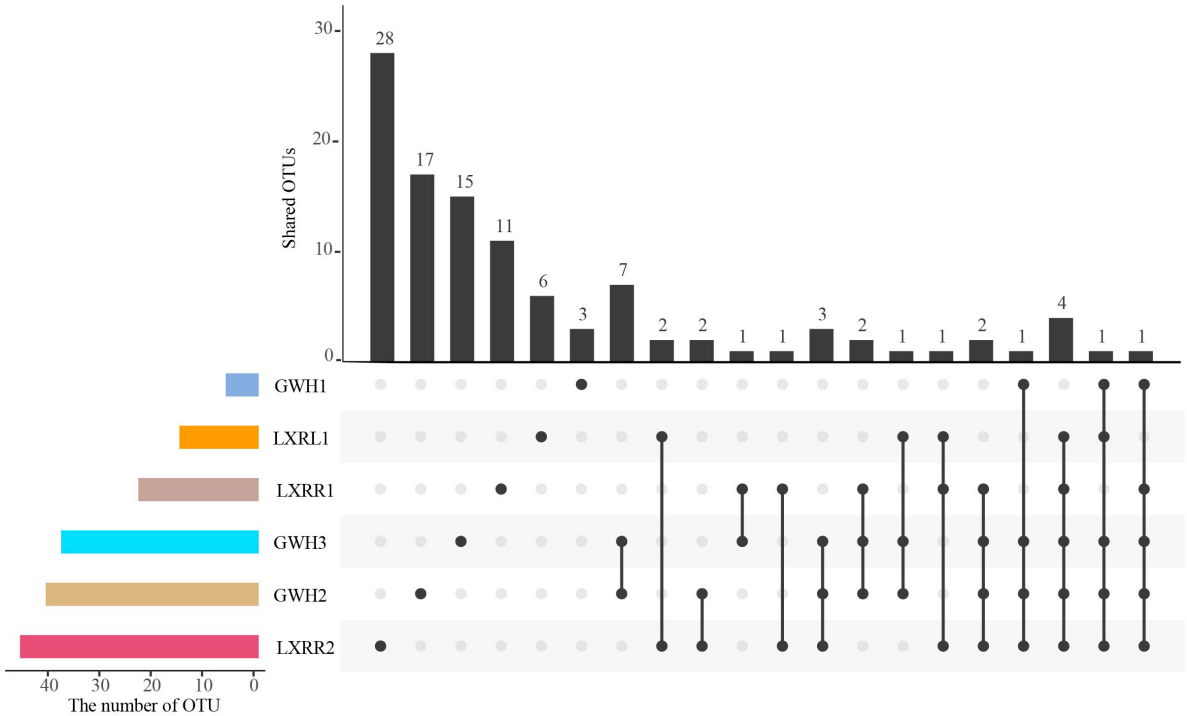


FIGURE 1 | An UpSetR plot showing the shared and unique OTUs at 97% similarity among the samples. UpSetR is an R package; it employs a matrix-based layout to show intersections of sets and their sizes, and it is implemented using ggplot2 and allows data analysts to easily generate UpSet plots for their own data. The strip at the bottom left shows the number of OTUs included in each sample. The dot and line at the bottom right represent the subsets of samples. The number of relevant OTUs in each subset is represented in the histogram, which is the upper part of the whole plot.

The Abundance of Anammox Bacteria

Anammox bacterial 16S rRNA genes were used to estimate the abundance of anammox bacteria by the qPCR method. In this study, melting curve analyses confirmed that fluorescent signals were derived from the specific PCR products during qPCR quantification. The qPCR results with 97.43% amplification efficiency showed that the abundance of anammox bacteria ranged from 1.60×10^4 to 1.20×10^7 copies L^{-1} (**Table 2**). Sample GWH1 (1.20×10^7 copies L^{-1}) had the highest abundance of anammox bacteria, followed by sample GWH3 (8.51×10^5 copies L^{-1}). The lowest abundance of anammox bacteria was observed in sample LXRL1 (1.60×10^4 copies L^{-1}).

Effects of Environmental Factors on Microbial Communities

CCA was conducted to show the potential relationships between environmental variables and the microbial communities of anammox bacteria in hot spring samples (**Figure 4**). The environmental factors in the first two CCA axes, respectively, explained 30.05 and 25.34% of the total variance for anammox bacterial OTU composition. Results showed that total organic carbon, ammonium, and nitrate were found to be the main factors affecting the distribution of anammox bacteria in Conghua hot spring.

Spearman's correlation analysis was used to illustrate the correlations among anammox gene abundances, diversity, and environmental factors. The results indicated that anammox bacterial abundance was negatively correlated with temperature ($p < 0.01$) and soluble reactive phosphorus ($p < 0.05$) (**Table 3**). Shannon diversity was negatively correlated with total nitrogen ($p < 0.05$). Anammox bacterial OTUs illustrated significantly positive correlations with dissolved inorganic carbon ($p < 0.05$). In addition, total organic carbon had a highly significant positive correlation with total carbon ($p < 0.01$), and soluble reactive phosphorus had a significantly positive correlation with temperature ($p < 0.05$) (**Figure 5**).

DISCUSSION

This study was the first time to investigate the diversity of anammox bacteria and to quantify the number of anammox 16S rRNA genes in hot springs of China. The influences of environmental factors on the distribution and diversity of anammox bacterial communities were also studied. Our data indicated that anammox bacteria can thrive in the hot spring of China, including putative novel taxa, which may imply that anammox bacteria could play a significant role in the nitrogen cycle of hot spring environments and a more in-depth research is necessary. Considering the detection of anammox bacteria from hot springs of California and Nevada (Jaeschke et al., 2009), we speculated that anammox bacteria might be widely distributed in the hot spring ecosystem. As expected, we reported

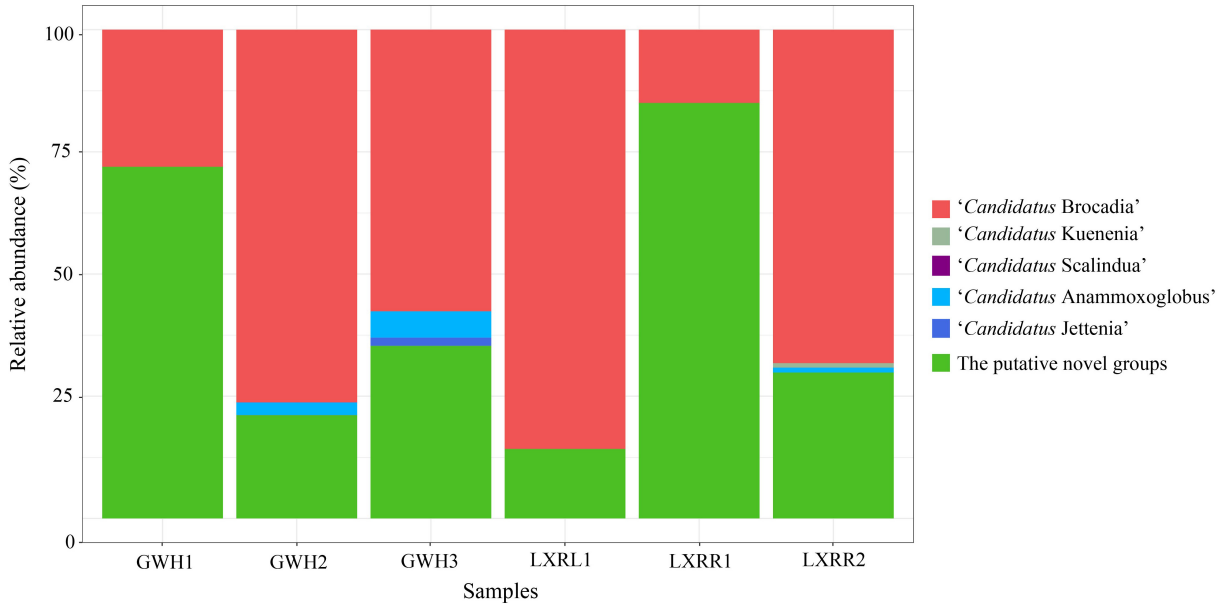


FIGURE 2 | Taxonomic classifications of anammox 16S rRNA gene sequences retrieved from the hot spring at the genus level. The colors indicate different taxa. The horizontal axis is the sample, and the vertical axis is the relative abundance.

the presence of anammox bacteria in Chinese hot springs during this study, which expanded our knowledge of the biogeography of anammox bacteria. Our study illustrated a high biodiversity of anammox bacteria in the hot spring of Conghua with five known genera of anammox bacteria ("*Candidatus* Scalindua," "*Candidatus* Brocadia," "*Candidatus* Kuenenia," "*Candidatus* Jettenia," and "*Candidatus* Anammoxoglobus"), and "*Candidatus* Brocadia" was identified as the dominant genus. To date, the diversity of anammox bacteria has been studied in various ecosystems. A low diversity of anammox bacteria was reported in marine ecosystems, which was mainly restricted to "*Candidatus* Scalindua" (Schmid et al., 2007). Anammox bacteria diversity in the freshwater ecosystem was higher than in the marine ecosystem; for example, the *hzsB* gene sequences were closely affiliated to "*Candidatus* Kuenenia" and "*Candidatus* Brocadia" in some freshwater extreme environments (Zhu et al., 2015). Anammox diversity in the terrestrial systems was higher in soil than in freshwater environments; for example, in wetland and dryland soils, three different genera of anammox bacteria were detected, including "*Candidatus* Kuenenia," "*Candidatus* Brocadia," and "*Candidatus* Jettenia" (Zhao et al., 2018). In acidic red soils, anammox bacteria were related to "*Candidatus* Brocadia," "*Candidatus* Scalindua," "*Candidatus* Kuenenia," and "*Candidatus* Anammoxoglobus" (Wu et al., 2018). The above discussions indicated that in natural ecosystems, the highest diversity of anammox bacteria occurred in terrestrial systems, followed by freshwater systems and marine systems, which was in accordance with our finding that anammox bacteria had a higher diversity in hot springs. Our finding was also consistent with the understanding that "*Candidatus* Kuenenia," "*Candidatus* Brocadia," "*Candidatus* Anammoxoglobus," and "*Candidatus* Jettenia" co-occurred in terrestrial environments (Gao et al., 2018). Furthermore, the microbial diversity of anammox bacteria exhibited a negatively significant correlation with total nitrogen, which was inconsistent with previous results that a higher total nitrogen concentration may provide an environment favorable for the distribution and growth of anammox bacteria (Shen et al., 2016).

The community composition of anammox bacteria in the six samples was different. The genera "*Candidatus* Brocadia," "*Candidatus* Scalindua," "*Candidatus* Anammoxoglobus," and "*Candidatus* Jettenia" were detected in GWH2, "*Candidatus* Brocadia," "*Candidatus* Anammoxoglobus," and "*Candidatus* Jettenia" were present in GWH3, and "*Candidatus* Brocadia," "*Candidatus* Anammoxoglobus," and "*Candidatus* Kuenenia" were observed in LXRR2. Among all the known anammox bacteria, "*Candidatus* Brocadia" was found in all the six samples; this finding might be related to the diverse metabolic pathways reported to "*Candidatus* Brocadia" (Gori et al., 2011), making it the most common anammox genus in terrestrial ecosystems, including the hot spring ecosystem. Interestingly, "*Candidatus* Brasilis" was found in 2011 (Viancelli et al., 2011), and it included only one *Candidatus* species to date. However, since 2011, there was no report about it, and among the six known anammox bacteria, "*Candidatus* Brasilis" was not detected in this study, which might be due to that "*Candidatus* Brasilis" was not widely distributed, or the primers might not suitable for "*Candidatus* Brasilis."

So far, there are 23 *Candidatus* species from six different genera of anammox bacteria that have been discovered from different environments. There are still a large number of anammox bacteria waiting to be excavated. For example, the putative novel anammox bacteria in surface sediment from the Dongjiang River through the Pearl River Estuary to the South

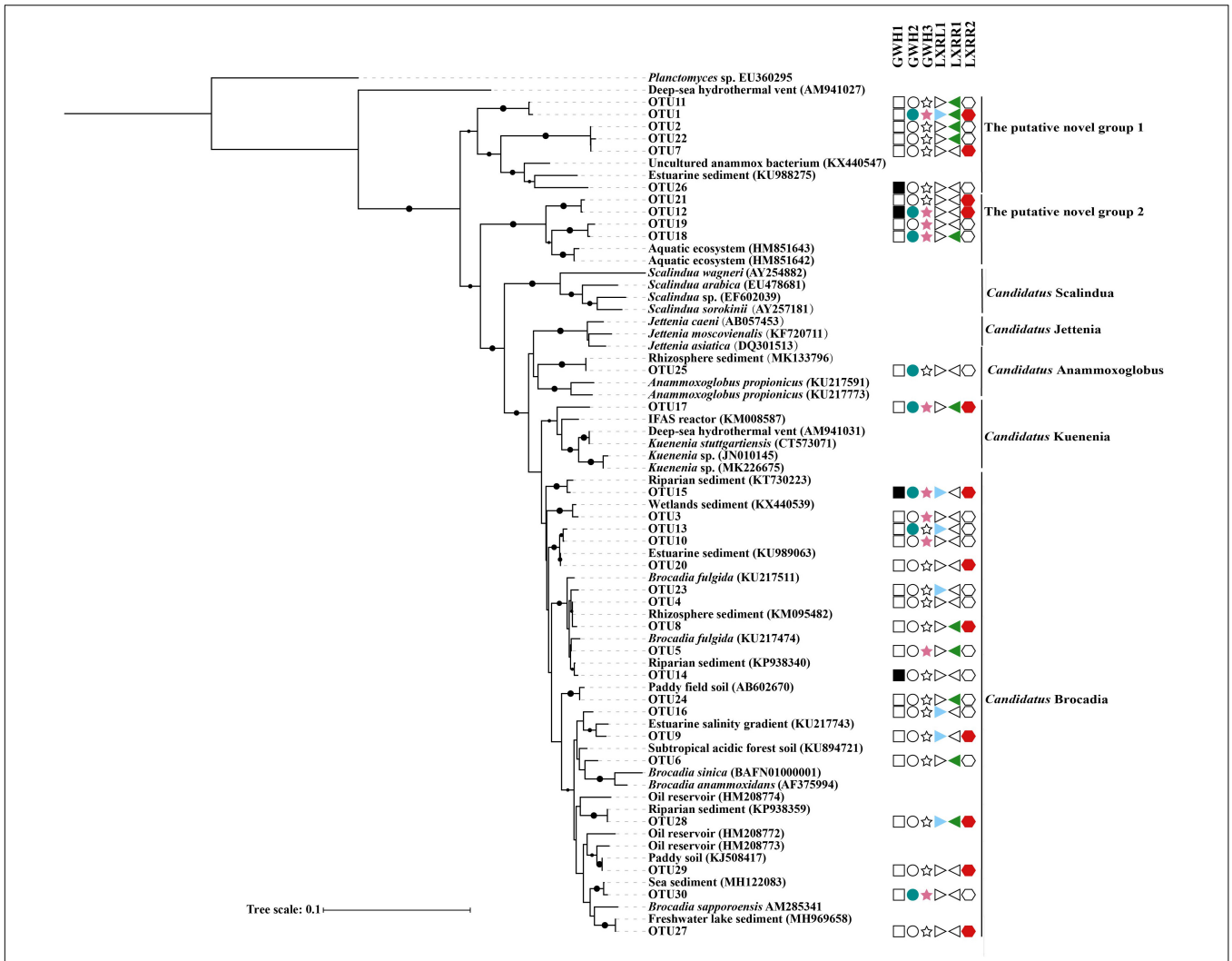


FIGURE 3 | The Neighbor-Joining tree showing the relationship among the top 30 dominant OTU sequences and the reference sequences from the NCBI database. Different shapes with different colors were used to code 6 samples, and empty symbols represent no sequence in the corresponding sample. Bootstrap values were 1,000 replicates. The scale bar represented 10% of sequence divergence.

China Sea accounted for 3.79% (Li et al., 2020). Similarly, the putative novel anammox bacteria accounted for 6.83% in the acidic red soils of southern China (Wu et al., 2018), and the proportion of putative novel anammox bacteria in sediment cores of the Pearl River Estuary was 45.12% (Wu et al., 2020). The putative novel anammox bacteria were also detected in freshwater wetlands of southeastern China (Shen et al., 2016) and Dongjiang River (Sun et al., 2014a). In our study, the putative novel anammox bacteria were up to 51.30%, which was higher than in the previous studies (Wu et al., 2018, 2020; Li et al., 2020). Additionally, two new clusters of putative anammox bacterial 16S rRNA genes were formed in the phylogenetic tree (**Figure 3**). Sequences of OTU1, OTU2, OTU7, OTU11, OTU22, and OTU26 showed 92.65–94.53% identity to the 16S rRNA genes of known anammox bacteria from estuarine sediments of southeastern China, and sequences of OTU12, OTU18, OTU19, OTU21, and OTU26 showed 96.12–96.77% identity to the 16S rRNA gene of anammox bacteria from aquatic ecosystems (Hirsch et al., 2011). It is worth noting that the predicted optimal growth temperatures of OTU1, OTU2, OTU7, OTU11, and OTU22 were ranging from 47.90 to 50.37°C (**Supplementary Table 1**). These results revealed that hot springs harbored a large number of novel anammox bacteria, indicating that anammox bacteria in the hot spring of Conghua were worthy for an in-deep mining. Spearman's rank correlation analysis showed that dissolved inorganic carbon had a significant positive correlation with OTUs, which was consistent with previous results that anammox bacteria are chemolithoautotrophic microorganisms (Strous et al., 2006; Jetten et al., 2010). This implied that dissolved inorganic carbon was the key environmental factor shaping the distribution of anammox bacteria in the hot spring of Conghua. Interestingly, no significant correlation was observed between the OTUs and

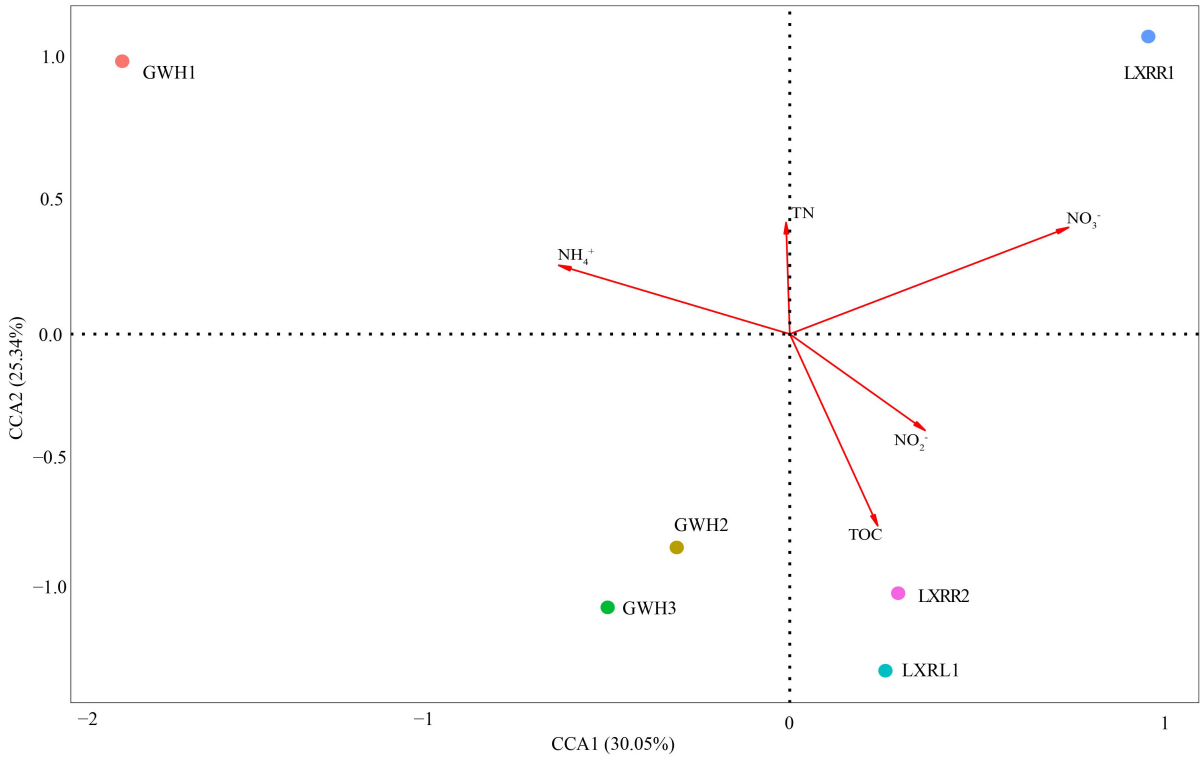


FIGURE 4 | Canonical correspondence analysis (CCA) ordination plots for the first two principal dimensions to show the relationship between anammox bacterial community composition and physicochemical parameters. The percentage indicates the interpretation of axis to the diversification of anammox bacterial community. Correlations between environmental variables and CCA axes are represented by the length and angle of arrows. Circles with different colors represent diverse samples. Arrow length expresses relative importance of an environmental variable; the longer the arrow, the greater influence of environmental factors. Dropping perpendicular to the arrows from each of the samples indicates the strength of the environmental factor's influence on the sample; the shorter distance, the stronger the effect of the environmental factor on the sample. The sample is located in the same direction as the arrow, indicating that the environmental factor is positively correlated with the change of the bacterial community, and the sample is located in the opposite direction of the arrow, indicating that the environmental factor is negatively related to the change of the bacterial community. TOC and TN mean total organic carbon and total nitrogen, respectively.

diversity, indicating that each sample had its own unique dominant group, which was further supported by the results from UpSetR analysis.

Anammox bacteria were active at 6–43°C with an optimal temperature of 35°C in laboratory bioreactors. However, anammox bacteria were also observed at 52°C in hot spring (Jaeschke et al., 2009), 72°C in petroleum reservoirs (Li et al., 2010), 75°C in freshwater (Zhu et al., 2015), and even 60–80°C in hydrothermal vents (Russ et al., 2013). The dominant anammox species in these high-temperature habitats were either *Brocadia* or *Kuenenia*. In our study, the samples have a temperature of up to 58°C with *Brocadia* as a dominant anammox species, which is consistent with the findings that some anammox species can endure high temperatures. This result indicated that anammox bacteria were not only mesophilic but also thermostable. The special ecophysiological feature of anammox bacteria should be attributed to their specific cellular structure. To maintain the constant membrane fluidity at different temperatures, anammox bacteria could modify their membrane composition to adapt to the change of temperature (Jaeschke et al., 2009). Overall, the temperature may not be the limiting factor for the occurrence of anammox bacteria, and anammox can occur in a wide temperature range. However, the temperature tolerance and the optimal temperature in high-temperature habitats need to be studied in future research.

Among the anammox bacteria analysis, many anammox bacterial DNAs were collected from the sediments or soil of different ecosystems. The abundance of anammox bacteria was measured by the number of gene copies per gram of soil. For example, the global distribution of anaerobic ammonia oxidation bacteria in wetland, dryland, and groundwater aquifer revealed by the *hzsB* gene showed that the gene abundance of anammox bacteria in wetlands was 5.42×10^4–9.56×10^6 copies/g of dry soil, sediments from freshwater rivers at 6.9×10^4–8.9×10^5 copies/g of dry soil, and freshwater lakes at 5.4×10^4–9.6×10^6 copies/g of dry soil, and that in groundwater aquifers was high, ranging from 5.93×10^5 to 9.12×10^6 copies/g of dry soil (Wang et al., 2019). However, some anammox bacterial DNA was extracted from the water, and the abundance of anammox bacteria was measured by the number of cells per milliliter sample (Byrne et al., 2009; Li et al., 2010; Sun et al., 2014a). In our study, the abundance of anammox bacteria ranged from 1.60×10^4 to 1.20×10^7 copies L^{-1}. Based on the estimation

TABLE 3 | Spearman rank correlation analysis of environmental factors with the abundance and diversity of anammox bacterial community in Conghua hot spring.

	Temperature	pH	TOC	TC	DIC	TN	NH_4^+	P	NO_3^-	NO_2^-	Abundance	Shannon diversity	OTUs
Temperature	1.00	−0.03	0.71	0.60	−0.26	0.14	−0.20	0.88*	0.41	0.77	−1.00**	−0.03	0.09
pH		1.00	−0.09	0.15	0.38	0.15	−0.62	0.36	0.58	−0.32	0.03	−0.15	0.15
TOC			1.00	0.94**	0.09	−0.14	−0.60	0.44	0.20	0.60	−0.71	0.49	0.26
TC				1.00	0.37	−0.26	−0.77	0.44	0.38	0.37	−0.60	0.60	0.49
IC					1.00	−0.77	−0.60	−0.09	0.58	−0.37	0.26	0.77	0.89*
TN						1.00	0.31	0.18	−0.46	−0.09	−0.14	−0.89*	−0.77
NH_4^+							1.00	−0.26	−0.55	−0.09	0.20	−0.54	−0.43
P								1.00	0.67	0.53	−0.88*	−0.18	0.18
NO_3^-									1.00	0.23	−0.41	0.32	0.64
NO_2^-										1.00	−0.77	0.09	−0.14
Abundance											1.00	0.03	−0.09
Shannon diversity												1.00	0.77
OTUs													1.00

Positive numbers indicate a positive correlation, and negative numbers indicate a negative correlation. "" denotes a p-value of < 0.05, and "**" denotes a p-value of < 0.01, the lower p-value, the more significant correlation.*
TOC, total organic carbon; TC, total carbon; DIC, resolved inorganic carbon; TN, total nitrogen; P, soluble reactive phosphorus; NH_4^+, ammonia; NO_3^-, nitrate; NO_2^-, nitrite.

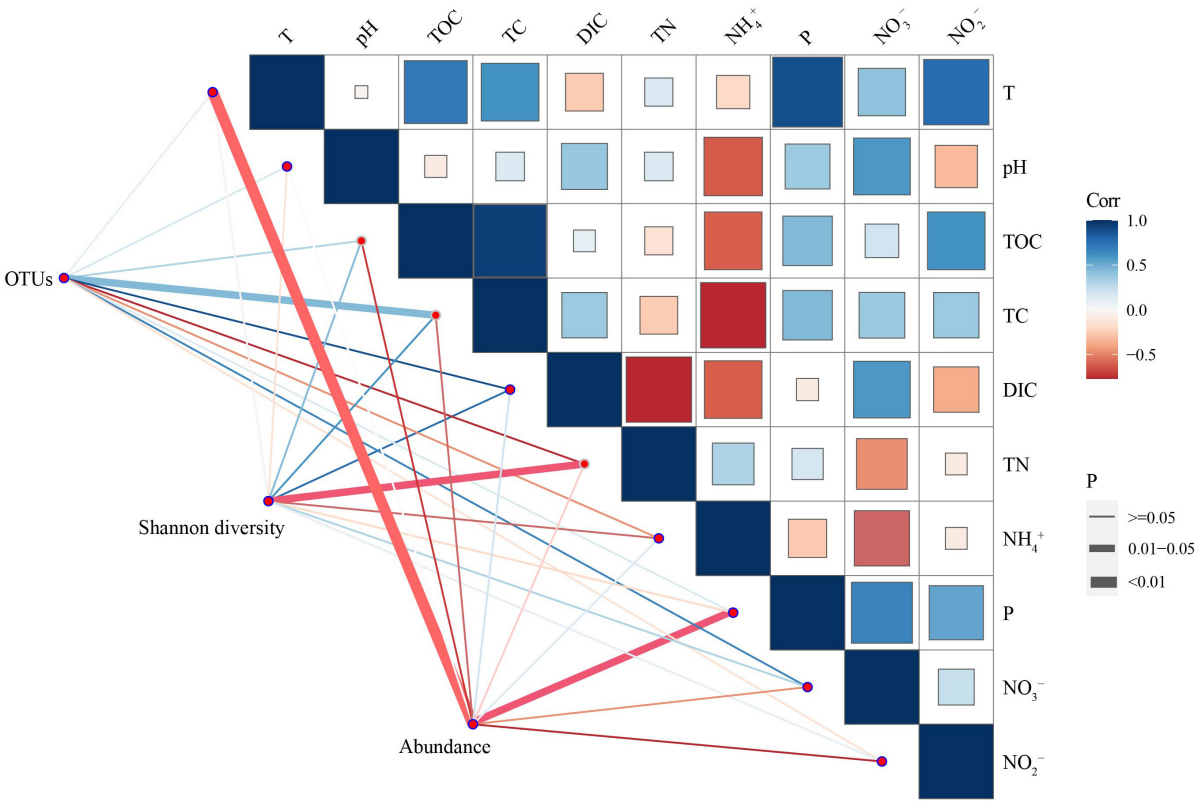


FIGURE 5 | Spearman's correlation analysis based on the 16S rRNA gene of anammox bacterial diversity, abundance, and physicochemical parameters (including T, pH, TOC, TC, DIC, TN, NH_4^+, P, NO_2^-, and NO_3^-) in six hot spring samples. The color scale denotes the nature of the correlation, with blue and red denoting the positive and negative correlations, respectively. The line thickness denotes different *p*-values; the thicker the line, the more significant correlation. On the contrary, the thinner the line, the less significant the correlation. T, temperature; TOC, total organic carbon; TC, total carbon; DIC, dissolved inorganic carbon; TN, total nitrogen; P, soluble reactive phosphorus; NH_4^+, ammonia; NO_3^-, nitrate; NO_2^-, nitrite.

of 3.6 copy numbers of 16S rRNA gene per bacterial cell genome (Harms et al., 2003), the anammox bacterial abundance was calculated to be 4.40–3.3 × 10^3 cells mL^{-1}. The overall abundance of anammox bacteria observed in this study was lower than the values reported for most deep-sea sediments (Schmid et al., 2007), and high-temperature oil reservoirs (Li et al., 2010), but almost had the same range as reported in Dongjiang River (Sun et al., 2014a). Temperature and soluble reactive phosphorus were negatively correlated with anammox abundance, suggesting that higher temperature and soluble

reactive phosphorus concentration could restrain the growth of anammox bacteria.

Previous studies have shown that many different environmental factors can affect anammox bacterial diversity and distribution. For example, temperature was a key environmental factor shaping the distribution and diversity of anammox bacteria in the coastal estuaries of China (Yang et al., 2017). Organic carbon influenced the distribution of anammox bacteria in Qiantang River sediments (Hu et al., 2012) and Yangtze estuary marsh sediment (Hou et al., 2013). Ammonium and nitrite amendments would change the community compositions of anammox bacteria in mangrove sediments to ammonium or nitrite (Li and Gu, 2013), and salinity was one of the key factors driving the biogeography of anammox bacteria (Sonthiphand et al., 2014). This selection is owing to the physiological properties of anammox bacteria, such as the optimum growth temperature and pH, and the affinity for ammonia and nitrite (Oshiki et al., 2011). In the present study, a CCA test was conducted to find the potential relationship between the distribution of anammox bacteria and the environmental factors as the steepest gradient of detruded correspondence analysis was higher than 4.0, and the result of CCA suggested that ammonium and total organic carbon were the main factors affecting the distribution of anammox bacteria in hot spring of Conghua, Additionally, nitrate content had a significant contribution to anammox bacterial community structure, which may be attributed to an increased supply of nitrite *via* reduction of nitrate, and have also been observed in Dongjiang River (Sun et al., 2014b). However, the role of other unidentified ecological parameters except for the present factors also cannot be ruled out.

CONCLUSION

Evidence for the occurrence of anammox in this study was demonstrated by the amplification of 16S rRNA gene sequences and quantitative PCR analysis, suggesting that anammox bacteria were present in the hot springs of Conghua. Phylogenetic analysis of the 16S rRNA gene sequences showed a higher diversity of anammox bacteria in hot springs of Conghua than other high-temperature habitats, such as deep-sea hydrothermal vent, hot spring, and the petroleum reservoir. The anammox community was dominated by "*Candidatus* Brocadia" and harbored putative novel anammox bacterial candidates. Nitrate played a key environmental factor in regulating the distribution of the anammox bacterial community in Conghua hot springs. These results extend our understanding of the community dynamics, environmental importance, and biogeography of anammox bacteria in hot spring ecosystems and would guide for future enrichment strategies of anammox bacteria in Conghua hot springs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI SRA BioProject, accession no: PRJNA751159; National Omics Data Encyclopedia, accession no: OEP002743.

AUTHOR CONTRIBUTIONS

W-JL, LL, and J-YJ jointly conceived the study. W-DX, Z-TL, and J-YZ performed the sample collection. LL designed the experiments. A-PL, Y-ZM, and M-ML did the experiments. LL, X-TZ, and J-YJ analyzed the culture-independent data. LL, J-YJ, MPNR, NS, and W-JL wrote the manuscript. W-JL gave the final approved of the manuscript to be published. All authors read and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (Nos. 91951205 and 32170101), the Natural Science Foundation of Guangdong Province, China (No. 2016A030312003), the China Postdoctoral Science Foundation (No. 2019M653156), the Fundamental Research Funds for the Universities (No. 19lgpy204), the Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515110227), and the National Science and Technology Fundamental Resources Investigation Program of China (No. 2021FY100900).

ACKNOWLEDGMENTS

We would like to thank Jia-Peng Wu from Guangzhou University for his kind help. Genewiz, Suzhou, China, is gratefully acknowledged for 16S rRNA gene amplicon sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.739234/full#supplementary-material

REFERENCES

Ali, M., Oshiki, M., Awata, T., Isobe, K., Kimura, Z., Yoshikawa, H., et al. (2015). Physiological characterization of anaerobic ammonium oxidizing bacterium '*Candidatus* Jettenia caeni'. *Environ. Microbiol.* 17, 2172–2189. doi: 10.1111/1462-2920.12674

Araujo, J., Campos, A., Correa, M., Silva, E., Matte, M., Matte, G., et al. (2011). Anammox bacteria enrichment and characterization from municipal activated sludge. *Water Sci. Technol.* 64, 1428–1434. doi: 10.2166/wst.2011.632

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon

sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90. doi: 10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0252-6

Botchkova, E., Litti, Y. V., Novikov, A., Grouzdev, D., Bochkareva, E., Beskorovayny, A., et al. (2018). Description of "*Candidatus* jettenia ecosi" sp. nov., a new species of anammox bacteria. *Microbiology* 87, 766–776. doi: 10.1134/S002626171806005X

Brandsma, J., van de Vossenberg, J., Risgaard-Petersen, N., Schmid, M. C., Engström, P., Eurenius, K., et al. (2011). A multi-proxy study of anaerobic ammonium oxidation in marine sediments of the Gullmar Fjord, Sweden. *Environ. Microbiol. Rep.* 3, 360–366. doi: 10.1111/j.1758-2229.2010.00233.x

Broda, E. (1977). Two kinds of lithotrophs missing in nature. *Z. Allg. Mikrobiol.* 17, 491–493. doi: 10.1002/jobm.3630170611

Byrne, N., Strous, M., Crépeau, V., Kartal, B., Birrien, J. L., Schmid, M., et al. (2009). Presence and activity of anaerobic ammonium-oxidizing bacteria at deep-sea hydrothermal vents. *ISME J.* 3, 117–123. doi: 10.1038/ismej.2008.72

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

Dang, H., Chen, R., Wang, L., Guo, L., Chen, P., Tang, Z., et al. (2010). Environmental factors shape sediment anammox bacterial communities in hypernutrified Jiaozhou Bay, China. *Appl. Environ. Microbiol.* 76, 7036–7047. doi: 10.1128/AEM.01264-10

Fuchsman, C. A., Staley, J. T., Oakley, B. B., Kirkpatrick, J. B., and Murray, J. W. (2012). Free-living and aggregate-associated Planctomycetes in the Black Sea. *FEMS Microbiol. Ecol.* 80, 402–416. doi: 10.1111/j.1574-6941.2012.01306.x

Gao, D., Wang, X., Liang, H., Wei, Q., Dou, Y., and Li, L. (2018). Anaerobic ammonia oxidizing bacteria: ecological distribution, metabolism, and microbial interactions. *Front. Environ. Sci. Eng.* 12:10. doi: 10.1007/s11783-018-1035-x

Gori, F., Tringe, S. G., Kartal, B., Machiori, E., and Jetten, M. S. (2011). The metagenomic basis of anammox metabolism in Candidatus 'Brocadia fulgida'. *Biochem. Soc. Trans.* 39, 1799–1804. doi: 10.1042/BST20110707

Hamersley, M. R., Lavik, G., Woebken, D., Rattray, J. E., Lam, P., Hopmans, E. C., et al. (2007). Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. *Limnol. Oceanogr.* 52, 923–933. doi: 10.4319/lo.2007.52.3.0923

Harms, G., Layton, A. C., Dionisi, H. M., Gregory, I. R., Garrett, V. M., Hawkins, S. A., et al. (2003). Real-time PCR quantification of nitrifying bacteria in a municipal wastewater treatment plant. *Environ. Sci. Technol.* 37, 343–351. doi: 10.1021/es0257164

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, F., Radax, R., Woebken, D., Holtappels, M., Lavik, G., Rapp, H. T., et al. (2009). Complex nitrogen cycling in the sponge *Geodia barretti*. *Environ. Microbiol.* 11, 2228–2243. doi: 10.1111/j.1462-2920.2009.01944.x

Hong, Y. G., Yin, B., and Zheng, T. L. (2011). Diversity and abundance of anammox bacterial community in the deep-ocean surface sediment from equatorial Pacific. *Appl. Microbiol. Biotechnol.* 89, 1233–1241. doi: 10.1007/s00253-010-2925-4

Hou, L., Zheng, Y., Liu, M., Gong, J., Zhang, X., Yin, G., et al. (2013). Anaerobic ammonium oxidation (anammox) bacterial diversity, abundance, and activity in marsh sediments of the Yangtze Estuary. *J. Geophys. Res. Biogeosci.* 118, 1237–1246. doi: 10.1002/jgrg.20108

Hu, B. L., Rush, D., van der Biezen, E., Zheng, P., van Mullekom, M., Schouten, S., et al. (2011). New anaerobic, ammonium-oxidizing community enriched from peat soil. *Appl. Environ. Microbiol.* 77, 966–971. doi: 10.1128/AEM.02402-10

Hu, B. L., Shen, L. D., Zheng, P., Hu, A. H., Chen, T. T., Cai, C., et al. (2012). Distribution and diversity of anaerobic ammonium-oxidizing bacteria in the sediments of the Qiantang River. *Environ. Microbiol. Rep.* 4, 540–547. doi: 10.1111/j.1758-2229.2012.00360.x

Hu, B. L., Zheng, P., Tang, C. J., Chen, J. W., van der Biezen, E., Zhang, L., et al. (2010). Identification and quantification of anammox bacteria in eight nitrogen removal reactors. *Water Res.* 44, 5014–5020. doi: 10.1016/j.watres.2010.07.021

Hua, Z. S., Wang, Y. L., Evans, P. N., Qu, Y. N., Goh, K. M., Rao, Y. Z., et al. (2019). Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea. *Nat. Commun.* 10:4574. doi: 10.1038/s41467-019-12574-y

Humbert, S., Tarnawski, S., Fromin, N., Mallet, M. P., Aragno, M., and Zopfi, J. (2010). Molecular detection of anammox bacteria in terrestrial ecosystems: distribution and diversity. *ISME J.* 4, 450–454. doi: 10.1038/ismej.2009.125

Jaeschke, A., Op den Camp, H. J., Harhangi, H., Klimiuk, A., Hopmans, E. C., Jetten, M. S., et al. (2009). 16S rRNA gene and lipid biomarker evidence for anaerobic ammonium-oxidizing bacteria (anammox) in California and Nevada hot springs. *FEMS Microbiol. Ecol.* 67, 343–350. doi: 10.1111/j.1574-6941.2008.00640.x

Jetten, M. S. M., Op den Camp, H. J. M., Kuenen, J. G., and Strous, M. (2010). "Description of the order Brocadiales," in *Bergey's Manual of Systematic Bacteriology*, Vol. 4, eds N. R. Krieg, J. T. Staley, B. P. Hedlund, B. J. Paster, N. Ward, W. Ludwig, et al. (Heidelberg: Springer). doi: 10.1007/s10482-013-0019-x

Jiao, J. Y., Liu, L., Hua, Z. S., Fang, B. Z., Zhou, E. M., Salam, N., et al. (2021b). Microbial dark matter coming to light: challenges and opportunities. *Natl. Sci. Rev.* 8:nwaa280. doi: 10.1093/nsr/nwaa280

Jiao, J. Y., Fu, L., Hua, Z. S., Liu, L., Salam, N., Liu, P. F., et al. (2021a). Insight into the function and evolution of the Wood-Ljungdahl pathway in Actinobacteria. *ISME J.* 15, 3005–3018. doi: 10.1038/s41396-021-00935-9

Kartal, B., Rattray, J., van Niftrik, L. A., van de Vossenberg, J., Schmid, M. C., Webb, R. I., et al. (2007). Candidatus "Anammoxoglobus propionicus" a new propionate oxidizing species of anaerobic ammonium oxidizing bacteria. *Syst. Appl. Microbiol.* 30, 39–49. doi: 10.1016/j.syapm.2006.03.004

Kartal, B., Van Niftrik, L., Rattray, J., Van De Vossenberg, J. L., Schmid, M. C., Sinninghe Damsté, J., et al. (2008). Candidatus 'Brocadia fulgida': an autofluorescent anaerobic ammonium oxidizing bacterium. *FEMS Microbiol. Ecol.* 63, 46–55. doi: 10.1111/j.1574-6941.2007.00408.x

Kimura, H., Sugihara, M., Kato, K., and Hanada, S. (2006). Selective phylogenetic analysis targeted at 16S rRNA genes of thermophiles and hyperthermophiles in deep-subsurface geothermal environments. *Appl. Environ. Microbiol.* 72, 21–27. doi: 10.1128/aem.72.1.21-27.2006

Kuypers, M. M., Sliekers, A. O., Lavik, G., Schmid, M., Jørgensen, B. B., Kuenen, J. G., et al. (2003). Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* 422, 608–611. doi: 10.1038/nature01472

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, H., Chen, S., Mu, B. Z., and Gu, J. D. (2010). Molecular detection of anaerobic ammonium-oxidizing (anammox) bacteria in high-temperature petroleum reservoirs. *Microb. Ecol.* 60, 771–783. doi: 10.1007/s00248-010-9733-3

Li, M., and Gu, J. D. (2013). Community structure and transcript responses of anammox bacteria, AOA, and AOB in mangrove sediment microcosms amended with ammonium and nitrite. *Appl. Microbiol. Biotechnol.* 97, 9859–9874. doi: 10.1007/s00253-012-4683-y

Li, Y., Hong, Y., Wu, J., Wang, Y., and Ye, F. (2020). Spatial variability pattern of the anaerobic ammonia-oxidizing bacterial community across a salinity gradient from river to ocean. *Ecotoxicology* 30, 1743–1753. doi: 10.1007/s10646-020-02282-5

Lu, X., Sun, S., Zhang, Y. Q., Hollibaugh, J. T., and Mou, X. (2015). Temporal and vertical distributions of bacterioplankton at the Gray's Reef National Marine Sanctuary. *Appl. Environ. Microbiol.* 81, 910–917. doi: 10.1128/AEM.02802-14

Luo, Z. H., Narsing Rao, M. P., Chen, H., Hua, Z. S., Li, Q., Hedlund, B. P., et al. (2020). Genomic insights of "*Candidatus Nitrosocaldaceae*" based on nine new metagenome-assembled genomes, including "*Candidatus Nitrosothermus*" gen nov. and two new species of "*Candidatus Nitrosocaldus*". *Front. Microbiol.* 11:608832. doi: 10.3389/fmicb.2020.608832

Mohamed, N. M., Saito, K., Tal, Y., and Hill, R. T. (2010). Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *ISME J.* 4, 38–48. doi: 10.1038/ismej.2009.84

Moore, T. A., Xing, Y., Lazenby, B., Lynch, M. D., Schiff, S., Robertson, W. D., et al. (2011). Prevalence of anaerobic ammonium-oxidizing bacteria in contaminated groundwater. *Environ. Sci. Technol.* 45, 7217–7225. doi: 10.1021/es201243t

Mulder, A., Van de Graaf, A. A., Robertson, L., and Kuenen, J. (1995). Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor. *FEMS Microbiol. Ecol.* 16, 177–183. doi: 10.1111/j.1574-6941.1995.tb00281.x

Narita, Y., Zhang, L., Kimura, Z. I., Ali, M., Fujii, T., and Okabe, S. (2017). Enrichment and physiological characterization of an anaerobic ammonium-oxidizing bacterium '*Candidatus* Brocadia sapporoensis'. *Syst. Appl. Microbiol.* 40, 448–457. doi: 10.1016/j.syapm.2017.07.004

Nikolaev, Y. A., Kozlov, M., Kevbrina, M., Dorofeev, A., Pimenov, N., Kallistova, A. Y., et al. (2015). Candidatus "*Jettenia moscovienalis*" sp. nov., a new species of bacteria carrying out anaerobic ammonium oxidation. *Microbiology* 84, 256–262. doi: 10.1134/s0026261715020101

Oshiki, M., Mizuto, K., Kimura, Z. I., Kindaichi, T., Satoh, H., and Okabe, S. (2017). Genetic diversity of marine anaerobic ammonium-oxidizing bacteria as revealed by genomic and proteomic analyses of '*Candidatus* Scalindua japonica'. *Environ. Microbiol. Rep.* 9, 550–561. doi: 10.1111/1758-2229.12586

Oshiki, M., Shimokawa, M., Fujii, N., Satoh, H., and Okabe, S. (2011). Physiological characteristics of the anaerobic ammonium-oxidizing bacterium '*Candidatus* Brocadia sinica'. *Microbiology* 157, 1706–1713. doi: 10.1099/mic.0.048595-0

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740. doi: 10.1126/science.276.5313.734

Quan, Z. X., Rhee, S. K., Zuo, J. E., Yang, Y., Bae, J. W., Park, J. R., et al. (2008). Diversity of ammonium-oxidizing bacteria in a granular sludge anaerobic ammonium-oxidizing (anammox) reactor. *Environ. Microbiol.* 10, 3130–3139. doi: 10.1111/j.1462-2920.2008.01642.x

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

Rich, J. J., Dale, O. R., Song, B., and Ward, B. B. (2008). Anaerobic ammonium oxidation (anammox) in Chesapeake Bay sediments. *Microb. Ecol.* 55, 311–320. doi: 10.1007/s00248-007-9277-3

Russ, L., Kartal, B., Op Den Camp, H., Sollai, M., Le Bruchec, J., Caprais, J. C., et al. (2013). Presence and diversity of anammox bacteria in cold hydrocarbon-rich seeps and hydrothermal vent sediments of the Guaymas Basin. *Front. Microbiol.* 4:219. doi: 10.3389/fmicb.2013.00219

Schmid, M., Twachtmann, U., Klein, M., Strous, M., Juretschko, S., Jetten, M., et al. (2000). Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst. Appl. Microbiol.* 23, 93–106. doi: 10.1016/S0723-2020(00)80050-8

Schmid, M., Walsh, K., Webb, R., Rijpstra, W. I., van de Pas-Schoonen, K., Verbruggen, M. J., et al. (2003). Candidatus "*Scalindua brodae*," sp. nov., Candidatus "*Scalindua wagneri*," sp. nov., two new species of anaerobic ammonium oxidizing bacteria. *Syst. Appl. Microbiol.* 26, 529–538. doi: 10.1078/072320203770865837

Schmid, M. C., Risgaard-Petersen, N., Van De Vossenberg, J., Kuypers, M. M., Lavik, G., Petersen, J., et al. (2007). Anaerobic ammonium-oxidizing bacteria in marine environments: widespread occurrence but low diversity. *Environ. Microbiol.* 9, 1476–1484. doi: 10.1111/j.1462-2920.2007.01266.x

Shen, L. D., Wu, H. S., Gao, Z. Q., Cheng, H. X., Li, J., Liu, X., et al. (2016). Distribution and activity of anaerobic ammonium-oxidising bacteria in natural freshwater wetland soils. *Appl. Microbiol. Biotechnol.* 100, 3291–3300. doi: 10.1007/s00253-015-7191-z

Sonthiphand, P., Hall, M. W., and Neufeld, J. D. (2014). Biogeography of anaerobic ammonia-oxidizing (anammox) bacteria. *Front. Microbiol.* 5:399. doi: 10.3389/fmicb.2014.00399

Speth, D. R., Lagkouvardos, I., Wang, Y., Qian, P. Y., Dutilh, B. E., and Jetten, M. S. (2017). Draft genome of Scalindua rubra, obtained from the interface above the discovery deep brine in the Red Sea, sheds light on potential salt adaptation strategies in anammox bacteria. *Microb. Ecol.* 74, 1–5. doi: 10.1007/s00248-017-0929-7

Strous, M., Fuerst, J. A., Kramer, E. H., Logemann, S., Muyzer, G., van de Pas-Schoonen, K. T., et al. (1999). Missing lithotroph identified as new planctomycete. *Nature* 400, 446–449. doi: 10.1038/22749

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M. W., et al. (2006). Deciphering the evolution, and metabolism of an anammox bacterium from a community genome. *Nature* 440, 790–794. doi: 10.1038/nature04647

Sun, W., Xia, C., Xu, M., Guo, J., Wang, A., and Sun, G. (2014a). Diversity and distribution of planktonic anaerobic ammonium-oxidizing bacteria in the Dongjiang River, China. *Microbiol. Res.* 169, 897–906. doi: 10.1016/j.micres.2014.05.003

Sun, W., Xu, M. Y., Wu, W. M., Guo, J., Xia, C. Y., Sun, G. P., et al. (2014b). Molecular diversity and distribution of anammox community in sediments of the Dongjiang River, a drinking water source of Hong Kong. *J. Appl. Microbiol.* 116, 464–476. doi: 10.1111/jam.12367

Tan, S., Liu, J., Fang, Y., Hedlund, B. P., Lian, Z. H., Huang, L. Y., et al. (2019). Insights into ecological role of a new deltaproteobacterial order *Candidatus* Acidulodesulfobacterales by metagenomics and metatranscriptomics. *ISME J.* 13, 2044–2057. doi: 10.1038/s41396-019-0415-y

Thamdrup, B. (2012). New pathways and processes in the global nitrogen cycle. *Annu. Rev. Ecol. Syst.* 43, 407–428. doi: 10.1146/annurev-ecolsys-102710-145048

Van De Vossenberg, J., Rattray, J. E., Geerts, W., Kartal, B., Van Niftrik, L., Van Donselaar, E. G., et al. (2008). Enrichment and characterization of marine anammox bacteria associated with global nitrogen gas production. *Environ. Microbiol.* 10, 3120–3129. doi: 10.1111/j.1462-2920.2008.01643.x

Viancelli, A., Kunz, A., Esteves, P. A., Bauermann, F. V., Furukawa, K., Fujii, T., et al. (2011). Bacterial biodiversity from an anaerobic up flow bioreactor with ANAMMOX activity inoculated with swine sludge. *Braz. Arch. Biol. Technol.* 54, 1035–1041. doi: 10.1590/S1516-89132011000500022

Wang, T., Cheng, L., Zhang, W., Xu, X., Meng, Q., Sun, X., et al. (2017). Anaerobic ammonium-oxidizing bacteria in cow manure composting. *J. Microbiol. Biotechnol.* 27, 1288–1299. doi: 10.4014/jmb.1702.02065

Wang, Y., Xu, L., Wang, S., Ye, F., and Zhu, G. (2019). Global distribution of anaerobic ammonia oxidation (anammox) bacteria - field surveys in wetland, dryland, groundwater aquifer and snow. *Front. Microbiol.* 10:2583. doi: 10.3389/fmicb.2019.02583

Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978. doi: 10.1126/science.1086909

Woebken, D., Lam, P., Kuypers, M. M., Naqvi, S. W. A., Kartal, B., Strous, M., et al. (2008). A microdiversity study of anammox bacteria reveals a novel *Candidatus* Scalindua phylotype in marine oxygen minimum zones. *Environ. Microbiol.* 10, 3106–3119. doi: 10.1111/j.1462-2920.2008.01640.x

Wu, J., Hong, Y., He, X., Jiao, L., Wen, X., Chen, S., et al. (2018). Anaerobic ammonium oxidation in acidic red soils. *Front. Microbiol.* 9:2142. doi: 10.3389/fmicb.2018.02142

Wu, J., Hong, Y., Wen, X., Li, Y., Wang, Y., and Chang, X. (2020). Activity, abundance, and community composition of anaerobic ammonia–oxidizing (anammox) bacteria in sediment cores of the Pearl River Estuary. *Estuar. Coast.* 43, 73–85. doi: 10.1007/s12237-019-00668-1

Yang, X. R., Weng, B. S., Li, H., Marshall, C. W., Li, H., Chen, Y. S., et al. (2017). An overlooked nitrogen loss linked to anaerobic ammonium oxidation in estuarine sediments in China. *J. Soils Sediments* 17, 2537–2546. doi: 10.1007/s11368-017-1728-y

Yoshinaga, I., Amano, T., Yamagishi, T., Okada, K., Ueda, S., Sako, Y., et al. (2009). Distribution and diversity of anaerobic ammonium oxidation (anammox) bacteria in the sediment of a eutrophic freshwater lake, Lake Kitaura, Japan. *Microbes Environ.* 26, 189–197. doi: 10.1264/jsme2.me10184

Zhang, Y., Ruan, X. H., Op den Camp, H. J., Smits, T. J., Jetten, M. S., and Schmid, M. C. (2007). Diversity and abundance of aerobic and anaerobic ammonium-oxidizing bacteria in freshwater sediments of the Xinyi River (China). *Environ. Microbiol.* 9, 2375–2382. doi: 10.1111/j.1462-2920.2007.01357.x

Zhao, S., Zhuang, L., Wang, C., Li, Y., Wang, S., and Zhu, G. (2018). High-throughput analysis of anammox bacteria in wetland and dryland soils along the altitudinal gradient in Qinghai-Tibet Plateau. *Microbiologyopen* 7:e00556. doi: 10.1002/mbo3.556

Zhu, G., Wang, S., Wang, Y., Wang, C., Risgaard-Petersen, N., Jetten, M. S., et al. (2011). Anaerobic ammonia oxidation in a fertilized paddy soil. *ISME J.* 5, 1905–1912. doi: 10.1038/ismej.2011.63

Zhu, G., Xia, C., Shanyun, W., Zhou, L., Liu, L., and Zhao, S. (2015). Occurrence, activity and contribution of anammox in some freshwater extreme environments. *Environ. Microbiol. Rep.* 7, 961–969. doi: 10.1111/1758-2229.12341

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Lv, Narsing Rao, Ming, Salam, Li, Liu, Zhang, Zhang, Xian, Jiao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) 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: 28 January 2022
doi: 10.3389/fmicb.2021.739625

Characterization of a Novel Hot-Spring Cyanobacterium *Leptodesmis sichuanensis* sp. Nov. and Genomic Insights of Molecular Adaptations Into Its Habitat

*Jie Tang[1], Lian-Ming Du[1], Meijin Li[2], Dan Yao[1], Ying Jiang[2], Malgorzata Waleron[3], Krzysztof Waleron[4] and Maurycy Daroch[2]**

[1] *Antibiotics Research and Re-Evaluation Key Laboratory of Sichuan Province, Sichuan Industrial Institute of Antibiotics, Chengdu University, Chengdu, China,* [2] *School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China,* [3] *Laboratory of Plant Protection and Biotechnology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, University of Gdańsk, Gdańsk, Poland,* [4] *Department of Pharmaceutical Microbiology, Faculty of Pharmacy Medical University of Gdańsk, Gdańsk, Poland*

OPEN ACCESS

Edited by:
Brian P. Hedlund, University of Nevada, United States

Reviewed by:
Margarita Grabovich, Voronezh State University, Russia
Dhiraj Kumar Chaudhary, Korea University, South Korea

***Correspondence:**
Maurycy Daroch
m.daroch@pkusz.edu.cn

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

Received: *11 July 2021*
Accepted: *25 November 2021*
Published: *28 January 2022*

Citation:
Tang J, Du L-M, Li M, Yao D, Jiang Y, Waleron M, Waleron K and Daroch M (2022) Characterization of a Novel Hot-Spring Cyanobacterium Leptodesmis sichuanensis sp. Nov. and Genomic Insights of Molecular Adaptations Into Its Habitat. Front. Microbiol. 12:739625. doi: 10.3389/fmicb.2021.739625

The newly described genus *Leptodesmis* comprises several strains of filamentous cyanobacteria from diverse, primarily cold, habitats. Here, we sequenced the complete genome of a novel hot-spring strain, *Leptodesmis* sp. PKUAC-SCTA121 (hereafter A121), isolated from Erdaoqiao hot springs (pH 6.32, 40.8°C), China. The analyses of 16S rRNA/16S-23S ITS phylogenies, secondary structures, and morphology strongly support strain A121 as a new species within *Leptodesmis, Leptodesmis sichuanensis* sp. nov. Notably, strain A121 is the first thermophilic representative of genus *Leptodesmis* and more broadly the first *Leptodesmis* sp. to have its genome sequenced. In addition, results of genome-scale phylogenetic analysis and average nucleotide/amino acid identity as well as *in silico* DNA-DNA hybridization and patristic analysis verify the establishment of genus *Leptodesmis* previously cryptic to *Phormidesmis*. Comparative genomic analyses reveal that the *Leptodesmis* A121 and *Thermoleptolyngbya sichuanensis* A183 from the same hot-spring biome exhibit different genome structures but similar functional classifications of protein-coding genes. Although the core molecular components of photosynthesis, metabolism, and signal transduction were shared by the two strains, distinct genes associated with photosynthesis and signal transduction were identified, indicating that different strategies might be used by these strains to adapt to that specific niche. Furthermore, the complete genome of strain A121 provides the first insight into the genomic features of genus *Leptodesmis* and lays the foundation for future global ecogenomic and geogenomic studies.

Keywords: *Leptodesmis*, 16S rRNA, genome, comparative genomics, hot spring, thermophilic cyanobacterium, adaptation, Leptolyngbyaceae

INTRODUCTION

Thermophilic cyanobacteria are unique photoautotrophic microorganisms that can survive and thrive in adverse, thermal environments (Elleuche et al., 2015) with optimum growth and metabolic activities at high temperatures with the support of highly stable enzymes (Patel et al., 2019). Thus, they exhibit considerable biotechnological potential, e.g., CO_2 sequestration and production of biofuels, bioactive compounds, and pigments (Liang et al., 2019; Patel et al., 2019).

Despite the abundance of thermal ecosystems around the world, the diversity of thermophilic cyanobacteria is still highly underestimated. Next-generation sequencing (NGS) is extensively employed to investigate the cyanobacterial diversity in thermal environments in recent years (Tang et al., 2018b; Alcorta et al., 2020; Chen et al., 2021). However, NGS results often provide ample abstract data, e.g., operational taxonomic units (OTUs), higher taxonomic groups. Reports on novel isolates of thermophilic cyanobacteria are scarce. Isolating these microrganisms from different ecosystems is crucial for multidisciplinary studies and provides potential strains and genetic sequence data for biotechnology and industrial applications. Further, isolated strains provide opportunity for in-depth study studies to understand the various characteristics, such as morphology, genomics, and adaptation (Cordeiro et al., 2020; Tang et al., 2021).

Leptodesmis is a newly described genus based on environmental, morphological, and molecular data (Raabova et al., 2019). Two species in this genus have been studied, *L. alaskaensis* (Strunecky et al., 2020) and *L. paradoxa* (Raabova et al., 2019). In our previous study, we isolated a *Leptodesmis* strain, A121, from an Erdaoqiao hot spring, Sichuan, China (Tang et al., 2018a). Strain A121 can grow at 50°C and/or at the concentration of 0.1 M $NaHCO_3$ (Tang et al., 2018a).

To date, there has been no genome sequence available for any of the *Leptodesmis* strains. To fill this gap and reveal thermal adaptations, we sequenced the entire genome of A121 in this study. This strain was also characterized using molecular and morphological analysis, and further comparative genomics was conducted to reveal the genome structures and adaptation mechanisms of the *Leptodesmis* sp. A121 and *Thermoleptolyngbya sichuanensis* A183 from the same hot-spring biome. This study provides a foundation for future ecogenomic and geogenomic studies on global thermophilic cyanobacteria.

MATERIALS AND METHODS

Strain Information and Genome Sequencing

The strain A121 was originally isolated by Dr. Md. Mahfuzur R. Shah from Erdaoqiao hot springs (pH 6.32, 40.8°C) in Ganzi Prefecture of Sichuan Province, China, as previously described (Tang et al., 2018b). Unicyanobacterial culture of A121 was used to establish experimental cultures essentially as described previously (Tang et al., 2018a). The strain was initially denoted and deposited in the Peking University Algae Collection as PKUAC-SCTA121 and has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession number FACHB-3319. Cultivation of the strain and basic physiological studies closely follow previously published material (Tang et al., 2021). In short, the strain was grown in BG11 shake-flasks in at 45°C, 100 rpm, photoperiod 16 L: 8D (45 $\mu mol\ m^{-2}\ s^{-1}$).

The short reads of A121 were sequenced using Illumina NovaSeq PE150 at the Beijing Novogene Bioinformatics Technology Co., Ltd. Illumina PCR adapter reads and low-quality reads from the paired ends were filtered by the step of quality control using the company's own compiling pipeline. In total, 495,688,800 bp data were generated. For Oxford Nanopore Technologies (ONT) sequencing, sequencing libraries were generated using SQK-LSK109 Kit following the manufacturer's recommendations and protocol. The long reads were sequenced using a PromethION sequencer (Oxford Nanopore Technologies, Oxford, United Kingdom), and as a result, a total of 1,792,271,158 bp for A121 was generated. The raw data from Nanopore sequencing was converted from fast5 to fastq using Albacore software in the MinKNOW package. The generated circular contig was error corrected with a Geneious Prime (2020.2.2) mapper® to generate the final genome draft.

Gene prediction and annotation were performed using the NCBI prokaryotic genome annotation pipeline and further using the RAST annotation system to minimize poor calls. The circular plot of the A121 genome was generated in Circos v0.68 (Krzywinski et al., 2009). Default settings were used for all bioinformatics tools unless stated otherwise.

Phylogeny

Sequences of the 16S rRNA gene were collected for A121 by extraction from the genome sequence and for cyanobacterial references through BLAST search. Muscle implemented in Mega7 (Kumar et al., 2016) was employed to generate multiple sequence alignments, and manual editing was carried out when necessary. Maximum-likelihood (ML) inference of 16S rRNA sequences was reconstructed using PhyML v3.3 (Guindon et al., 2010). Parameter settings in PhyML were followed as described (Tang et al., 2019).

To infer the phylogenomic relationship, genomes of A121 and 13 focus taxa (nine representatives from family Leptolyngbyaceae and three from family Oculatellaceae as references and one from family Synechococcaceae as the outgroup) were included for analysis. The quality of the genomes was evaluated using CheckM (Parks et al., 2015) to ensure a high-quality data set with high completeness ($\geq$ 95%) and low contamination ($<$ 5%). Single-copy genes shared by all genomes were extracted from the homologous gene clusters identified by OrthoMCL (Li et al., 2003) and concatenated employing custom Perl script. Multisequence alignment was performed using MAFFT v7.453 (Standley, 2013). The supergene alignment was subjected to phylogenomic inference using IQ-TREE v2.1.3 (Minh et al., 2020). The optimal substitution model for phylogenomic analysis was selected from 546 protein models by ModelFinder implemented in IQ-TREE. Bootstrap tests (1,000 replicates) were

conducted for evaluation of tree topologies using UltraFast Bootstrap (Hoang et al., 2018).

Multi Locus Sequence Analysis encompassing 14 loci (11,869 bp): *cpc*A (465 bp), *cpc*B (486 bp), *cpc*H (204 bp), *gyr*A (597 bp), *kai*A (408 bp), *kai*B (303 bp), *kai*C (1,475 bp), *nir* (1,457 bp), *rbc*C (318 bp), *rbc*L (1,241 bp), *rec*A (672 bp), *rec*N (1,513 bp), *rpo*C (689 bp), *sec*A (2,039 bp) was performed using MEGA X software[1] using the ML method and the GTR + G + I with Gamma distribution as the best nucleotide substitution model. The bootstrap consensus was inferred from 1,000 replicates. Phylogenetic analysis was carried out using 34 strains on the basis of sequences retrieved from genomes of *Leptodesmis* and closely related Leptolyngbyaceae strains that genomic or good quality metagenomic data were available at GenBank (in June 2021).

Analysis of 16S-23S ITS

The 16S-23S ITS region was extracted from the A121 genome, and cyanobacterial sequences were retrieved from GenBank through BLAST search as references. Phylogenetic analysis of 16S-23S ITS was performed using the aforementioned pipeline of 16S rRNA analysis. For the strains ascribable to *Leptodesmis*, the conserved domains of the 16S-23S ITS region: D1-D1′, D2, D3, boxA, and D4; and its variable regions (V2, boxB, and V3) were identified as previously described (Iteman et al., 2000). The tRNAs presented in the spacer were identified by tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997). The secondary structures of the identified fragments were individually determined by Mfold web server (Zuker, 2003). Except for the use of the structure draw mode untangle with loop fix, default conditions in Mfold were used in all cases.

Morphology Investigation

Strain A121 was inspected at 400 × magnification using light microscopy (LM, DP72, OLYMPUS, Japan) equipped with an image acquisition system (U-TV0.63XC, OLYMPUS, Japan). Microscopic investigations were also conducted using scanning electron microscopy (SEM) (SU8100, HITACHI, Japan) and transmission electron microscopy (TEM) (HT7800, HITACHI, Japan). All microscopic operation was performed following the procedures as described by Tang et al. (2021).

Genome Analysis

Pairwise genome comparisons were performed to calculate average nucleotide identity (ANI) and average amino acid identity (AAI) using the ANI/AAI calculator (Rodriguez-R and Konstantinidis, 2016). Further, clustering analysis was conducted for genomes based on AAI distances using the unweighted pair group method with algorithmic mean (UPGMA).

In silico DDH comparison was performed employing the genome-to-genome distance calculator (GGDC 2.0)[2] (Tang et al., 2021) using the recommended BLAST + alignment and formula 2 (identities/HSP length) (Johansen et al., 2011). Patristic distances between and within the *Leptodesmis* sp. A121 and other Leptolyngbyaceae strains were calculated with PATRISTICv1.0 software (Zhang et al., 2019) using concatenated sequences of 14 loci: *cpc*B, *cpc*A, *cpc*H, *gyr*A, *kai*A, *kai*B, *kai*C, *nir*, *rbc*C, *rbc*L, *rec*A, *rec*N, *rpo*C, *sec*A.

Comparative genomics was performed between *Leptodesmis* sp. A121 (CP075171) and *T. sichuanensis* A183 (CP053661). The two strains originated from the same sampling site (Tang et al., 2018a). We speculated that similar molecular adaptations of the two strains were shaped or evolutionarily driven by the thermal ecosystem. The homologous genes between genomes were identified by OrthoMCL (Li et al., 2003). The customized Venn diagram was drawn using an online tool to exhibit the orthologous and unique genes between the strains. The EggNOG database v4 (Jaime et al., 2016) was used for functional classification of protein sequences using the following thresholds: *E*-value cutoff of 1E-6, ≥ 30% identity and 70% coverage. Pairwise genome alignments were performed using MUMmer v3.2.3 (Kurtz et al., 2004).

RESULTS AND DISCUSSION

Complete Genome Sequence of *Leptodesmis* sp. A121

Whole-genome sequencing of *Leptodesmis* sp. A121 generated a total of 3,304,592 filtered paired-end reads (clean data), providing approximately 93-fold coverage of the genome. Bioinformatic assembly obtained a single circular chromosome with a size of 5,348,817 bp (GC content, 50.32%) and no plasmid. The annotation of the A121 genome indicated two ribosomal RNA operons, 45 tRNA genes, and 5345 protein-coding genes (**Figure 1** and **Supplementary Table 1**). Approximately 52.6% of protein-coding genes were identified as hypothetical proteins (**Supplementary Table 2**). No strong patterns of GC-skew were observed in the chromosome organization of A121 (**Figure 1**). This result was similar to other hot-spring cyanobacteria, e.g., *Thermosynechococcus elongatus* BP-1 (Yasukazu et al., 2002) or *T. sichuanensis* A183 (Tang et al., 2021).

Phylogenetic Reconstruction of 16S rRNA Gene

The ML phylogenetic tree (**Figure 2** and **Supplementary Figure 1**) inferred by 16S rRNA gene sequences categorized the cyanobacterial strains into 17 well-supported genera within the family Leptolyngbyaceae and two genera within family Oculatellaceae and *Gloeobacter* as an outgroup. In addition, the tree indicates that five *Leptolyngbya* strains, strain Greenland 10, E412, E811, A2 and LBK, were quite divergent from the genus *Leptolyngbya sensu stricto* and could be classified into three new genera. The actual taxonomic delineation of these *Leptolyngbya* strains must be carefully investigated using polyphasic approaches (e.g., more molecular markers, morphology, etc.) in a separate study because genus *Leptolyngbya* has been recognized as polyphyletic (Johansen et al., 2011).

Strain A121 clustered with six strains, forming a well-defined clade of genus *Leptodesmis*. Of the six strains, *L. alaskaensis* V20 (Strunecky et al., 2020) and *L. paradoxa*

[1] www.megasoftware.net

[2] http://ggdc.dsmz.de/distcalc2.php

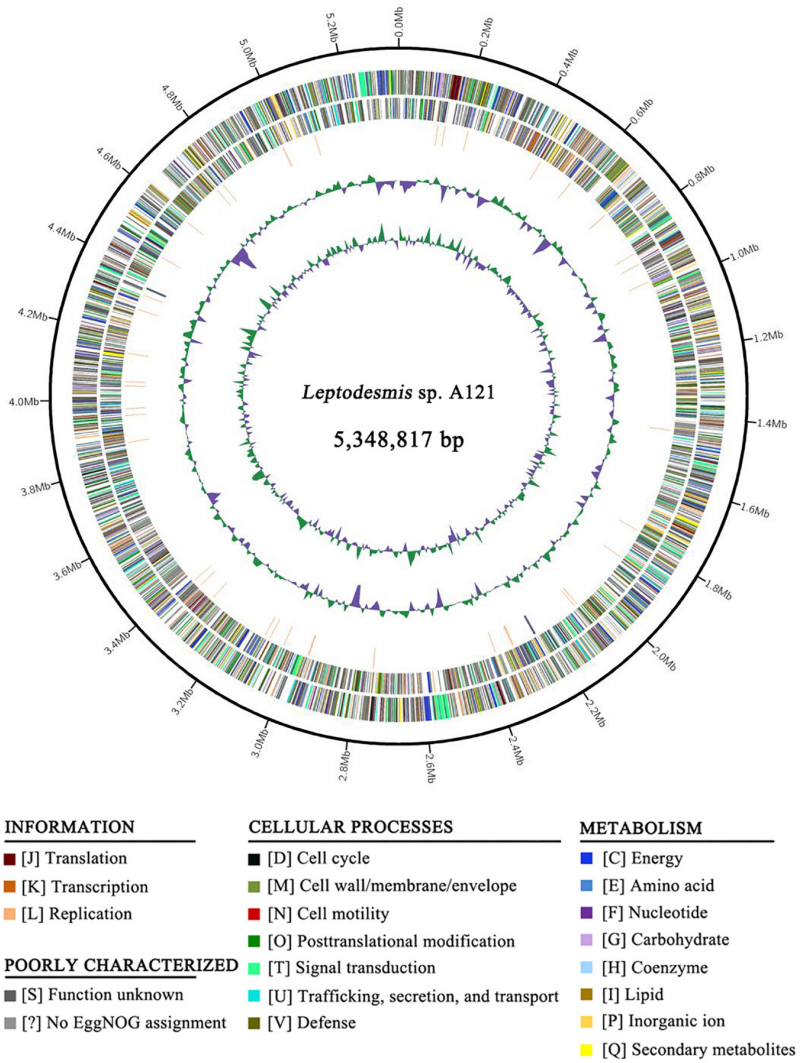


FIGURE 1 | Genome illustration of *Leptodesmis* sp. A121. Rings are as follows (outer–inner): CDS on plus strand; CDS on minus strand; rRNA (orange) and tRNA (blue); the last two circles represent GC content and GC skew both calculated for a 10 kb window with 1 kb stepping. The CDS are color-coded by functional categories of EggNOG database.

LK021 (Raabova et al., 2019) are two recently proposed species; *Leptolyngbya* sp. Greenland 9 (Roeselers et al., 2007) and CCNUM2 (Zhang et al., 2019) were probably misclassified as indicated by the clear conflict between phylogeny and taxonomy; strain EY05-AM1 and EY05-AM2 were taxonomically unassigned and just labeled as Leptolyngbyaceae cyanobacterium. The 16S rRNA sequence identities between A121 and the six strains ranged from 94.9 to 96.9% (**Table 1**), suggesting A121 as a novel species within *Leptodesmis* based on the recommended threshold of 16S rRNA gene identity for bacterial species (98–99%) or genera (94.5–95%) demarcation (Rodriguez-R et al., 2018). Strain Greenland 9 and CCNUM2 can be proposed to be another new species within *Leptodesmis* based on the 16S rRNA phylogeny and identity (**Figure 2** and **Table 1**), but the affiliation of strain EY05-AM1 and EY05-AM2 to the same species as Greenland 9 and CCNUM2 was not strongly supported in light of the 16S rRNA identity values (97.6–98.3%). Besides this, the shorter length of 16S rRNA sequence of strain EY05-AM1 and EY05-AM2 (**Table 1**) might affect the identity value and phylogenetic positions on the tree of the current study. Therefore, the actual taxonomy of strain EY05-AM1 and EY05-AM2 should be thoroughly studied in the future.

Intriguingly, strain A121 (Tang et al., 2018a) and Greenland 9 (Roeselers et al., 2007) were originally isolated from hot springs in sharp contrast to habitat niches of strain V20 (Strunecky et al., 2020) and LK021 (Raabova et al., 2019) recovered from polar regions. The distinct habitat niches suggest that these strains might be different ecotypes. Strain CCNUM2 appeared to be an intermediate ecotype between thermophiles and psychrophiles because it was isolated from humid moss on forest limestone where temperature ranges from 29 to 39°C in the summer

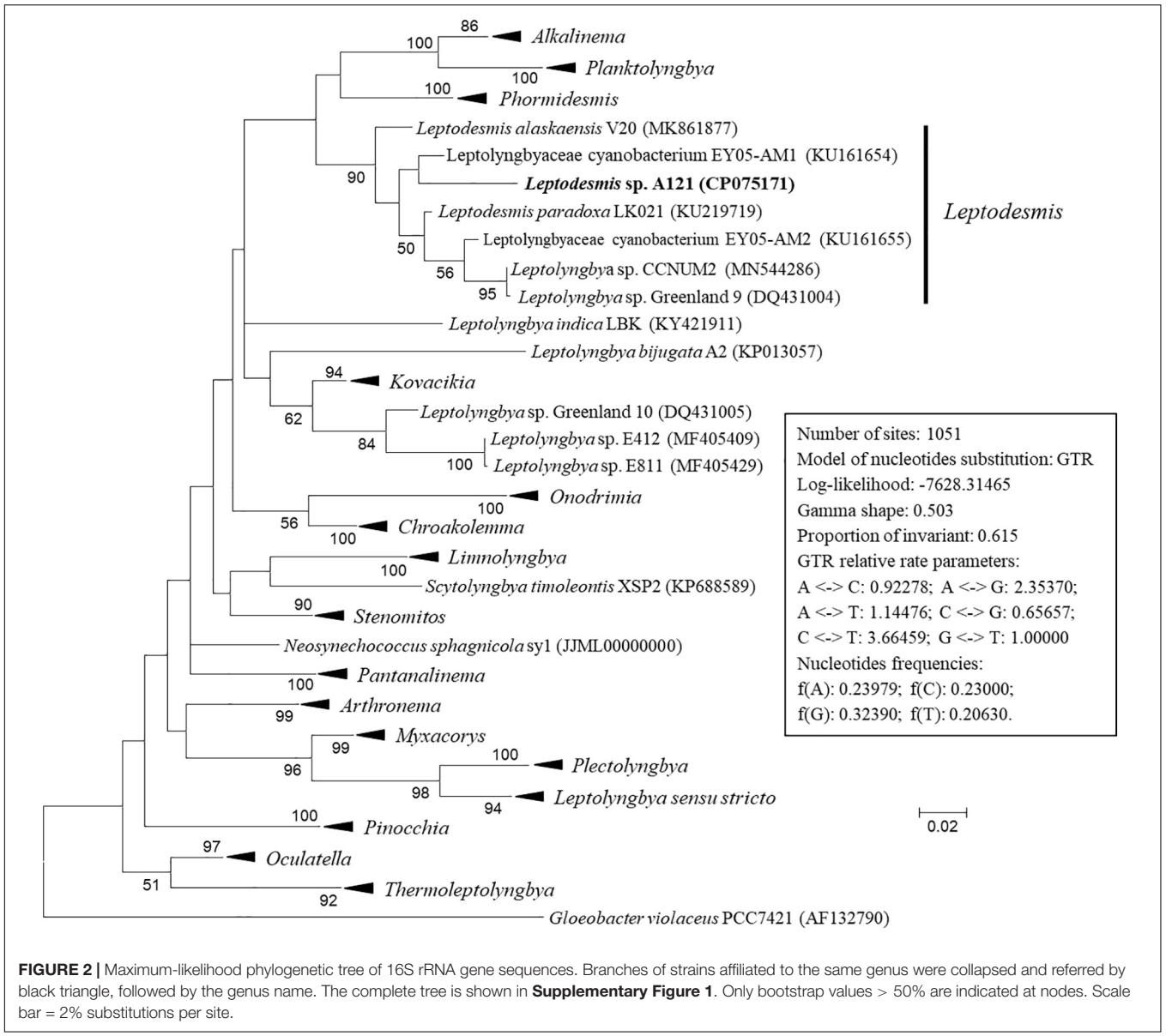


FIGURE 2 | Maximum-likelihood phylogenetic tree of 16S rRNA gene sequences. Branches of strains affiliated to the same genus were collapsed and referred by black triangle, followed by the genus name. The complete tree is shown in **Supplementary Figure 1**. Only bootstrap values > 50% are indicated at nodes. Scale bar = 2% substitutions per site.

to 1–8°C in the winter (Zhang et al., 2019). Strain EY05-AM1 and EY05-AM2 originated from a wet waterfall wall in El Yunque national forest, Puerto Rico. The acclimation of *Leptodesmis* strains to diverse habitats strongly implies the underlying genetic diversity within this genus. Verifying this speculation will be an interesting topic using phylogenomic and ecogenomic approaches with the help of more genome sequences of this genus. Unfortunately, to date there is only one genome sequence available for genus *Leptodesmis*, that presented in this study, strain A121.

Phylogeny and Secondary Structures of 16S-23S ITS

In addition to the 16S rRNA gene, the 16S-23S ITS region is another commonly used molecular marker for delineating cyanobacterial ecotypes or species (Brito et al., 2017). In light of sequence unavailability, the sequence data set of 16S-23S ITS is smaller than that of 16S rRNA. Although the phylogenetic inference of 16S-23S ITS (**Figure 3** and **Supplementary Figure 2**) show a distinct topology from that of the 16S rRNA gene, the classification of species was consistent at the genus level. The three strains ascribable to *Leptodesmis* were placed into a well-separated clade but showed evident genetic divergence as indicated by branch length (**Figure 3**), suggesting that the three strains are different species within *Leptodesmis*. However, 16S-23S ITS sequences are extraordinarily divergent and probably bring about misleading taxonomic assignment without secondary structure domain analysis (Johansen et al., 2011). Therefore, secondary structure analyses of 16S-23S ITS were also carried out as an essential complement to further resolve the taxonomy within *Leptodesmis*.

TABLE 1 | Sequence identities (%) of 16S rRNA gene of *Leptodesmis* strains studied.

Strain	A121	V20	LK021	CCNUM2	Greenland 9	EY05-AM1	EY05-AM2
Leptodesmis sp. A121	100 (1,453)						
Leptodesmis alaskaensis V20	96.1 (1,147)	100 (1,145)					
Leptodesmis paradoxa LK021	94.9 (1,148)	95.6 (1,149)	100 (1,140)				
Leptolyngbya sp. CCNUM2	96.5 (1,420)	96.0 (1,115)	96.5 (1,116)	100 (1,419)			
Leptolyngbya sp. Greenland 9	96.6 (1,425)	96.1 (1,120)	96.6 (1,101)	99.9 (1,419)	100 (1,425)		
Leptolyngbyaceae cyanobacterium EY05-AM1	96.9 (1,147)	96.8 (1,149)	96.4 (1,149)	97.6 (1,114)	97.7 (1,120)	100 (1,147)	
Leptolyngbyaceae cyanobacterium EY05-AM2	95.8 (1,110)	96.7 (1,110)	96.4 (1,110)	98.3 (1,110)	98.2 (1,110)	97.4 (1,110)	100 (1,110)

The alignment length (bp) is indicated inside brackets.

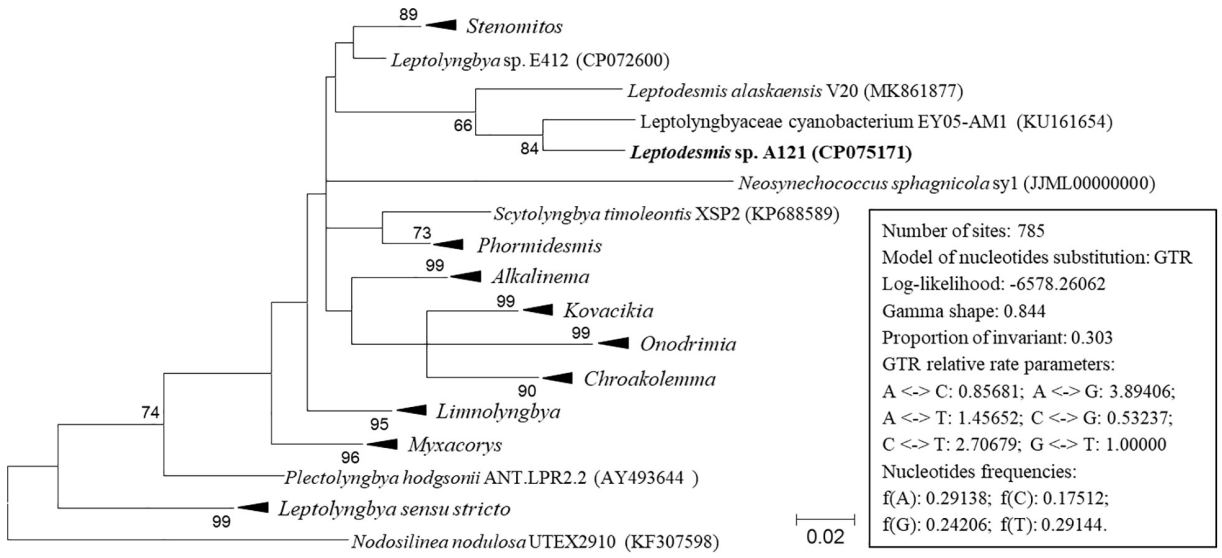


FIGURE 3 | ML phylogenetic tree of 16S-23S ITS sequences. Branches of strains affiliated to the same genus are collapsed and referred by black triangle, followed by the genus name. The complete tree is shown in **Supplementary Figure 2**. Only bootstrap values > 50% are indicated at nodes. Scale bar = 2% substitutions per site.

TABLE 2 | The length (bp) summary of regions within16S-23S ITS of *Leptodesmis* strains studied.

Strain	ITS length (tRNA removed)	D1-D1′ helix	D2	D3	V2 helix	Boxb helix	Boxa	D4	V3 helix
Leptodesmis sp. A121	325	63	12	5	81	33	12	7	98
Leptolyngbyaceae cyanobacterium EY05-AM1	236	62	12	5	4	33	12	7	44
Leptodesmis alaskaensis V20	NA	77	12	5	4	NA	NA	NA	NA

NA, not available.

Unfortunately, only strain A121 and EY05-AM1 had available full-length ITS sequences, whereas strain V20 had partial sequence (**Table 2**). Excluding two highly conserved tRNAs from ITS sequences, the length of the remaining ITS sequences were 325 and 236 bp for strain A121 and EY05-AM1, respectively (**Table 2**). Among the three *Leptodesmis* strains, identical sequences were observed in conserved domains D3 (GGTTT), whereas conserved domain D2 exhibited three sequence types with one-nucleotide difference, CTTTCAAACTA in A121, CTTTCAAACTT in V20 and CTTTCAAGCTA in EY05-AM1. Strain A121 and EY05-AM1 both showed identical sequences of conserved domain boxA (GGACCTTGAAAA) and D4 (CTGCATA).

The inferred D1-D1′ helices (**Figure 4A**) of the three *Leptodesmis* strains showed different lengths (62–77 bp) and distinct structures to each other. The V2 helix of strain A121 was 81 residues long (**Figure 4B**), whereas that of strain EY05-AM1 and V20 only contained four residues, resulting in no secondary structure predicted. Strain A121 and EY05-AM1 shared a basal stem structure (GUC-GAC) of V3 helices, but the considerable difference in length of V3 helices generated distinct structures (**Figure 4C**). Although strain A121 showed the same residue length (33 bp) of boxB helix to EY05-AM1, the structures of boxB helix (**Figure 4D**) were clearly incompatible with each other.

In summary, the phylogeny and secondary structures undoubtedly differentiate A121 from the other *Leptodesmis*

TABLE 3 | Comparison of morphological features of *Leptodesmis* strains.

Strain	Isolation source	Filaments	Motility	Cell width (μm)	Cell length (μm)	Sheaths	Number of thylakoids	Color of trichomes	References
Leptodesmis sp. A121	Green mat, hot spring, Sichuan, China	Straight, curved, flexuous or wavy, solitary, one filament per sheath dark green	Motile	0.9–1.4	2.0–3.0	Thick, colorless, Multilayered	3–5	Blue-green	This study
Leptodesmis alaskaensis V20	Green mat in dry channels, Lake NE2 in Toolik lake area, Alaska, United States	Straight, curved, flexuous or wavy, solitary, one filament per sheath pale-blue green to dark green	Motile	1–1.5	1.4–4.3	Thick, colorless	NA	Pale blue-green to green	Strunecky et al., 2020
Leptodesmis paradoxa LK021	Mud in pool, Deception Island, Antarctica	Straight, curved, flexuous or wavy, solitary, one filament per sheath pale blue-green to dark green	None	2.5–3.5	1.0–1.5	Thick, colorless	NA	Pale blue green	Raabova et al., 2019
Leptolyngbya sp. CCNUM2	Humid moss on forest limestone, Wuhan, China	Non-branched	NA	1.57–1.71	1.38–1.96	NA	NA	NA	Zhang et al., 2019

NA, not available.

strains, confirming the delineation of the A121 strain as a new species of *Leptodesmis*. The secondary structure analysis of the variable regions (D1-D1′, V2, boxB, and V3 helices) proved to be an effective approach for species-level identification within *Leptodesmis*. Although the V2 helix was the most inconsistent (**Table 2**), it was the least taxonomic informative in light of its extreme variability and absence in some cyanobacterial strains (Sciuto and Moro, 2016). In addition, more sequences are essential for accurately resolving the other *Leptodesmis* strains included in the 16S rRNA tree.

Morphological Features of *Leptodesmis* sp. A121

Light microscopy (**Figure 5A**) indicated straight, wavy, curved, and occasionally bent trichomes of strain A121. The SEM indicated long, thin, unbranched filaments, whereas the TEM (**Figures 5B,C**) showed that some shorter trichomes of A121 are comprised of 10–25 individual barrel-shaped cells. The trichomes observed in the longitudinal direction were solitary filaments composed of long cylindrical cells with contraction in the transverse wall (**Figure 5B**). In addition, TEM revealed the ultrastructure of A121, a single cell with a length of 2.0–3.0 μm and a width of 0.9–1.4 μm (**Figures 5C,D**). The long cylindrical cells were separated by the centripetal invagination of the cell wall, and no intracellular connection between the vegetative cells was observed (**Figure 5D**). The TEM analysis also showed that the three to five thylakoid layers were located in parallel at the inner periphery of the cells (**Figures 5C,D**). Sheaths, septum, carboxysomes, phycocyanin granules, and polyphosphate bodies were present in the cytoplasm (**Figures 5C,D**). The trichomes were surrounded by a clearly visible layer of extracellular polymeric substances (EPS).

Comparisons of morphological features of known *Leptodesmis* strains are summarized in **Table 3**. In general, the three *Leptodesmis* strains were blue-green and were straight, curved, flexuous, or wavy, solitary filaments with one sheath of one cell. Dissimilarities, however, were also observed. Strains A121 and *L. alaskaensis* were motile, whereas *L. paradoxa* was immotile. The cell width of *L. paradoxa* was considerably wider than that of A121 and *L. alaskaensis*. A broader range of cell lengths was noticed in *L. alaskaensis*, whereas A121 exhibited a medium cell length. Meanwhile, the CCNUM2 strain showed moderate cell widths and lengths compared with other representatives of *Leptodesmis*.

Phylogenomic and Genomic Distance Analysis

To further resolve relationships between *Leptodesmis* strains and related taxa, phylogenomic and genome distances were compared. Analysis of homologous gene clusters indicated 845 single-copy genes present in the 13 strains studied. The concatenated alignments of these genes exhibited 266,815 aligned amino acid sites. ML inference of the supergene alignment generated a phylogeny with strong bootstrap support (100%) for all branches (**Figure 6**), clearly defining each genus. This phylogenomic topology was almost consistent with that of the 16S rRNA gene (**Figure 2**). There was only one exception. The phylogenetic analysis of the 16S rRNA gene clearly indicated the affiliation of *Neosynechococcus sphagnicola* sy1 to the family Leptolyngbyaceae, whereas sy1 was placed in an ambiguous position from the genome-scale phylogeny. Further studies are required to ascertain the actual taxonomy of this strain using, e.g., polyphasic approaches. Multilocus sequence analysis (MLSA) of 14 loci from 34 strains (**Supplementary Figure 3**) further support the findings of the previous phylogenomic and 16S rRNA gene

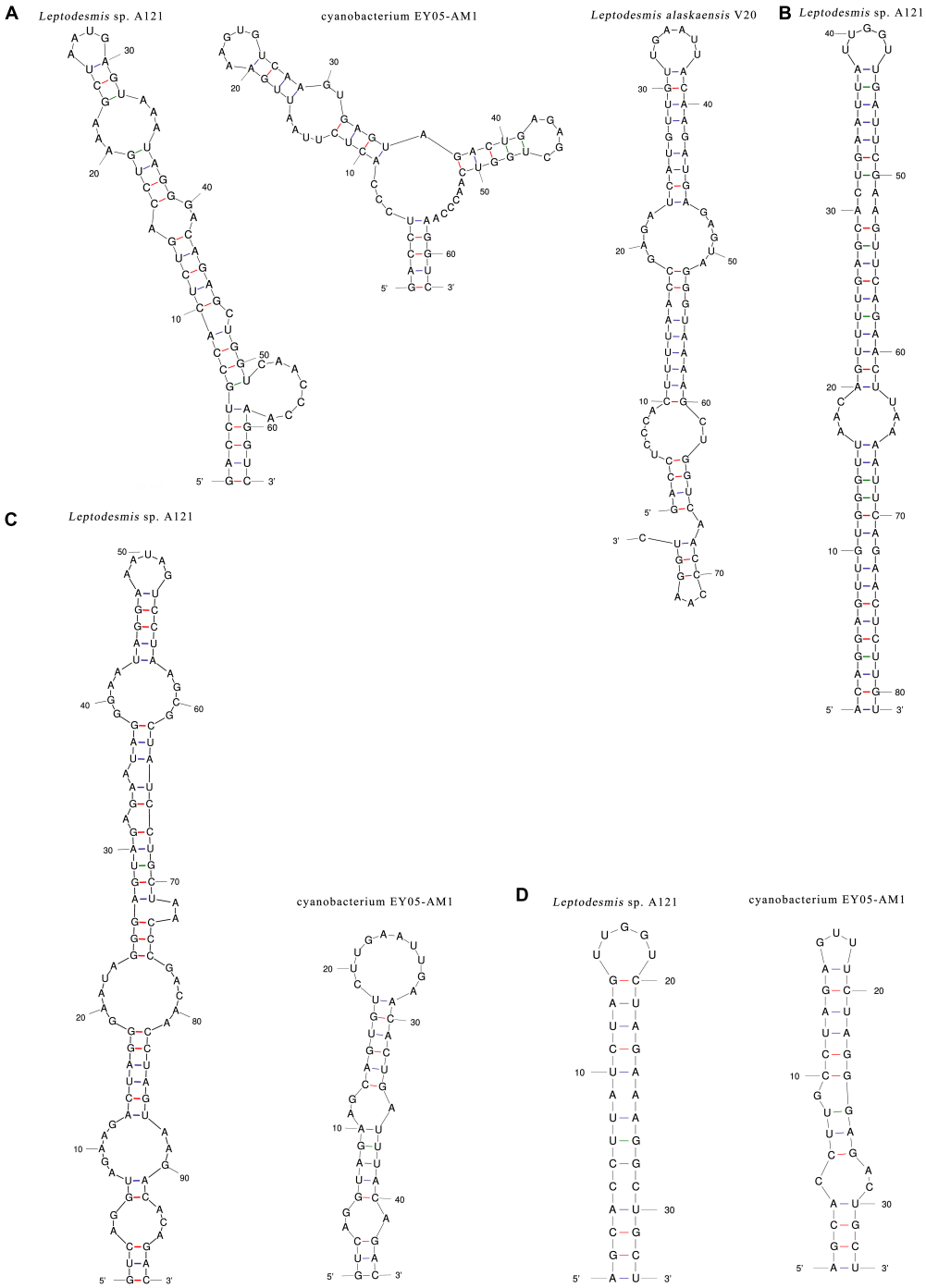


FIGURE 4 | Hypothetical secondary structures of D1-D1′ helix **(A)**, V2 helix **(B)**, V3 helix **(C),** and boxB **(D)** of 16S-23S ITS of *Leptodesmis* strains.

analyses. To summarize, the phylogenomic trees supported the new delineation of *Leptodesmis*, which was a cryptic genus of *Phormidesmis* (Raabova et al., 2019).

Results of genome-wide ANI and AAI (**Supplementary Figure 4**) all conformed to the suggested values for species (ANI > 96%, AAI ≥ 95%) and genus (ANI < 83%, AAI ≤ 70%) delimitation (Walter et al., 2017; Jain and Rodriguez, 2018), respectively. The UPGMA clustering based on genome-scale AAI distances (**Supplementary Figure 4**) was similar to the molecular phylogenies. To shed more light on global sequence similarity between *Leptodesmis* sp. A121 and different Leptolyngbyaceae, combined *in silico* DNA-DNA hybridization (**Supplementary Table 3**) and patristic distance (**Supplementary Table 4**) analyses were performed. The

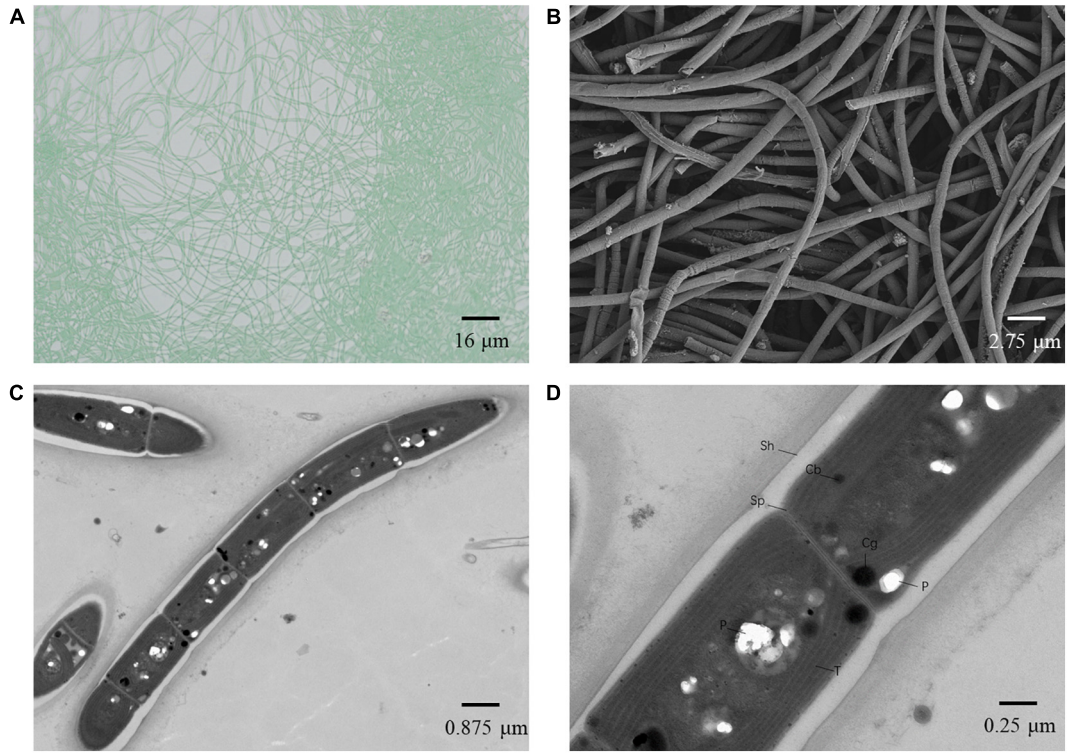


FIGURE 5 | Microscopic images of *Leptodesmis* sp. A121. **(A)** light microscopy image. **(B)** SEM image. **(C,D)** TEM images. Cb, carboxysome; Cg, cyanophycin granule; P, polyphosphate body; Sh, sheath; Sp, septum; T, thylakoid membrane. Magnifications were 400× **(A)**, 5,000× **(B)**, 8,000× **(C),** and 12,000× **(D)**.

estimated *is*DDH values dropped below 32% when the genome of *Leptodesmis* sp. A121 was compared with other Leptolyngbyaceae strains that were used as queries (**Supplementary Table 3**). When patristic distances were compared between *Leptodesmis* sp. A121 and other Leptolyngbyaceae strains, significantly higher values of 0.24–0.35 were noticed (**Supplementary Table 4**). Thus, the results indicate that there are no other representatives described until now, which might belong or be assigned to *Leptodesmis* based on genomic data deposited in the public databases. These findings combined further reinforce that *Leptodesmis* was correctly delineated as a separate genus from *Phormidesmis*. Meanwhile, within the *Leptodesmis* genus, the strain A121 could be delineated as a separate species based on currently available sequence information.

Genome Comparison

To elucidate the genomic features of *Leptodesmis* sp. A121 and *T. sichuanensis* A183 isolated from the same Erdaoqiao hot spring, a comparative analysis was performed. Genome characteristics and related information for the two hot-spring strains were shown in **Supplementary Table 1**. Venn diagram (**Supplementary Figure 5**) indicated that 3056 genes were shared by the two genomes, and 2177 genes were present only in the *Leptodesmis* sp. A121 genome. Functional classification showed that the distribution patterns of EggNOG categories (**Supplementary Figure 6**) were similar between the two genomes. Pairwise genome alignments (**Supplementary Figure 7**) suggest considerably low conservation in chromosomal organization between the two genomes. These results indicate that the two strains from the same ecosystem showed different genome structures but similar functional classifications of protein-coding genes.

Photosynthesis

The core sets of genes coding for both photosystem I and photosystem II were conserved in genomes of the two hot-spring strains. However, these genes also exhibit variations in distribution and copies. For example, two copies of cytochrome c-550 genes (*psbV1* and *psbV2*) were tandemly clustered in the *Leptodesmis* genome while separately arranged in the *Thermoleptolyngbya* genome. The two genomes contained *psbA* genes (*psbA1*, *psbA2*, and *psbA3*) encoding the reaction center D1 protein of photosystem II, but the copies of *psbA* genes were three for the *Thermoleptolyngbya* genome and four for the *Leptodesmis* genome. For the auxiliary components in the light energy transmission system of photosynthesis, homologous genes encoding allophycocyanin were present in the *Leptodesmis* genome, while there were allophycocyanin and phycocyanin in the *Thermoleptolyngbya* genome. The different types of phycobilisome proteins between the two hot-spring strains indicate that an alternative light-absorbing strategy was utilized for acclimation to the same environments, possibly enabling a degree of their stratification within the biofilm.

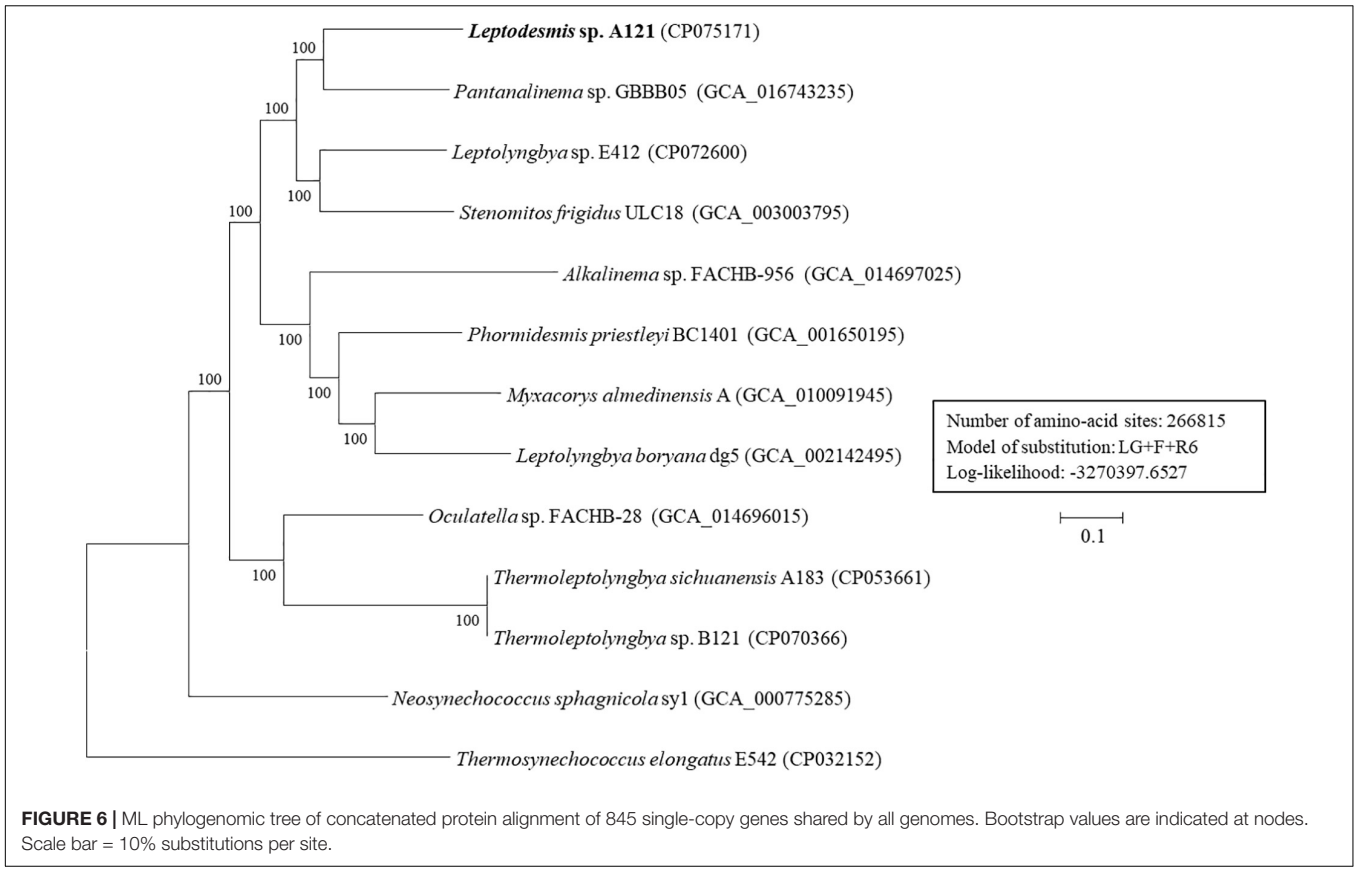


FIGURE 6 | ML phylogenomic tree of concatenated protein alignment of 845 single-copy genes shared by all genomes. Bootstrap values are indicated at nodes. Scale bar = 10% substitutions per site.

Essential genes required for the carbon-dioxide concentrating mechanism (CCM) were also conserved between the two hot-spring strains. However, some variations were found in the components of CCM-related genes. The arrangements of *ndhA-M* genes encoding the NADPH dehydrogenase (NDH-1) complexes responsible for uptake of gaseous CO_2 systems (Ogawa and Mi, 2007) were completely distinct between genomes. Eleven copies of *Hat/HatR* gene encoding high-affinity carbon uptake protein were present in the *Leptodesmis* genome and 13 in the *Thermoleptolyngbya* genome. For the bicarbonate transport system, *bicA1*, a low-affinity Na^+-dependent bicarbonate transporter gene was present in the *Leptodesmis* genome, whereas both *bicA1* and *bicA2* were present in the *Thermoleptolyngbya* genome. The *cmpABCD* operon encoding high-affinity ABC-type bicarbonate transport system was present in both genomes. In contrast, high-affinity Na^+-dependent bicarbonate transport family permease genes, *sbtA* and *sbtB*, were present only in the *Thermoleptolyngbya* genome. The different systems of carbon assimilation in the two strains might be flexibly activated to cater to the varying demands of carbon uptake in light of an economy of regulation from a survival aspect. Therefore, the increased genetic repertoire of bicarbonate transporters can indicate A183 strain's preference for this source of inorganic carbon vs. A183 (Tang et al., 2018a).

Interestingly, the flavodoxin gene (*fldA*) was not found in the *Leptodesmis* genome but present the *Thermoleptolyngbya* genome. Under iron-deficient conditions, flavodoxins can substitute ferredoxin to function as an electron transfer, and the iron stress-inducible proteins (*isiA*) increase the light-absorbing efficiency of PSI in the form of PSI-*isiA*-flavodoxin supercomplex (Cao et al., 2020). Thus, this kind of energy and electron transfer is not the prevailing acclimation and regulation mechanism for the *Leptodesmis* strain to overcome usually iron-deficient natural environments. In addition, the *isiA* gene in the *Leptodesmis* genome showed a significant degree of similarity to that of mesophilic *Synechocystis* sp. PCC6803 (37.3% amino acid sequence identity). Further study is required to elucidate the physiological significance of structural differences.

Metabolism

The ability to utilize various organic and inorganic sources of substances is vital for hot-spring strains to survive in oligotrophic aquatic environments (Esteves-Ferreira et al., 2018). The core metabolism genes were shared by genomes of the two hot-spring strains. The strains are likely able to import and utilize nitrate and urea as nitrogen sources, in light of hosting complete gene sets of *nrtABCD* (nitrate assimilation), *urtABCDE* (ABC-type urea transport system) and *ureA*, *-BC*, *-EF*, *-G* (urease genes). The *Leptodesmis* genome contained genes encoding for nitrogenases (*nif*) similar to those in the *Thermoleptolyngbya* genome, suggesting that the *Leptodesmis* strain is a nitrogen-fixing cyanobacteria. The ability of nitrogen fixation was experimentally confirmed by nitrogenase

activity (**Supplementary Figure 8**). Moreover, the two genomes possessed homologs associated with ammonium transporter, glutamine synthetase and glutamine amidotransferase, and nitrogen assimilation transcriptional activator, which played roles in ammonium ion assimilation (Bolay et al., 2018) and nitrogen control for efficient utilization of intracellular resources responding to ambient low nitrogen (Vega-Palas et al., 2010), respectively.

Hot-spring strains are usually capable of sulfur assimilation as many geothermal systems possess abundant sulfur. The ABC-type sulfate transport system (the sulfate-thiosulfate permease, *sulT*) was present in the two genomes. However, the distribution of these coding genes varied significantly. In the *Leptodesmis* genome, two of the four genes encoding the subunits of *sulT* formed a *cysTW* operon, and the other two genes (*cysA*, *cysP*) were individually arranged away from the operon. By contrast, the *Thermoleptolyngbya* genome comprised a *cysPTW* operon and *cysA* located far away from the operon. In addition, the *Leptodesmis* genome harbored three distinct copies of sulfate permease encoded by *sulP*, whereas the *Thermoleptolyngbya* genome contained one. The high-affinity ModABC molybdate transport system encoded by the *modABC* operon was present in the two genomes and can be also used for sulfur transport.

Two distinct sets of gene clusters for the phosphate-specific transport system (encoded by *pstSCAB* operon) were detected in the *Leptodesmis*, whereas the *Thermoleptolyngbya* genome hosted only one operon. One *pstSCAB* operon of the *Leptodesmis* genome shared high amino acid sequence similarity with that of the *Thermoleptolyngbya* genome. The other one was quite distinct, indicating that this operon might be obtained from different origins through horizontal gene transfer. An arsenical resistance operon (glyceraldehyde-3-phosphate dehydrogenase, MFS-type efflux pump *arsJ*, arsenical-resistance protein ACR3) with an upstream repressor was present in the two genomes. This result indicated that the gene operon may confer regulable arsenate resistance to the two strains (Chen et al., 2016). Overall, the *Leptodesmis* and *Thermoleptolyngbya* strains appeared to have similar metabolism systems to acclimate to the shifting conditions in hot-spring environments.

Signal Transduction

The perception of environmental stress and the subsequent transduction of stress signals are conducted through the two-component regulatory systems in cyanobacteria (Wuichet and Zhulin, 2010). In the *Leptodesmis* genome, 32 and 42 genes were identified as potential genes encoding histidine kinases and response regulators, respectively. However, the association between histidine kinases and response regulators cannot directly be elucidated due to the scattered distribution of these genes in the genome. This is in sharp contrast to several cases in prokaryotes with which a single two-component system was tandemly arranged into operons or closely clustered (Wuichet et al., 2010; Aguilar et al., 2014). Most of the histidine kinases and response regulators were conserved between the two genomes, but variations also existed. For example, the *Leptodesmis* genome only had a complete set of motility-related two-component signaling systems, including *cheA*, methyl-accepting chemotaxis protein, *cheW* and *cheY*, whereas the *Thermoleptolyngbya* genome had two sets. In addition, the GGDEF/EAL domain proteins in the genomes varied in gene numbers from 7 to 17, suggesting the complexity of cyclic-di-GMP signaling pathways might be different between the two strains.

Other Features in the *Leptodesmis* Genome

The Ni-Fe bidirectional hydrogenase enzyme (encoded by *hoxEFUYH*) was present in the *Leptodesmis* genome but absent in the *Thermoleptolyngbya* genome. The *hoxW* gene, involved in subunit assembly of *hoxH*, was located downstream the *hoxEFUYH* around 5.5 kb in the *Leptodesmis* genome. In addition, the two genomes hosted polyketide synthase (PKS) module-related proteins, a family of multidomain enzymes that produce polyketides (a large class of secondary metabolites) (Maurya et al., 2019). The genes encoding acyl-[acyl-carrier-protein] (ACP) reductase and aldehyde decarbonylase were tandemly arranged in the *Leptodesmis* genome but scattered in the *Thermoleptolyngbya* genome. These genes are involved in the biosynthesis of pentadecane and heptadecane alkanes that are the major constituents of various fuels, such as diesel, jet fuel, and gasoline (Peralta-Yahya et al., 2012). The *Leptodesmis* genome comprised genes encoding Rpn family recombination-promoting nuclease/putative transposase that was absent in the *Thermoleptolyngbya* genomes. The Rpn family can promote the acquisition process of obtaining new genes from other bacteria to adapt to ecological niches and survive under stressful conditions (Kingston and Ponkratz, 2017). Therefore, this result indicated that hot-spring strains may conduct gene regulation to accelerate the process of horizontal gene transfer. Furthermore, only the *Leptodesmis* genome contained VapC family toxin that belongs to type II toxin-antitoxin system and induces RNA cleavage (McKenzie et al., 2012). Further study may be required to elucidate the significance of these distinct genes.

CONCLUSION

In this manuscript, the polyphasic approach allowed us to propose a new species, *Leptodesmis sichuanensis*, and the delineation of strain A121 as representative of this new taxon. This proposal was strongly supported by 16S rRNA/16S-23S ITS phylogenies, secondary structures, and morphology. Further, strain A121 is the first thermophilic representative of genus *Leptodesmis*. From the perspective of genomics, the results of phylogenomics, MLSA, ANI/AAI, *in silico* DNA-DNA hybridization and patristic distance assays reinforced the recent delineation of genus *Leptodesmis* from genus *Phormidesmis*. Comparative genomic analyses revealed that the *Leptodesmis* and *Thermoleptolyngbya* strains from the same hot-spring biome exhibited different genome structures but similar functional classifications of protein-coding genes. Although the core molecular components of photosynthesis, metabolism, and signal transduction were shared by the two strains, distinct features of

genes were also noticed in terms of photosynthesis, and signal transduction, etc., indicating that different strategies might be used by these strains to adapt to the specific niche. In addition, the complete genome of strain A121 provides first insight into the genomic feature of genus *Leptodesmis* and lays foundation for future globally ecogenomic and geogenomic studies.

Taxonomic Treatment and Description of *Leptodesmis sichuanensis* Daroch, Tang, and Shah et al. sp. nov.

The classification system that was applied was based on Komárek et al. (2014).

Taxon description in accordance with the prescriptions of the International Code of Nomenclature for Algae, Fungi and Plants (Shenzhen code) (Turland et al., 2018).

Phylum: Cyanobacteria

Order: Synechococcales

Family: Leptolyngbyaceae

Description: Observed under the light microscopy the filaments of A121 strain were blue-green, thin, straight, bent, wavy, and occasionally curved (**Figure 5**). Trichomes observed with TEM show barrel shaped cells forming a single filament without any branching; meanwhile, SEM and light microscopy photos indicate much longer filaments. Trichomes observed in longitudinal direction were characterized by long cylindrical cells with contraction in the transverse wall. Typical length of each cell ranged from 2.0 to 3.0 μm and had a width of 0.9–1.4 μm (**Figure 5**). Each of the cylindrical cells was separated from a neighboring cell with centripetal invagination of the cell wall, and no connections between the cells were observed. Each of the cells had between three and five thylakoid layers located in parallel on the inner periphery of the cells. Cells were observed to contain carboxysomes, cyanophycin granules, and polyphosphate bodies. The strain was motile. The strain was capable of growing in standard BG11 medium at a maximal temperature of 50°C and in the presence of 0.1 M $NaHCO_3$. The strain can be successfully cryopreserved in 10% DMSO for a period exceeding 24 months.

Type strain: is A121 (= FACHB-3319).

Type species: *Leptodesmis sichuanensis Daroch, Tang, and Shah et al.* sp. nov. (see below).

Etymology: Species epithet derives from the name of the collection site.

Type locality: Ecology of type locality: the sample occurred as macroscopic dark green mat attached to the sinter around the pond with a small amount of mucilage around the entire mat. Sample collection was done in 12.05.2016 with the humidity being close to 71%. Ambient temperature at the time of collection was 15°C and the light intensity was around 1000 lux. The temperature of the hot spring, its pH, and concentration of total dissolved solids were, 40.8°C, 6.32, and 447 mmol L^{-1}, respectively.

Habitat: thermal springs in Ganzi Prefecture of the spring was 6.32 and of Sichuan Province, China (30°05′14″N, 101°56′55″ E) *Leptodesmis* species. The strain is diazotrophic and exhibits experimentally verified nitrogenase activity with acetylene reduction assay, and the corresponding gene cluster has been identified in strains genome. Development of heterocysts was not observed.

Holotype here designated: the culture of *Leptodesmis sichuanensis Daroch, Tang et Shah* sp. *nov*. was initially denoted and deposited in Peking University Algae Collection as PKUAC-SCTA121 has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession number FACHB-3319 as *Leptodesmis* species after identification and authentication on the basis of the full-length sequencing of the 16S rRNA gene along with folding of the secondary structures of the 16S–23S ITS region. After proper identification and authentication, the culture is being maintained in the FACHB under the accession number FACHB-3391.

DATA AVAILABILITY STATEMENT

The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, CP075171.

AUTHOR CONTRIBUTIONS

JT: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing – original draft, writing – review, and editing, visualization, supervision, project administration, funding acquisition. L-MD: methodology, software, data curation. ML: formal analysis, investigation, data curation, writing – original draft. DY: formal analysis, investigation, data curation. YJ: data curation, writing – review and editing. MW and KW: methodology, data curation, formal analysis, visualization, writing – review, and editing. MD: conceptualization, methodology, resources, data curation, writing – original draft, writing – review and editing, supervision, project administration, funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (31970092 and 32071480), the Shenzhen Fundamental Research Program (GXWD20201231165807007-20200806170221001), and the Tenure-Track Fund to MD. Funding bodies had no influence over design and execution of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.739625/full#supplementary-material

REFERENCES

Aguilar, P. S., Hernandez-Arriaga, A. M., Cybulski, L. E., Erazo, A. C., and De, M. (2014). Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*. *EMBO J.* 20, 1681–1691. doi: 10.1093/emboj/20.7.1681

Alcorta, J., Alarcón-Schumacher, T., Salgado, O., and Díez, B. (2020). Taxonomic novelty and distinctive genomic features of hot spring Cyanobacteria. *Front. Genet.* 11:568223. doi: 10.3389/fgene.2020.568223

Bolay, P., Muro-Pastor, M. I., Florencio, F. J., and Klähn, S. (2018). The distinctive regulation of cyanobacterial glutamine synthetase. *Life* 8:52. doi: 10.3390/life8040052

Brito, Â, Ramos, V., Mota, R., Lima, S., Santos, A., Vieira, J., et al. (2017). Description of new genera and species of marine Cyanobacteria from the Portuguese Atlantic coast. *Mol. Phylogenet. Evol.* 111, 18–34. doi: 10.1016/j.ympev.2017.03.006

Cao, P., Cao, D., Si, L., Su, X., Tian, L., Chang, W., et al. (2020). Structural basis for energy and electron transfer of the photosystem I–IsiA–flavodoxin supercomplex. *Nat. Plants* 6, 167–176. doi: 10.1038/s41477-020-0593-7

Chen, J., Yoshinaga, M., Garbinski, L. D., and Rosen, B. P. (2016). Synergistic interaction of glyceraldehydes-3-phosphate dehydrogenase and ArsJ, a novel organoarsenical efflux permease, confers arsenate resistance. *Mol. Microbiol.* 100, 945–953. doi: 10.1111/mmi.13371

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., et al. (2021). Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J.* 15, 211–227. doi: 10.1038/s41396-020-00775-z

Cordeiro, R., Luz, R., Vasconcelos, V., Gonçalves, V., and Fonseca, A. (2020). Cyanobacteria phylogenetic studies reveal evidence for polyphyletic genera from thermal and freshwater habitats. *Diversity* 12:298. doi: 10.3390/d12080298

Elleuche, S., Schfers, C., Blank, S., Schrder, C., and Antranikian, G. (2015). Exploration of extremophiles for high temperature biotechnological processes. *Curr. Opin. Microbiol.* 25, 113–119. doi: 10.1016/j.mib.2015.05.011

Esteves-Ferreira, A. A., Inaba, M., Fort, A., Araújo, W. L., and Sulpice, R. (2018). Nitrogen metabolism in cyanobacteria: metabolic and molecular control, growth consequences and biotechnological applications. *Crit. Rev. Microbiol.* 44, 541–560. doi: 10.1080/1040841X.2018.1446902

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

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Iteman, I., Rippka, R., Tandeau, D. M. N., and Herdman, M. (2000). Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. *Microbiology* 146, 1275–1286. doi: 10.1099/00221287-146-6-1275

Jaime, H. C., Damian, S., Kristoffer, F., Helen, C., Davide, H., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248

Jain, C., and Rodriguez, R. L. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Johansen, J. R., Kovacik, L., Casamatta, D. A., Iková, K. F., and Kaštovský, J. (2011). Utility of 16S-23S ITS sequence and secondary structure for recognition of intrageneric and intergeneric limits within cyanobacterial taxa: *Leptolyngbya corticola* sp. nov. (Pseudanabaenaceae, Cyanobacteria). *Nova Hedwigia* 92, 283–302. doi: 10.1127/0029-5035/2011/0092-0283

Kingston, A. W., and Ponkratz, C. (2017). Rpn (YhgA-Like) proteins of *Escherichia coli* K-12 and their contribution to recA-independent horizontal transfer. *J. Bacteriol.* 199, e787–e716. doi: 10.1128/JB.00787-16

Komárek, J., Kaštovský, J., Mares, J., and Johansen, J. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86, 295–335.

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

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

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

Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liang, Y., Tang, J., Luo, Y., Kaczmarek, M. B., Li, X., and Daroch, M. (2019). *Thermosynechococcus* as a thermophilic photosynthetic microbial cell factory for CO2 utilisation. *Bioresour. Technol.* 278, 255–265. doi: 10.1016/j.biortech.2019.01.089

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Maurya, S. K., Niveshika, and Mishra, R. (2019). "Chapter 24–Importance of bioinformatics in genome mining of cyanobacteria for production of bioactive compounds," in *Cyanobacteria*, eds A. K. Mishra, D. N. Tiwari, and A. N. Rai (Cambridge, MA: Academic Press), 477–506. doi: 10.1016/b978-0-12-814667-5.00024-6

McKenzie, J. L., Robson, J., Berney, M., Smith, T. C., Ruthe, A., Gardner, P. P., et al. (2012). A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. *J. Bacteriol.* 194, 2189–2204. doi: 10.1128/JB.06790-11

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Ogawa, T., and Mi, H. (2007). Cyanobacterial NADPH dehydrogenase complexes. *Photosynth. Res.* 93, 69–77. doi: 10.1007/s11120-006-9128-y

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

Patel, A., Matsakas, L., Rova, U., and Christakopoulos, P. (2019). A perspective on biotechnological applications of thermophilic microalgae and Cyanobacteria. *Bioresour. Technol.* 278, 424–434. doi: 10.1016/j.biortech.2019.01.063

Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328.

Raabova, L., Kovacik, L., Elster, J., and Strunecky, O. (2019). Review of the genus *Phormidesmis* (Cyanobacteria) based on environmental, morphological, and molecular data with description of a new genus *Leptodesmis*. *Phytotaxa* 395, 1–16. doi: 10.11646/phytotaxa.395.1.1

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1901.

Rodriguez-R, L. M., Santosh, G., Harvey, W. T., Ramon, R.-M., Tiedje, J. M., Cole, J. R., et al. (2018). The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* 46, W282–W288. doi: 10.1093/nar/gky467

Roeselers, G., Norris, T. B., Castenholz, R. W., Rysgaard, S., Glud, R. N., Kühl, M., et al. (2007). Diversity of phototrophic bacteria in microbial mats from Arctic hot springs (Greenland). *Environ. Microbiol.* 9, 26–38. doi: 10.1111/j.1462-2920.2006.01103.x

Sciuto, K., and Moro, I. (2016). Detection of the new cosmopolitan genus *Thermoleptolyngbya* (Cyanobacteria, Leptolyngbyaceae) using the 16S rRNA gene and 16S–23S ITS region. *Mol. Phylogenet. Evol.* 105, 15–35. doi: 10.1016/j.ympev.2016.08.010

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

Strunecky, O., Raabova, L., Bernardova, A., Ivanova, A. P., Semanova, A., Crossley, J., et al. (2020). Diversity of Cyanobacteria at the Alaska North Slope with description of two new genera: Gibliniella and Shackletoniella. *FEMS Microbiol. Ecol.* 96:fiz189. doi: 10.1093/femsec/fiz189

Tang, J., Du, L.-M., Liang, Y.-M., and Daroch, M. (2019). Complete genome sequence and comparative analysis of *Synechococcus* sp. CS-601 (SynAce01), a cold-adapted cyanobacterium from an oligotrophic Antarctic habitat. *Int. J. Mol. Sci.* 20: 152. doi: 10.3390/ijms20010152

Tang, J., Jiang, D., Luo, Y., Liang, Y., Li, L., Shah, M. M. R., et al. (2018a). Potential new genera of cyanobacterial strains isolated from thermal springs of western Sichuan, China. *Algal Res.* 31, 14–20. doi: 10.1016/j.algal.2018.01.008

Tang, J., Liang, Y., Jiang, D., Li, L., Luo, Y., Shah, M. M. R., et al. (2018b). Temperature-controlled thermophilic bacterial communities in hot springs of western Sichuan, China. *BMC Microbiol.* 18:134. doi: 10.1186/s12866-018-1271-z

Tang, J., Li, L., Li, M., Du, L., Shah, M. R., Waleron, M., et al. (2021). Description, taxonomy, and comparative genomics of a novel species, *Thermoleptolyngbya sichuanensis* sp. nov., isolated from Hot Springs of Ganzi, Sichuan, China. *Front. Microbiol.* 12:696102. doi: 10.3389/fmicb.2021.696102

Turland, N., Wiersema, J., Barrie, F. R., Greuter, W., and Smith, G. F. (2018). International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Königstein: Koeltz Botanical Books.

Vega-Palas, M. A., Flores, E., and Herrero, A. (2010). NtcA, a global nitrogen regulator from the cyanobacterium Synechococcus that belongs to the Crp family of bacterial regulators. *Mol. Microbiol.* 6, 1853–1859. doi: 10.1111/j.1365-2958.1992.tb01357.x

Walter, J. M., Coutinho, F. H., Dutilh, B. E., Swings, J., Thompson, F. L., and Thompson, C. C. (2017). Ecogenomics and taxonomy of Cyanobacteria phylum. *Front. Microbiol.* 8:2132. doi: 10.3389/fmicb.2017.02132

Wuichet, K., Cantwell, B. J., and Zhulin, I. B. (2010). Evolution and phyletic distribution of two-component signal transduction systems. *Curr. Opin. Microbiol.* 13, 219–225. doi: 10.1016/j.mib.2009.12.011

Wuichet, K., and Zhulin, I. B. (2010). Origins and diversification of a complex signal transduction system in prokaryotes. *Sci. Signal.* 3:ra50. doi: 10.1126/scisignal.2000724

Yasukazu, N., Takakazu, K., Shusei, S., Masahiko, I., Hiroshi, K., Shigemi, S., et al. (2002). Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res.* 9, 123–130.

Zhang, Z. C., Li, Z. K., Yin, Y. C., Li, Y., Jia, Y., and Chen, M. (2019). Widespread occurrence and unexpected diversity of red-shifted chlorophyll producing cyanobacteria in humid subtropical forest ecosystems. *Environ. Microbiol.* 21, 1497–1510. doi: 10.1111/1462-2920.14582

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tang, Du, Li, Yao, Jiang, Waleron, Waleron and Daroch. 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: 21 March 2022
doi: 10.3389/fmicb.2022.803241

A Large-Scale Genome-Based Survey of Acidophilic Bacteria Suggests That Genome Streamlining Is an Adaption for Life at Low pH

*Diego Cortez[1], Gonzalo Neira[1], Carolina González[1], Eva Vergara[1] and David S. Holmes[1,2]**

[1] Center for Bioinformatics and Genome Biology, Centro Ciencia & Vida, Fundación Ciencia & Vida, Santiago, Chile,
[2] Facultad de Medicina y Ciencia, Universidad San Sebastian, Santiago, Chile

OPEN ACCESS

Edited by:
Rafael R. de la Haba, University of Seville, Spain

Reviewed by:
Jeremy Dodsworth, California State University, San Bernardino, United States
Sophie R. Ullrich, Freiberg University of Mining and Technology, Germany
Huaqun Yin, Central South University, China

***Correspondence:**
David S. Holmes
dsholmes2000@yahoo.com

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

Received: *27 October 2021*
Accepted: *07 February 2022*
Published: *21 March 2022*

Citation:
Cortez D, Neira G, González C, Vergara E and Holmes DS (2022) A Large-Scale Genome-Based Survey of Acidophilic Bacteria Suggests That Genome Streamlining Is an Adaption for Life at Low pH. Front. Microbiol. 13:803241. doi: 10.3389/fmicb.2022.803241

The genome streamlining theory suggests that reduction of microbial genome size optimizes energy utilization in stressful environments. Although this hypothesis has been explored in several cases of low-nutrient (oligotrophic) and high-temperature environments, little work has been carried out on microorganisms from low-pH environments, and what has been reported is inconclusive. In this study, we performed a large-scale comparative genomics investigation of more than 260 bacterial high-quality genome sequences of acidophiles, together with genomes of their closest phylogenetic relatives that live at circum-neutral pH. A statistically supported correlation is reported between reduction of genome size and decreasing pH that we demonstrate is due to gene loss and reduced gene sizes. This trend is independent from other genome size constraints such as temperature and G + C content. Genome streamlining in the evolution of acidophilic bacteria is thus supported by our results. The analyses of predicted Clusters of Orthologous Genes (COG) categories and subcellular location predictions indicate that acidophiles have a lower representation of genes encoding extracellular proteins, signal transduction mechanisms, and proteins with unknown function but are enriched in inner membrane proteins, chaperones, basic metabolism, and core cellular functions. Contrary to other reports for genome streamlining, there was no significant change in paralog frequencies across pH. However, a detailed analysis of COG categories revealed a higher proportion of genes in acidophiles in the following categories: "replication and repair," "amino acid transport," and "intracellular trafficking". This study brings increasing clarity regarding the genomic adaptations of acidophiles to life at low pH while putting elements, such as the reduction of average gene size, under the spotlight of streamlining theory.

Keywords: genome reduction, genome streamlining, extremophile, acidophile, chemolithoautotroph, gene gain and loss, protein size reduction and expansion, evolution of acid resistance

INTRODUCTION

Significant differences in genome sizes (number of base pairs per genome) have been detected between closely related lineages of prokaryotes isolated from a broad spectrum of environments, with genome sizes down to 1.2 Mbp in free-living bacteria (Konstantinidis and Tiedje, 2004; Dufresne et al., 2005; Lynch, 2006; Giovannoni et al., 2014; Bentkowski et al., 2015; Martínez-Cano et al., 2015; Rodríguez-Gijón et al., 2021). Small or reduced genomes, also termed streamlined genomes, have been widely observed in microorganisms adapted to live in low-nutrient niches, such as cosmopolitan marine bacterioplankton (Giovannoni et al., 2005; Schneiker et al., 2006; Swan et al., 2013; Luo et al., 2014; Sun and Blanchard, 2014; Graham and Tully, 2021), rivers (Nakai et al., 2016), slow growers in anoxic subsurfaces (Chivian et al., 2008; McMurdie et al., 2009), and in a wide range of extremophiles such as bacteria adapted to supersaturated silica (Saw et al., 2008), halophiles (López-Pérez et al., 2013; Min-Juan et al., 2016), thermophiles (Sabath et al., 2013; Saha et al., 2015; Gu et al., 2021), psychrophiles (Dsouza et al., 2014; Goordial et al., 2016), and alkaliphiles (Suzuki et al., 2014). Differences in genome size have been reported for aerobes vs. anaerobes (Nielsen et al., 2021) and for microorganisms living in warmer vs. cooler environments (Lear et al., 2017; Sauer and Wang, 2019) and in bacterial pathogens (Murray et al., 2021).

The streamlining theory proposes that genome reduction is a selective process that these organisms undergo that promotes their evolutionary fitness (reviewed in Giovannoni et al., 2014). The theory suggests that a smaller genome reduces the energy cost of replication, and by encoding fewer gene products, there is a concomitant reduction of cell size that could optimize transport and nutrient acquisition (Button, 1991; Sowell et al., 2009). Some marine microorganisms with streamlined genomes have been found to have proportionately fewer genes encoding transcriptional regulators and an overall lower abundance of mRNA transcripts per cell, potentially reducing the cost of transcription and translation (Cottrell and Kirchman, 2016). These results are congruent with the observed correlation between regulatory network complexity and genome size (Konstantinidis and Tiedje, 2004). Genome size reduction is also observed in symbiotic microorganisms (Baker et al., 2010; Gao et al., 2014), but it has been theorized that this phenomenon differs to the streamlining of free-living bacteria as the former lose genes by genetic drift due to function redundancy between the host and the symbiont, while the latter would lose them by intense selective pressure (McCutcheon and Moran, 2012; Giovannoni et al., 2014), although recent evidence has argued otherwise (Gu et al., 2021).

Any organism that grows optimally at low pH can technically be classified as an acidophile. However, because there are many neutrophiles (optimum growth ∼pH 7) that successfully grow at around pH 6 or lower, it is useful from a practical point of view to define acidophiles as those microorganisms that grow optimally below pH 5 and make a distinction between moderate acidophiles that grow optimally between pH 5 and about pH 3.0 (Foster, 2004; Dopson, 2016; Benison et al., 2021) and extreme acidophiles that grow below pH 3 (Johnson, 2007). The latter are particularly challenged for survival and growth as they face a proton concentration across their membranes of over 4 orders of magnitude (Baker-Austin and Dopson, 2007; Slonczewski et al., 2009). Acidophilic microorganisms have been identified in all three domains of life (Johnson and Hallberg, 2003), but currently more genomic information is available for prokaryotic acidophiles (Archaea and Bacteria) (Cárdenas et al., 2016; Neira et al., 2020).

Our current understanding about genome streamlining in acidophiles comes from a limited number of observations. It has been reported that the genomes of several acidophilic microorganisms, such as *Methylacidiphilum*, *Ferrovum*, *Leptospirillum* (domain Bacteria) and *Picrophilus* (domain Archaea), are smaller (2.3, 1.9, 2.3, and 1.5 Mb, respectively) compared to their closest neutrophilic phylogenetic relatives (Angelov and Liebl, 2006; Hou et al., 2008; Ullrich et al., 2016; Vergara et al., 2020). Genome reduction in acidophiles has been discussed as a mechanism to reduce energy costs to survive in extremely low-pH environments where organisms must deploy multiple energy-intensive acid resistance mechanisms to maintain a circum-neutral cytoplasmic pH (Hou et al., 2008; Ullrich et al., 2016; Zhang et al., 2017; Vergara et al., 2020) while thriving in often nutrient-scarce and heavy-metal-polluted low-pH environments (Johnson, 1998; Dopson et al., 2003; Johnson and Hallberg, 2008). Despite this progress, there remains much to be discovered about genome reduction in acidophiles. With the increased availability of genome sequences of acidophiles (Cárdenas et al., 2016; Neira et al., 2020), we aim to determine whether there is a statistically supported correlation of genome reduction with low pH and, if so, what are the elements influencing this tendency. We also analyze and comment on the differences in genetic functions between acidophiles and neutrophiles that are involved in these changes.

MATERIALS AND METHODS

Data Procurement and Management

Genome Information

Genomes of 345 bacterial acidophiles together with their associated growth and taxonomic data were obtained from AciDB[1] (Neira et al., 2020). This set of genomes was modified for the present study in two ways: (i) organisms without an identified phylum affiliation were discarded and (ii) seven new genomes and their associated metadata from acidophiles have been added since the publication of AciDB. This resulted in an initial dataset of 342 genomes of acidophiles. In addition, 339 genomes were collected from non-acidophiles (growth optima, pH 5–8). These included 222 genomes of neutrophiles (growth optima, pH 6–8) that were the closest phylogenetic relatives to the acidophiles as identified using the National Center for Biotechnology Information (NCBI) taxonomy (Schoch et al., 2020), GTDB (Chaumeil et al., 2020), and AnnoTree (Mendler et al., 2019), resulting in an equal taxonomic representation

[1]https://acidb.cl/

of genomes of acidophiles and their neutrophilic phylogenetic relatives (**Supplementary Table 1**). The genome sequences were downloaded from NCBI and the Joint Genome Institute (JGI). The genomes were filtered for quality using CheckM v1.0.12, with cutoffs for completeness at > 80% and contamination at < 5% (Parks et al., 2015). This resulted in a final data set of 597 high-quality bacterial genomes, comprising 264 genomes from acidophiles (pH < 5) and 333 genomes from non-acidophiles (pH 5–8). The genome information is provided in **Supplementary Table 2**.

Genome average nucleotide identity was determined using fastANI v1.3 with 4 threads (Jain et al., 2018). A cutoff of 95% average nucleotide identity was defined (Kim et al., 2014) to group identical or highly similar genomes into species clusters. The genomic characteristics, proteomic data, and associated metadata are reported as the means of each group for all plots. This reduced data bias due to over-representation of some highly sequenced species.

Growth pH and Temperature

Data on the optimal growth pH and temperature of a species were downloaded from AciDB (Neira et al., 2020). For new species with sequenced genomes not yet deposited in AciDB, information for optimal growth pH and temperature was extracted from the literature. When no description of these optima was available, they were defined as the midpoint of the growth range reported for the strain or closely related strain as described by Neira et al. (2020). For metagenomes, the reported environmental data were used to determine optimum pH and temperature.

Proteome Analyses

Protein Annotations

The genome annotations were downloaded from NCBI[2] or JGI.[3] Genomes without an existing annotation were annotated with prokka v1.13.3 (Seemann, 2014). A proteome table was generated for each genome, which includes information for each predicted protein, including size, predicted subcellular localization, functional annotation with Clusters of Orthologous Genes (COGs) and Pfams, COG category, and presence of signal peptide and ortholog group. Unless stated, all software was run with default options.

Ortholog Groups

To define ortholog groups, reciprocal BLASTP was performed within each genome by using all the proteins in its predicted proteome as queries against a database of the same proteins. A coverage of 50%, a sequence identity of 50%, and an *e*-value of 10^{-5} were used as cutoffs (Tettelin et al., 2005; Naz et al., 2020). The protein pairs that follow these conditions were assigned to the same ortholog family if one or both were the best-scored BLASTP hit of the other. Ortholog groups will also be referred to as protein families.

[2]www.ncbi.nlm.nih.gov

[3]img.jgi.doe.gov

Subcellular Localization

Subcellular locations were assigned to each predicted protein using PSORTb v3.0 (Yu et al., 2010), which predicts either cytoplasmatic, inner membrane, exported, outer membrane, periplasmic for gram-negative bacteria, or cell wall for gram-positive bacteria. An "unknown" tag is assigned to proteins whose subcellular location could not be predicted. This was complemented with signal peptide identification, which was assigned using SignalP v5.0b that predicts the presence of signal peptides for translocation across the plasmatic membrane by either the Sec/SPI (standard system), Sec/SPII (lipoprotein signal peptide system), or Tat/SPI (alternative system) translocation/signal peptidases (Almagro et al., 2019). All three positive predictions were binned together and tagged as "has signal peptide". The proteins were sorted by both subcellular localization and signal peptide presence.

Pfam and Clusters of Orthologous Genes Functional Annotations

Pfams were assigned to predicted proteins using Pfam_scan v1.6 (Finn et al., 2016) under Pfam version 32.0 (El-Gebali et al., 2019), which contains a total of 17,929 different functional annotations, including protein families and clans. An *e*-value of $< 10^{-5}$ was applied as a cutoff for Pfam predictions of protein function. The Pfam with the lowest *e*-value was assigned to each protein. COG annotations were assigned with the web tool eggNOG-mapper v5.0 (Huerta-Cepas et al., 2019) under the December 2014 version of the COG database, which contains 4,632 functional annotations (Galperin et al., 2015). The percentage of ortholog groups that have a Pfam assignment (Mistry et al., 2021) or a COG assignment (Galperin et al., 2021) was calculated for each proteome. The percentage of ortholog groups belonging to each COG category was also calculated. In addition, Pfam assignments were used for the analysis of intra-protein family size variation and to determine the percentage of proteins with an annotation.

Paralog Frequencies

Paralog families were defined as ortholog groups with two or more proteins from the same proteome. The percentage of proteins that belong in paralog families was calculated for each COG category in relation to the total number of proteins in the category. The same procedure was repeated for the full proteome.

Statistical Analyses

A python script was developed to gather, filter, organize, and analyze the data from the organisms' genomes and proteomes. Data distributions were statistically analyzed using the following methods. The scipy library (Virtanen et al., 2020) was used for linear fittings (with the "linregress" module), binomial test (with the "stats.binom_test" module), and Pearson's linear correlation coefficient (with the "stats.pearsonr" module). A two-sided mode was used for all the tests. The *P*-value thresholds used for statistical significance were 0.05, 0.01, and 0.001. For estimation of correlation in potentially heteroscedastic distributions, generalized least squares was applied using the module

"regression.linear_model.GLS" within the statsmodels library (Seabold and Perktold, 2010). For multi-testing analyses, false discovery rate was used to determine the statistical significance using the Benjamini/Hochberg procedure (Benjamini and Hochberg, 1995) with the "stats.multitest.multipletests" module also within the statsmodels library. A q-value of 0.05 was used for Pearson's correlation p-values. The q-value is the upper limit of the rate of the findings (null hypothesis rejections) that is expected to be a false positive. Principal component analysis (PCA) was performed with the "decomposition. PCA" module within the sklearn library (Pedregosa et al., 2011). The number of components for dimensionality reduction was set to 2. Data was plotted using the matplotlib library (Hunter, 2007).

RESULTS AND DISCUSSION

Phylogenetic Distribution and Associated Metadata of the Genomes Interrogated

From the 342 publicly available genomic sequences (264 high-quality plus 78 low-quality genomes) of acidophilic bacteria, 331 genomes with well-defined taxonomy (phylum and class) were mapped onto a rooted cladogram (**Figure 1**). The genome sequences come from 177 species distributed in 17 classes and 8 phyla out of a total of 37 recognized bacterial phyla (55 if candidate phyla are included) (Schoch et al., 2020; **Figure 1** and **Supplementary Table 3**). The acidophiles are widely distributed in the cladogram, supporting the idea that acidophile lineages have emerged independently multiple times during evolution (Cárdenas et al., 2016; González et al., 2016; Colman et al., 2018; Khaleque et al., 2019; Vergara et al., 2020).

Supplementary Figure 1 shows the distribution of acidophilic species with sequenced genomes by phylum across pH, where pH represents the optimum for growth for each species. The total number of species declines from about 60 species in the range pH 4–5 to about 10 at pH 0.5–1.5, consistent with the observation that species diversity declines in low-pH environments (Bond et al., 2000; Baker and Banfield, 2003; Johnson and Hallberg, 2003; Méndez-García et al., 2014; Lukhele et al., 2020; Hedrich and Schippers, 2021). These estimates are based on the distribution of acidophiles with publicly available sequenced genomes; the true richness of acidophile diversity is likely to be much higher and will probably increase as more acidic econiches are sampled using metagenomics approaches.

Figure 2 shows the distribution of species by percentage across pH. The results have been divided into three sections (a–c) for discussion. Section (a) with a pH range of 1.0–2.0 is dominated by species in the phyla Proteobacteria, Firmicutes, and Nitrospirae in approximately equal proportions at around pH 2 and by Firmicutes at pH 1. Section (b) shows the species distribution in the range pH 2–4. Acidophilic species of phylum Proteobacteria are the most prevalent in this range but exhibit a declining percentage with decreasing pH. Species of Actinobacteria and Verrucomicrobia are represented about equally, but both phyla have few representatives below pH 2. Species of Aquificae are present in a low percentage (∼3%), down to about pH 3, beyond which there are no representative genomes. Section (c) shows the species distribution in the range pH 4–5. All seven phyla (eight, including the one species from Armatimonadetes) have species in this range, but Acidobacteria show a declining percentage from pH 5–4, below which there are no representative genomes.

Genome Size as a Function of pH

A scatterplot of genome size across optimal growth pH shows declining genome sizes from about 4.5 Mb for circum-neutrophiles to an average of about 3.4 Mb for extreme acidophiles (**Figure 3**). There are no large genomes (> 5 Mb) for bacteria that grow below about pH 4, whereas large genomes including up to about 10 Mb are present in acidophiles that grow between pH 4 pH 5 and in neutrophilic relatives of the acidophiles that grow from pH 5 to 8. A linear regression model fitted to the data shows a tendency that is statistically significant with a positive Pearson's correlation coefficient of 0.19 and a p-value of 2.97×10^{-5}, implying that genomes are smaller at a lower pH. However, there is evidence of heteroscedasticity[4] in the plot, which means that the variance is not constant across one of the variables (in this case, the pH), which invalidates Pearson's correlation tests. We applied generalized least squares regression to take into account heteroscedasticity, and a p-value of 1.8×10^{-3} was obtained, supporting the proposed relationship between pH and genome size.

However, the presence of heteroscedasticity suggests the possibility that other variables, in addition to pH, may contribute to the determination of genome size. To address this issue, we investigated the potential contributions of growth temperature and genomic G + C content on the distribution of genome size across pH. Many acidophiles are also moderate or even extreme thermophiles (Johnson and Hallberg, 2003; Capece et al., 2013; Colman et al., 2018), and temperature has been suggested to be a driving force for genome reduction (Sabath et al., 2013). Genome size has also been associated with G + C content, where organisms with relatively low genomic G + C content tend to have smaller genomes (Veloso et al., 2005; Almpanis et al., 2018).

We evaluated how these factors are correlated with genome size and pH (**Supplementary Figure 2**). Temperature is negatively correlated with genome size (Pearson's correlation coefficient, −0.34; p-value, 2.9×10^{-13}), and G + C is positively correlated with genome size (Pearson's correlation coefficient, 0.48; p-value, 1.91×10^{-25}). A negative correlation between genome size and temperature has recently been reported for extreme acidophiles of the *Acidithiobacillus* genus (Sriaporn et al., 2021). However, no statistically supported correlation is observed between temperature and pH (Pearson's correlation coefficient, −0.01; p-value, 0.84) nor between G + C content and pH (Pearson's correlation coefficient, −0.06; p-value, 0.22). Therefore, while both temperature and G + C content

[4] en.wikipedia.org/wiki/Heteroscedasticity

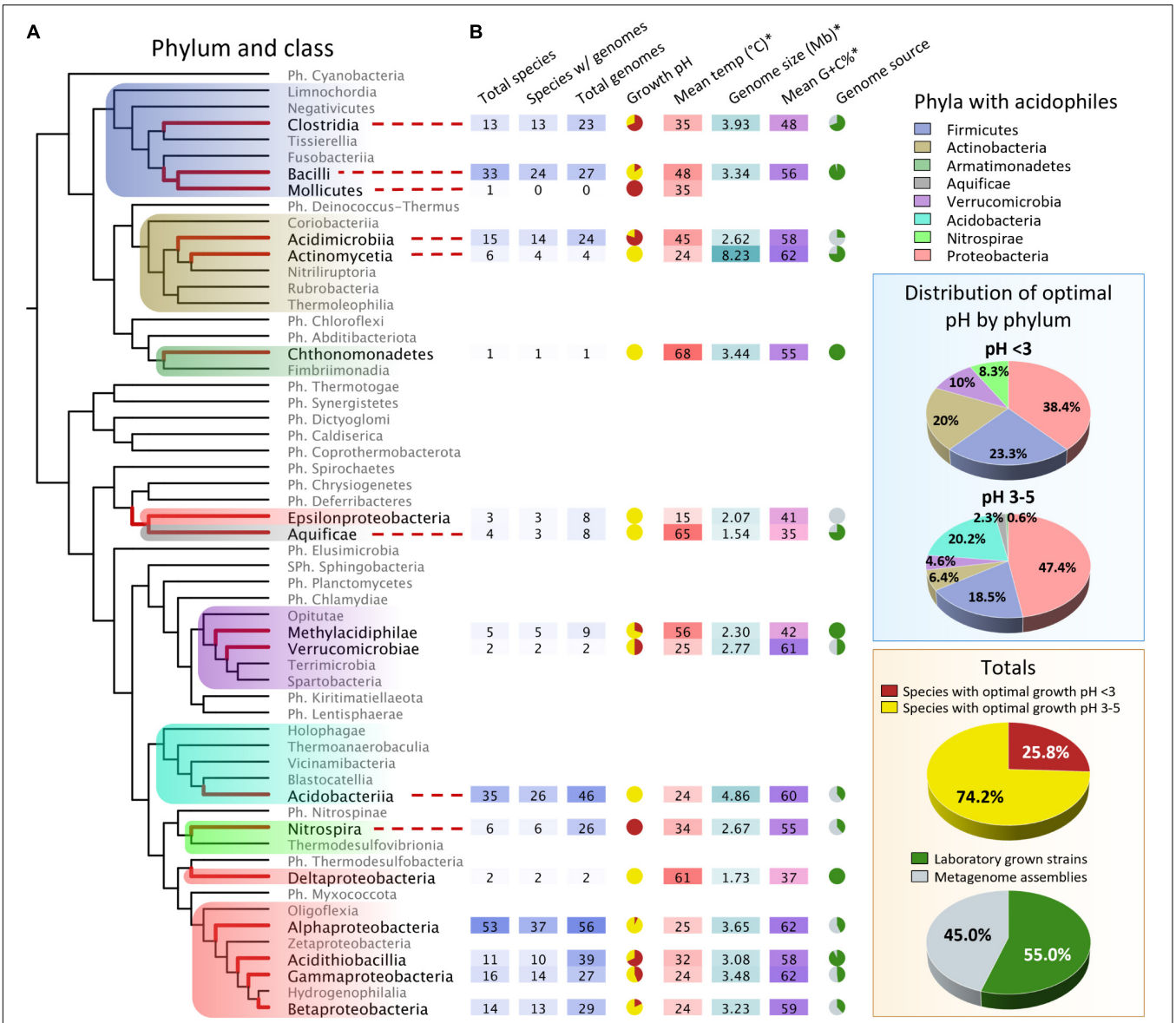


FIGURE 1 | Taxonomic distribution of the acidophilic genomes interrogated. **(A)** A rooted cladogram displaying the phyla, classes, and metadata of acidophiles with genomic data. The cladogram was constructed using AnnoTree (Mendler et al., 2019) as a guide for phylogenetic positioning and rooted as described by Parks et al. (2018). The phyla with acidophiles were broken down into classes. Lineages with known acidophiles are highlighted, and their branches are shown with thick red lines. **(B)** Genomic and growth data of the taxa with acidophiles. Dashed lines connect the acidophilic lineages with the taxon's information when necessary. Growth pH pie charts represent the percentage of species that grow optimally at pH < 3 (red) and at pH 3–5 (yellow). Genome source pie charts represent the percentage of acidophilic genomes sequenced from laboratory pure strains (dark green) vs. metagenome assemblies (gray). **(C)** Percentage of acidophilic species by phyla for both pH ranges (< 3 and 3–5). **(D)** Totals of both pie charts from **(B)** for all the phyla combined. Ph., phylum; Sph., superphylum. The asterisk indicates the mean values for the acidophiles in the taxon. A more detailed table with the classes' information can be found in **Supplementary Table 3**.

have a strong influence on genome size, they appear to act independently of the relationship between pH and genome size.

To investigate further the interplay of pH, temperature, and G + C content with genome size, we performed dimensionality reduction and visualization *via* PCA (Jolliffe, 2005). As seen in **Figure 4**, the directions of the loading vectors show that temperature is negatively correlated with both G + C content and genome size, while genome size is positively correlated with both G + C content and pH. This is also depicted in how the smallest genomes are found in thermophiles (optimal temperature: > 55°C, rightmost cluster) followed by extreme acidophiles (optimal pH: < 3, upmost cluster), while the biggest genomes are found in a high-G + C-content group (leftmost cluster). Conversely, the orthogonality of the loading vectors suggests that no correlation is observed between pH and temperature or between pH and G + C content. Therefore,

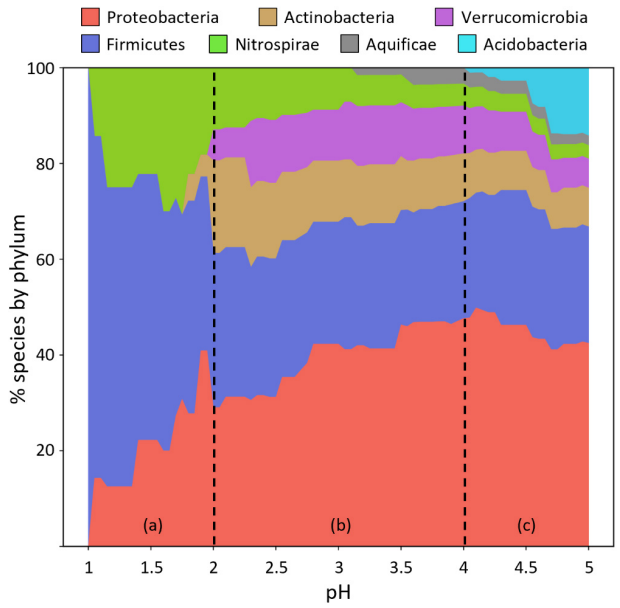


FIGURE 2 | Distribution of acidophilic species with sequenced genomes by phylum across pH. Cumulative plot of relative abundance (%) of acidophiles across pH. Percentages indicate species that can live at or below a given pH. **(a–c)** Indicate pH ranges 1–2, 2–4, and 4–5, respectively. The phyla are color-coded. Phylum *Armatimonadetes* has only one acidophilic species and is not shown.

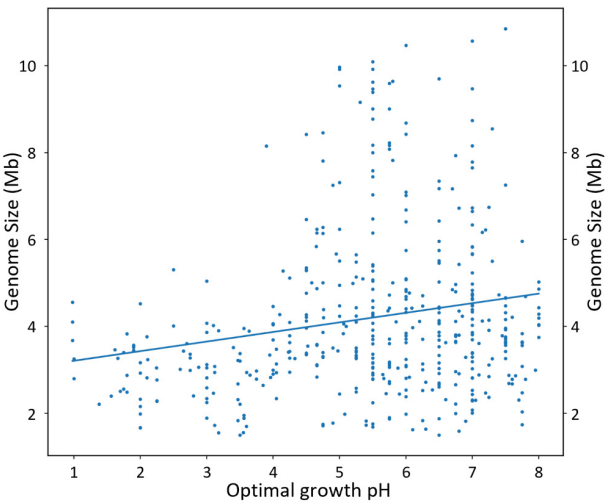


FIGURE 3 | Scatterplot of the genome size (Mb) of bacterial acidophiles and their most closely related extant, circum-neutral relatives vs. optimal growth pH. Each point corresponds to a different species. A linear regression curve has been fitted to the data with a Pearson's correlation coefficient of 0.19 and a *p*-value of 2.97×10^{-5}. The generalized least squares *p*-value was 1.8×10^{-3}.

when considering all variables at once, the same results are observed as when the variables were individually assessed (**Supplementary Figure 2**), providing additional evidence that neither G + C content nor temperature affects the correlation between pH and genome size; rather, multiple driving forces can independently exert their influence on genome size.

Genetic Mechanisms Affecting Genome Sizing

Given the observation that genome size is negatively correlated with pH in acidophiles, we aimed to determine what genomic processes influence this relationship. **Figure 5A** shows a diagrammatic representation of genetic mechanisms that have been postulated to be involved in genome expansion or reduction in Bacteria and Archaea (Keeling and Slamovits, 2005; Sabath et al., 2013; Giovannoni et al., 2014; Gillings, 2017; Kirchberger et al., 2020; Rodríguez-Gijón et al., 2021; Westoby et al., 2021). Genome size changes could result from having changes in the number of orthologous families (i, **Figure 5A**) or paralogous genes (ii, **Figure 5A**), in genome compaction/expansion resulting from changes in the number of intergenic nucleotides, including alteration in the frequency of overlapping genes (iii, **Figure 5A**; reviewed in Kirchberger et al., 2020), and in smaller or larger genes, including loss/gain of domains (iv, **Figure 5A**).

Based on the schema shown in **Figure 5A**, we investigated the contribution of the different mechanisms in genome size changes in acidophiles across pH. Annotated open reading frames (ORFs) were used as surrogates for "genes". A caveat is that ORF prediction depends on the quality of the genome sequence, where poor-quality genomes frequently have incorrectly annotated chimeric and truncated ORFs that confound the subsequent identification of genes (Klassen and Currie, 2013). We minimized these potential errors by analyzing only genomes that had passed a high-quality CheckM filter (Parks et al., 2015), yielding the 597 genomes used in our genomic analyses. However, even high-quality genomes are prone to errors of ORF annotation, especially in the identification of correct translation start sites (Korandla et al., 2020), which will impact the predictions of gene and intergenic spacer sizes. Currently, there is no computational program for ORF prediction that is flawless, including GenBank (Korandla et al., 2020), and we expect that future work will improve the annotations of ORFs used in our study.

Reduction/Expansion of Gene Number

The number of protein coding genes (ORFs) of each genome under interrogation was plotted as a function of the optimal growth pH of the species (**Figure 5B**). The results indicate that there is a statistically significant reduction (Pearson's coefficient: 0.18; *P*-value: 1.25×10^{-4}) of the average number of ORFs per organism across pH from an average of about 4,100 ORFs/organism at pH 7 to about 3,200 ORFs/organism at pH 2 (**Figure 5B**). This has been regarded as possibly the most predominant mechanism for genome size changes (Konstantinidis and Tiedje, 2004), and this is likely also true for our dataset (**Supplementary Figure 3**).

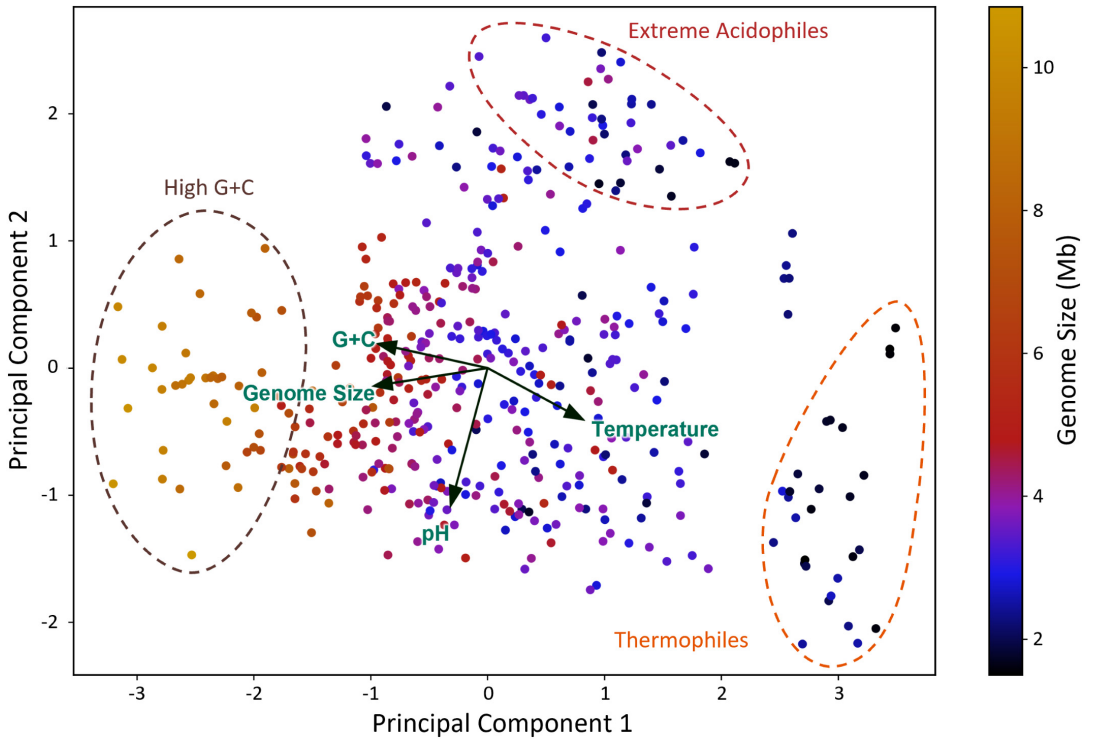


FIGURE 4 | Principal component analysis (PCA) of multiple variables potentially influencing the genome size. Dimensionality reduction was performed by PCA, inputting the optimal growth pH, optimal growth temperature, G + C content, and genome size of each species in the dataset. A biplot was constructed, which shows the loadings of each variable as arrows at the center of the plot and the distribution of the principal components. The average genome size of each species is shown as a color scale. Three clusters within the dotted circles are highlighted for their distinctive features.

Reduction of Intergenic Spacers as a Possible Contributor to Genome Compactness

It is well established that bacteria have compact genomes with an average protein-coding density of 87%, with a typical range of 85–90% (McCutcheon and Moran, 2012). Genome size reduction could occur by decreasing the amount of DNA occupied by intergenic spacers—for example, by promoting the frequency of overlapping genes (Veloso et al., 2005; Saha et al., 2015; Kreitmeier et al., 2021). This strategy has been especially exploited in compacting viral genomes (Pavesi, 2021).

To evaluate whether a reduction in the fraction of the genome dedicated to non-protein-coding DNA contributed to the smaller genomes observed in acidophiles, we calculated the percentage of intergenic spaces (IG) dedicated to the total genome content across pH. IG was calculated as genome size (Mbp)—$\sum$Mbps of all ORFs in a genome, expressed as a percentage of the total Mbps in the genome. A smaller % IG implies greater genome compaction. A tendency was observed for % IG to increase as pH growth optima declines (**Figure 5C**), which is borderline statistically significant (Pearson's coefficient = −0.11; *p*-value, 0.06). An increase in intergenic space in acidophiles is an interesting finding that might be explored further in future studies and indicates that this element is most likely not contributing to the reduced genome sizes of acidophiles. This result is particularly sensitive to the aforementioned errors of ORF annotation, and this influences the estimation of the percentage of intergenic genomic DNA.

Reduction/Increase of Protein Size

The average protein size was plotted as a function of pH (**Figure 5D**). There is a statistically supported positive correlation (p-value: 4.03×10^{-8}) between average protein size and pH, with an average size of 320 amino acids at pH 7 to 300 at pH 2. This indicates that acidophiles have shorter proteins on average, which could be produced by a loss of larger proteins or by protein size reduction (**Figure 5A**, mechanism iv) or possibly both.

To quantify protein size reduction in acidophiles, we analyzed the protein sizes of several conserved Pfams (> 90% of the species) in the dataset (**Figure 6**). We observed that the conserved Pfams with reduced protein sizes in acidophiles are over 5 times as many as the conserved Pfams with increased sizes (**Figure 6A**; binomial test p-value, 2.1×10^{-13}). This result accounts mainly for changes in the predominant domain architectures, implying that these proteins in acidophiles likely have fewer domains—for example, the Pfam for the biotin attachment domain was mainly found without additional domains below pH 5, while in neutrophiles it can often be found next to other domains, such as dihydrolipoamide acyltransferase (**Supplementary Table 4**). This inclination toward protein size reduction is also observed in a collection of conserved Pfams that are also in single copy and predominantly in single-domain architectures (**Figure 6B**;

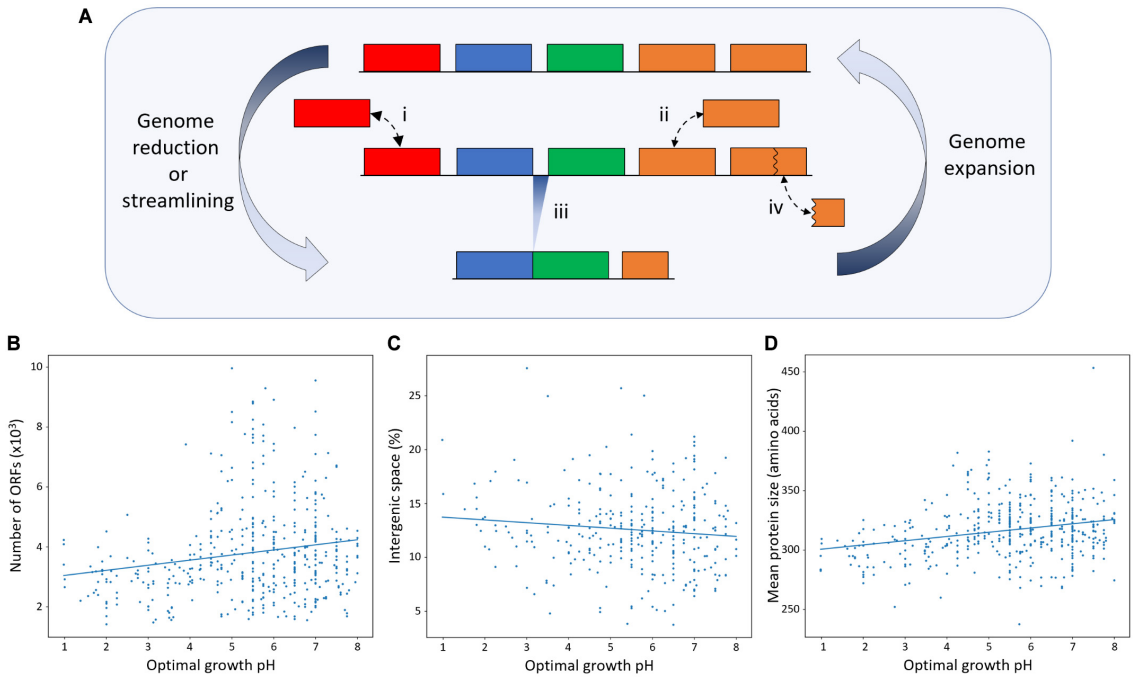


FIGURE 5 | Mechanisms involved in genome size changes. **(A)** Diagrammatic representation of the genetic mechanisms involved in genome size changes. Five genes of a hypothetical genome are shown, where the top, middle, and bottom rows represent an expanded genome, a transition genome, and a streamlined genome, respectively. The orange boxes indicate paralogous genes. The processes involved in genome size changes are shown, where (i) and (ii) represent gene loss/gain of single-copy genes or paralogous genes, respectively, (iii) shows the intergenic space reduction or expansion, which we refer to as genome compaction, and (iv) shows the gene size reduction or increase. **(B)** Number of genes (ORFs, open reading frames) across pH. Pearson's correlation coefficient is 0.18, with p-value 1.25×10^{-4}. **(C)** Intergenic space vs. pH. Intergenic space is defined as the genome size minus the sum of the nucleotide length of all protein-coding genes as defined by the ORFs of a genome divided by genome size, in percentage. A stricter genome quality filter of 97% completeness and 2% contamination was used in this analysis to minimize mis-annotation errors due to fragmented genomes. In total, 394 genomes from 317 species passed the filter. Pearson's correlation coefficient is −0.11, with p-value 0.06. **(D)** Average protein length across pH. Pearson's correlation coefficient is 0.25, with p-value 4.03×10^{-8}.

binomial test p-value, 7.4×10^{-3}). This result accounts mainly for loop size reductions and domain size reductions. Such is the case of the ribosomal protein L19 that, in acidophiles, lacks long loops and is 4 amino acids shorter on average (**Supplementary Table 5**). As for the possible contribution of gene gain/loss into the reduction of the average protein size in acidophiles (by gain of smaller proteins or loss of larger proteins), we estimate that it had a much less significant contribution than protein size reduction (**Supplementary Figure 4**).

Gene Categories Over- and Underrepresented in Acidophiles

Having established that there is a statistically supported positive correlation between genome size and optimal pH for growth and that gene gain and loss events likely contributed to this correlation, we investigated in more detail what types of genes were involved in these events.

Changes in Ortholog Group Representativity in Acidophiles

To gain insight into the contribution of gains or losses of genes in the observed genome size changes of acidophiles (mechanism i, **Figure 5A**), we first clustered the genes into ortholog families and systematically classified the predicted proteomes of each genome by (i) subcellular location and (ii) functional category as predicted by Pfam annotations (Mistry et al., 2021) and COG categories (Galperin et al., 2015). Subsequently, we mapped the frequencies of ortholog families of these categories in the genomes across pH.

Changes in Ortholog Frequencies by Subcellular Location

Figure 7A shows the frequency of occurrence of protein families with subcellular location and/or signal peptide predictions expressed as a percentage of the total protein families per genome. The frequency of predicted cytoplasmic proteins does not change across pH. However, there is a statistically significant decrease (Pearson's correlation coefficient, 0.22; p-value, 1.4×10^{-6}) in the frequency of proteins predicted to have a signal peptide with decreasing pH and a statistically significant increase (Pearson's correlation coefficient, −0.19; p-value, 4.4×10^{-5}) in the frequency of inner membrane proteins with decreasing pH. There is a small but, nevertheless, statistically significant decrease (Pearson's correlation coefficient, 0.21; p-value, 7.5×10^{-6}) in the frequency of proteins predicted to be in the category "periplasm, outer membrane, cell wall, and exported" with decreasing pH.

The decrease in proportion of proteins with signal peptides at low pH is consistent with the observation that there

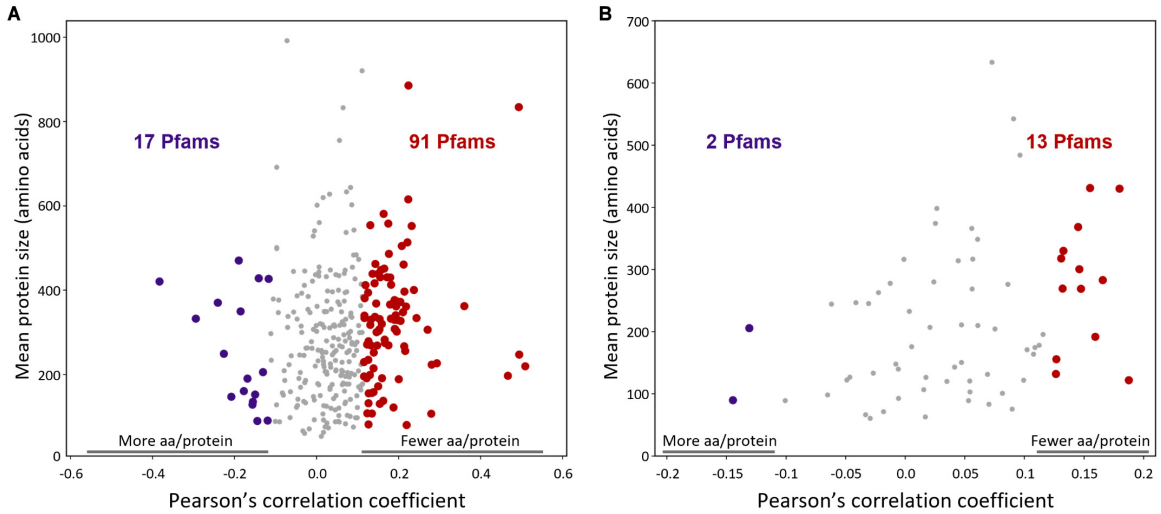


FIGURE 6 | Protein size vs. pH correlations for conserved Pfams. **(A)** Pfams present in over 90% of species and in a pH span of at least 6 pH units were selected for analysis. For each Pfam, the Pearson's correlation coefficient for protein size vs. organism optimal growth pH was calculated using the species averages as data. Each point corresponds to a different Pfam. Positive correlations (91 red points to the right) indicate Pfams whose proteins are shorter at low pH, while negative correlations (17 purple points to the left) are Pfams whose proteins are larger at low pH. The 25 Pfams with the lowest *p*-values are listed in **Supplementary Table 4**. **(B)** Analog to **(A)** but for a list of Pfams that in addition to being present in over 90% of the species and in a span of at least 6 pH units were also in a unique copy in the genomes (proteins with the Pfam per genome < 1.1), and only one domain architecture was dominant in the proteins. These Pfams are listed in **Supplementary Table 5**. For both plots, a false discovery rate *q*-value of 0.05 was used for statistical significance. Significant correlations are shown as big points, which are red for positive correlations and purple for negative correlations. Non-significant correlations are shown as small gray points.

are correspondingly fewer proteins predicted in the category "periplasm, outer membrane, cell wall, and exported" at low pH since most of these proteins require a signal peptide export mechanism to pass through the periplasmic membrane (Green and Mecsas, 2016). We hypothesize that the decrease in relative frequency of proteins found outside the inner membrane in acidophiles could be due to physico-chemical challenges that such proteins would encounter as they are exposed to high concentrations of protons at low pH, potentially limiting the diversity of proteins that have evolved to confront such challenges (D'Abusco et al., 2005; Chi et al., 2007; Duarte et al., 2009, 2011; Panja et al., 2020; Chowhan et al., 2021). We speculate that the observed enrichment of predicted inner membrane protein families in acidophiles (**Figure 7A**) reflects the importance of such proteins in acid stress management since the inner membrane is the barrier that separates the neutral (pH ∼7) cytoplasm from the extreme acid conditions of the periplasm or extracellular space (Slonczewski et al., 2009; Lund et al., 2014; Zhang et al., 2016; Hu et al., 2020; Vergara et al., 2020). This is also supported by the lack of correlation of the representativity of inner membrane proteins with genome size in neutrophiles (**Supplementary Figure 5**), suggesting that this is a specific adaptation to low-pH environments rather than a general streamlining element.

Changes in Ortholog Frequencies by Functional Category

The contribution of gene gain or loss to genome size changes across pH was also analyzed using gene functional classification using COG and Pfam annotations. In total, 25 functional categories are recognized in the 2014 COG database (Galperin et al., 2015), and Pfam v32.0 contains a total of 17,929 families (El-Gebali et al., 2019).[5] The combination of COG and Pfam analyses provides deep and accurate coverage for searching for predicted protein function in our dataset. **Figure 7B** shows that the percentage of proteins per genome with a COG or Pfam annotation decreases at a lower pH with statistical significance (Pearson's correlation coefficients, 0.24 and 0.14; *p*-values, 2×10^{-7} and 2.6×10^{-3}), which is not observed for small neutrophilic genomes (**Supplementary Figure 6**). This indicates that acidophiles have a higher proportion of putative protein-coding genes that are not recognized by either COG or Pfam. These proteins can be classified as non-conserved, hypothetical proteins with no functional prediction, which do not have protein clusters with sufficient entries to have their own functional annotation in the COG or Pfam databases. It is possible that some of these represent poorly annotated sequences and pseudogenes. However, an intriguing possibility is that some could correspond to validated protein-coding genes that are enriched in acidophiles. Their analysis could potentially yield clues about novel acid tolerance mechanisms and other functions enriched in acidophiles. Examples of such proteins have recently been detected, although their functions remain unknown (González et al., 2016; Vergara et al., 2020).

An analysis of the distribution of functional categories across pH using COGs shows that acidophiles are enriched in several functions that could possibly be attributed to their distinctive metabolisms and environmental challenges (**Table 1**)—for example, enrichment in proteins assigned to

[5]https://pfam.xfam.org

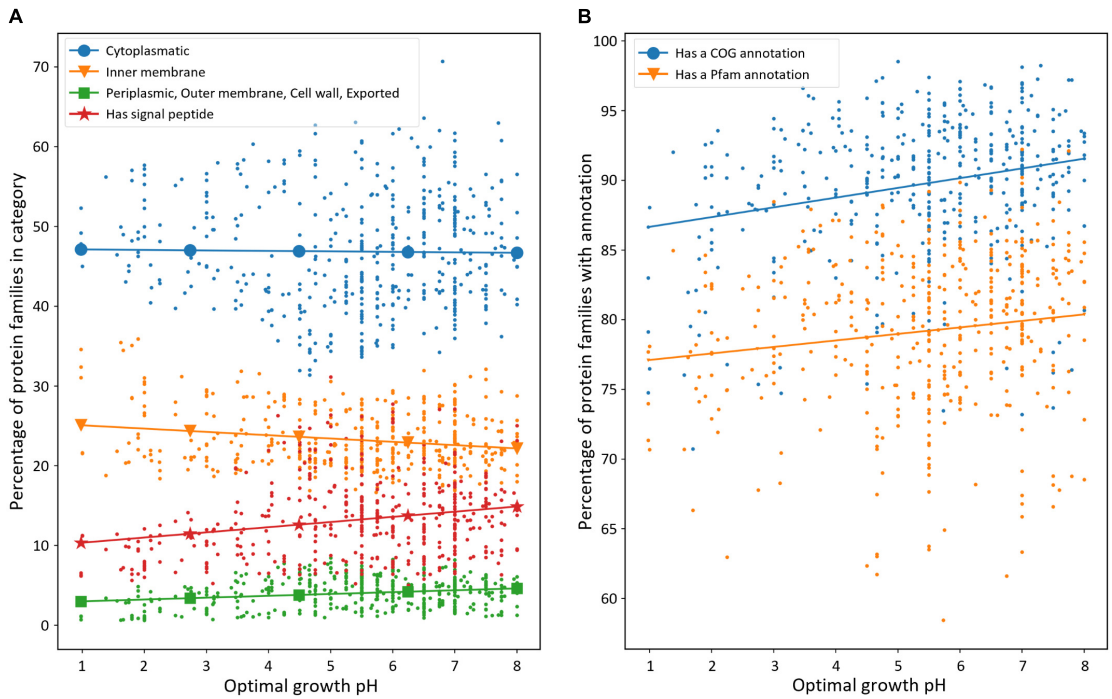


FIGURE 7 | Distribution of protein families across pH. Each point corresponds to a species. **(A)** Subcellular localization and signal peptide presence of protein families across pH. PSORTb and SignalP were used to predict the subcellular location of proteins and signal peptide, respectively. Either subcellular localization or signal peptide presence is expressed in terms of percentage of the protein families (ortholog groups). Pearson's correlation coefficient and *p*-value, respectively, are −0.01 and 0.77 for cytoplasmic (blue), −0.19 and 4.4×10^{-5} for inner membrane (orange), 0.21 and 7.5×10^{-6} for periplasmic, outer membrane, cell wall, and exported (green), and 0.22 and 1.4×10^{-6} for proteins with a signal peptide (red). **(B)** Percentage of protein families with functional classification across pH. The blue data points and the blue line correspond to proteins with a Clusters of Orthologous Genes (COG) annotation, and the orange data points and the orange line correspond to proteins with a Pfam annotation. Pearson's correlation coefficients and *p*-values are, respectively, 0.24 and 2×10^{-7} for proteins with a COG annotation and 0.14 and 2.6×10^{-3} for proteins with a Pfam annotation.

COG L (replication, recombination, and repair) and COG O (chaperone, post-translational modification) might reflect their need for DNA repair and protein refolding when confronted by potentially damaging stresses, such as low pH, high metal concentrations, and oxidative stress (Crossman et al., 2004; Baker-Austin and Dopson, 2007; Cárdenas et al., 2012; Dopson and Holmes, 2014). The increase in the frequency of proteins assigned to COGs C, F, and H (energy production and transport; nucleotide metabolism and transport, and coenzyme metabolism and transport, respectively) could reflect enzyme and pathway requirements associated with obligate autotrophic metabolism that has been found in many acidophiles (Johnson, 1998; Johnson and Hallberg, 2008). As for COG J, it is possible that as ribosomal proteins are very conserved across prokaryotic life (Lecompte et al., 2002), they are less likely to be discarded. Future research could investigate what functions in this category are overrepresented in acidophiles.

On the contrary, the genomes of acidophiles are depleted in proteins assigned to COG T (Signal transduction mechanisms). A depletion of signal transduction mechanisms has been observed in some marine microbes, especially those that are slow-growing types (Gifford et al., 2013; Cottrell and Kirchman, 2016), in the streamlined genome of the extreme acidophile *Methylacidiphilum infernorum* (Hou et al., 2008) and in the metagenomic profiling data of acidic environments (Chen et al., 2015). The abundancy of signal transduction mechanisms generally declines with decreasing genome size, as it has been found that the number of one- and two-component signal transduction systems is proportional to the square of the genome size (Konstantinidis and Tiedje, 2004; Galperin, 2005; Ulrich et al., 2005). Extensive research has been conducted on the different signal pathways and regulatory networks of acidophiles (Rzhepishevska et al., 2007; Shmaryahu et al., 2009; Moinier et al., 2017; Díaz et al., 2018; Osorio et al., 2019). However, additional research is needed to uncover what signal pathways are not present in these organisms. Acidophiles possess several features which may explain their underrepresentation in proteins from this category, such as having small genomes and having a relatively slow growth speed (Fang et al., 2006; Mykytczuk et al., 2010). The genomes of acidophiles also have a proportionately reduced number of proteins assigned to COG S (unknown function). These are proteins with unknown function that are conserved across multiple species.

Paralog Frequency Across pH

We next examined whether the gain or loss of paralogs contributed to genome size changes (mechanism ii, **Figure 5A**). In contrast to what has been described above concerning gain

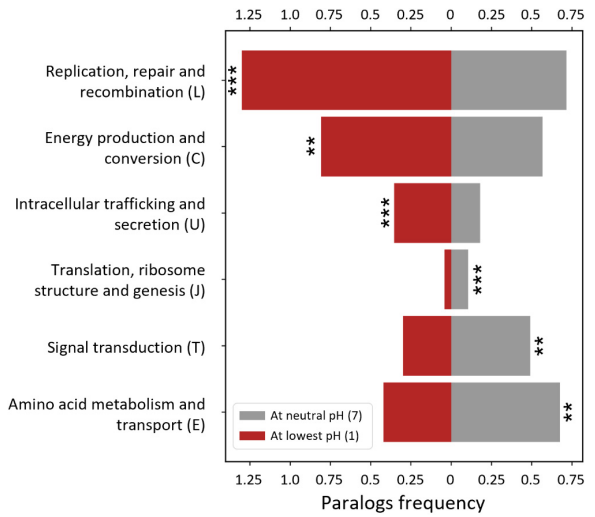


FIGURE 8 | Paralog frequency vs. pH by Clusters of Orthologous Genes (COG) category. The percentage of genes (relative to the proteome size) belonging to paralog families (paralog frequency) was calculated for each COG category. The categories where the paralog frequency had a statistically significant correlation with pH (p-value < 0.01) are shown. The mean duplication frequencies at pH 1 and 7 are displayed, calculated with linear regression (**Supplementary Figure 7**). $^{**}p$-value < 0.01, $^{***}p$-value < 0.001.

TABLE 1 | Genomic representativity of protein families by function as defined by Clusters of Orthologous Genes (COG) categories in acidophile genomes.

COG category	Pearson's correlation coefficient	p-value
Increased representativity in acidophiles (p-value < 0.01)		
(L) Replication, recombination, and repair	−0.25	3.6×10^{-8}
(F) Nucleotide metabolism and transport	−0.21	5.4×10^{-6}
(C) Energy production and conversion	−0.21	8.0×10^{-6}
(H) Coenzyme metabolism and transport	−0.19	3.0×10^{-5}
(D) Cell cycle control and cell division	−0.16	5.2×10^{-4}
(J) Translation and ribosome	−0.15	1.1×10^{-3}
(O) Chaperones, post-translational mod.	−0.13	6.3×10^{-3}
Decreased representativity in acidophiles (p-value < 0.01)		
(S) Function unknown	0.30	1.3×10^{-10}
(T) Signal transduction mechanisms	0.26	3.4×10^{-8}

or loss of specific COG and Pfam gene functions, here we explored how genome size could be influenced by the expansion or contraction of the number of genes in such families. Gene duplication, followed by functional diversification, has been invoked as a major contributor to gene evolution (reviewed in Innan and Kondrashov, 2010; Copley, 2020), and gene paralogs can be present as a significant proportion of a genome (Swan et al., 2013). An increase in the number of paralogous protein copies (including in- and out-paralogs and xenologs; Remm et al., 2001; Darby et al., 2017) has been observed to be correlated with a better performance in a specific function, such as heavy metal resistance or adaptation to other multiple stressors (Kondratyeva et al., 1995; Dulmage et al., 2018). Relatively high paralog frequencies for proteins linked to acid resistance mechanisms have been detected in acidophiles (Ullrich et al., 2016; Vergara et al., 2020).

We analyzed the paralog frequency changes in genomes across pH by COG categories. The COG annotation has been proved useful for gene enrichment analyses across several genomes (Galperin et al., 2021). As can be seen in **Figure 8** and **Supplementary Figure 7**, acidophiles have relatively high paralog frequencies in the COG categories "replication, repair, and recombination", "intracellular trafficking and secretion", and "energy production and conversion" but low frequencies in the COG categories "signal transduction", "translation and ribosome" and "amino acid metabolism", as shown by statistically significant correlations (p-value < 0.01).

High paralog frequencies were found in the "replication, repair, and recombination" category in acidophiles, which add to their overrepresentation of protein families from this category (**Table 1**). This might be attributed to a large number of transposases and integrases and also to DNA repair proteins. The high prevalence of mobile elements and horizontal gene transfer in acidophilic genomes has been previously pointed out as key factors for acidophilic evolution (Aliaga et al., 2009; Acuña et al., 2013; Navarro et al., 2013; Ullrich et al., 2016; Zhang et al., 2017; Colman et al., 2018; Vergara et al., 2020). DNA repair proteins have been found to protect against oxidative stress and heavy metal stress, which acidophiles are exposed to in higher levels (Crossman et al., 2004; Baker-Austin and Dopson, 2007; Cárdenas et al., 2012). As for the increased number of paralogous proteins from the "intracellular trafficking and secretion" category, this could result from an abundance of type II secretory systems involved in conjugation or vesicle-related proteins. The former are frequently associated with mobile elements and are particularly abundant in the flexible genomes of acidophiles (Acuña et al., 2013; Beard et al., 2021). In addition, vesicle-related proteins are linked to biofilm formation (Jan, 2017), which, in turn, has been widely observed in acidophiles (Baker-Austin et al., 2010; González et al., 2013; Díaz et al., 2018; Vargas-Straube et al., 2020). Ultimately, a more detailed examination of what specific functions are duplicated is necessary and remains a topic for future research.

Similar to the results of genome representativity (**Table 1**), the increased paralog frequencies of proteins from the "energy production and conversion" category in acidophiles might be related to their overrepresentation of chemolithotrophic metabolism. Some of the enzymes involved in iron or sulfur oxidation belong to this category, such as cytochrome C, heterodisulfide reductase, and quinone-related proteins (Quatrini et al., 2009; Zhan et al., 2019). Additionally, several proteins in this category are involved in proton exporting functions, such as the H^+-ATPase, and the overall electron transfer chain proteins, such as ubiquinone oxidoreductase

(Walker, 1992; Fütterer et al., 2004; Feng et al., 2015). This indicates that some genes in this category might be in high copy numbers to increase the acid resistance of acidophiles. Alternatively, it could be a consequence of the high energy requirements of maintaining a neutral internal pH (Baker-Austin and Dopson, 2007; Slonczewski et al., 2009).

The reduced paralog frequencies in the "signal transduction" category are concordant with their reduced genome representativity in acidophiles and thus might be accounted by the same phenomena exposed in the previous section about the depletion of these proteins in streamlined organisms. As for the "amino acid transport and metabolism" category, this might be accounted for by a reduction in the number of amino acid importers that are not common in acidophiles. The predominancy of autotrophic metabolism in acidophiles could result in an inclination of these organisms toward the biosynthesis of amino acids rather than uptake by active transporters. Additionally, uptake of amino acids could be harmful to acidophiles as organic acids carry protons into the cytoplasm of these organisms, thus short-circuiting acid resistance mechanisms (Kishimoto et al., 1990; Lehtovirta-Morley et al., 2014; Carere et al., 2021). The current hypothesis is that organic acids are protonated in the extremely acid medium where acidophiles grow (pH < 3), becoming non-ionic and soluble in bacterial membranes and permitting diffusion into the cytoplasm where they uncouple from the proton. A similar phenomenon could occur with amino acids but involve membrane transporters, as amino acids are unlikely to diffuse passively through the membrane.

As for COG J "translation and ribosome," their reduced paralog frequency is opposite to the increased representativity of protein families from this category in the genomes of acidophiles (**Table 1**). In other words, acidophiles tend to discard (or not evolve) duplicated genes from this category rather than losing core functions by relinquishing unique protein families. Further exploration is needed to identify the changes that acidophiles exhibit in this category.

Concordantly, as there was an equilibrium between COG categories with increased and decreased paralog frequencies in acidophiles, the overall paralog frequency had no statistically significant correlation with optimal pH and remained at a relatively constant 8% average, ranging from 2 to 20% (**Supplementary Figure 8**). These relatively low percentages indicate that paralog frequencies are only a minor contributor to genome size changes in our dataset. The constant paralog frequency across pH still contradicts what has been found for other streamlined organisms, which have a relatively low number of paralogs (Giovannoni et al., 2005; Swan et al., 2013). This unusual finding could be partially a consequence of acid resistance genes in multiple copies that would compensate the evolutionary pressure of discarding paralogs.

CONCLUSION

We have shown that acidophilic bacteria possess several streamlining features, such as having smaller genomes, fewer ORFs, smaller proteins, and an underrepresentation of signal transduction proteins. Some features that have been described as important in genome reduction in several systems were not detected in acidophiles, such as lower intergenic space percentages and lower overall paralog frequencies. Our study had a statistical approach in contraposition to other streamlining studies which focus on single clades. When considering a dataset of several hundred genomes, our results suggest that the organisms lose genes in the process of adapting to low-pH environments. The reduction in average protein size is an element that has not been the focus of other streamlining studies and is an interesting topic to be developed further in future studies. In addition, several of our findings shed light on the ever-expanding knowledge about acidophile ecology and their acid resistance systems. Mainly, the higher representativity of inner membrane proteins and increased paralog frequencies in COG categories possibly related to energy production, DNA repair, and biofilm formation. The investigation of which functions might be in higher copy number in acidophiles is an interesting topic for future research, as it may uncover novel survival mechanisms for acidophiles. Similarly, acid-related genes shared between acidophiles could be hidden among the proteins without functional annotation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

DC, GN, and DH designed the research and analyzed the data. DC performed the research. DC and DH wrote the manuscript. CG and EV participated in the construction of the final manuscript. All authors read and approved the final manuscript.

FUNDING

DH was supported by the Fondecyt 1181717 and Programa de Apoyo a Centros con Financiamiento Basal AFB170004 to Fundación Ciencia and Vida.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.803241/full#supplementary-material

REFERENCES

Acuña, L. G., Cárdenas, J. P., Covarrubias, P. C., Haristoy, J. J., Flores, R., Nuñez, H., et al. (2013). Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus* caldus. *PLoS One* 8:11. doi: 10.1371/journal.pone.0078237

Aliaga, D. S., Denef, V. J., Singer, S. W., VerBerkmoes, N. C., Lefsrud, M., Mueller, R. S., et al. (2009). Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing "*Leptospirillum rubarum*"(Group II) and "*Leptospirillum ferrodiazotrophum*"(Group III) bacteria in acid mine drainage biofilms. *Appl. Environ. Microbiol.* 75, 4599–4615. doi: 10.1128/AEM.02943-08

Almagro, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Almpanis, A., Swain, M., Gatherer, D., and McEwan, N. (2018). Correlation between bacterial G+ C content, genome size and the G+ C content of associated plasmids and bacteriophages. *Microb. Genom.* 4:4. doi: 10.1099/mgen.0.000168

Angelov, A., and Liebl, W. (2006). Insights into extreme thermoacidophily based on genome analysis of *Picrophilus torridus* and other thermoacidophilic archaea. *J. Biotechnol.* 126, 3–10. doi: 10.1016/j.jbiotec.2006.02.017

Baker, B. J., and Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* 44, 139–152. doi: 10.1016/S0168-6496(03)00028-X

Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl. Acad. Sci. USA* 107, 8806–8811. doi: 10.1073/pnas.0914470107

Baker-Austin, C., and Dopson, M. (2007). Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* 15, 165–171. doi: 10.1016/j.tim.2007.02.005

Baker-Austin, C., Potrykus, J., Wexler, M., Bond, P. L., and Dopson, M. (2010). Biofilm development in the extremely acidophilic archaeon '*Ferroplasma acidarmanus*'. *Fer1. Extremophiles* 14, 485–491. doi: 10.1007/s00792-010-0328-1

Beard, S., Ossandon, F. J., Rawlings, D. E., and Quatrini, R. (2021). The Flexible Genome of Acidophilic Prokaryotes. *Curr. Issues Mol. Biol.* 40, 231–266. doi: 10.21775/cimb.040.231

Benison, K. C., O'Neill, W. K., Blain, D., and Hallsworth, J. E. (2021). Water activities of acid brine lakes approach the limit for life. *Astrobiology* 21, 729–740. doi: 10.1089/ast.2020.2334

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bentkowski, P., Van Oosterhout, C., and Mock, T. (2015). A model of genome size evolution for prokaryotes in stable and fluctuating environments. *Genome Biol. Evol.* 7, 2344–2351. doi: 10.1093/gbe/evv148

Bond, P. L., Druschel, G. K., and Banfield, J. F. (2000). Comparison of acid mine drainage microbial communities in physically and geochemically distinct ecosystems. *Appl. Environ. Microbiol.* 66, 4962–4971. doi: 10.1128/AEM.66.11.4962-4971.2000

Button, D. K. (1991). Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the Michaelis constant. *Appl. Environ. Microbiol.* 57, 2033–2038. doi: 10.1128/aem.57.7.2033-2038.1991

Capece, M. C., Clark, E., Saleh, J. K., Halford, D., Heinl, N., Hoskins, S., et al. (2013). *Polyextremophiles and the constraints for terrestrial habitability. In Polyextremophiles*. Dordrecht: Springer, 3–59. doi: 10.1007/978-94-007-6488-0_1

Cárdenas, J. P., Moya, F., Covarrubias, P., Shmaryahu, A., Levicán, G., Holmes, D. S., et al. (2012). Comparative genomics of the oxidative stress response in bioleaching microorganisms. *Hydrometallurgy* 127, 162–167.

Cárdenas, J. P., Quatrini, R., and Holmes, D. S. (2016). *Progress in acidophile genomics. Acidophiles: life in extremely acidic environments*. Norfolk: Caister Academic Press, 179–197. doi: 10.21775/9781910190333

Carere, C. R., Hards, K., Wigley, K., Carman, L., Houghton, K. M., Cook, G. M., et al. (2021). Growth on formic acid is dependent on intracellular pH homeostasis for the thermoacidophilic methanotroph *Methylacidiphilum* sp. RTK17.1. *Front. Microbiol.* 12:651744. doi: 10.3389/fmicb.2021.651744

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinform* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, L. X., Hu, M., Huang, L. N., Hua, Z. S., Kuang, J. L., Li, S. J., et al. (2015). Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. *ISME J.* 9, 1579–1592. doi: 10.1038/ismej.2014.245

Chi, A., Valenzuela, L., Beard, S., Mackey, A. J., Shabanowitz, J., Hunt, D. F., et al. (2007). Periplasmic proteins of the extremophile *Acidithiobacillus ferrooxidans*: a high throughput proteomics analysis. *Mol. Cell. Proteom.* 6, 2239–2251. doi: 10.1074/mcp.M700042-MCP200

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., et al. (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 322, 275–278. doi: 10.1126/science.1155495

Chowhan, R. K., Hotumalani, S., Rahaman, H., and Singh, L. R. (2021). pH induced conformational alteration in human peroxiredoxin 6 might be responsible for its resistance against lysosomal pH or high temperature. *Sci. Rep.* 11, 1–10. doi: 10.1038/s41598-021-89093-8

Colman, D. R., Poudel, S., Hamilton, T. L., Havig, J. R., Selensky, M. J., Shock, E. L., et al. (2018). Geobiological feedbacks and the evolution of thermoacidophiles. *ISME J.* 12, 225–236. doi: 10.1038/ismej.2017.162

Copley, S. D. (2020). Evolution of new enzymes by gene duplication and divergence. *FEBS J.* 287, 1262–1283. doi: 10.1111/febs.15299

Cottrell, M. T., and Kirchman, D. L. (2016). Transcriptional control in marine copiotrophic and oligotrophic bacteria with streamlined genomes. *Appl. Environ. Microbiol.* 82, 6010–6018. doi: 10.1128/AEM.01299-16

Crossman, L., Holden, M., Pain, A., and Parkhill, J. (2004). Genomes beyond compare. *Nat. Rev. Microbiol.* 2, 616–617. doi: 10.1038/nrmicro961

D'Abusco, A. S., Casadio, R., Tasco, G., Giangiacomo, L., Giartosio, A., Calamia, V., et al. (2005). Oligomerization of Sulfolobus solfataricus signature amidase is promoted by acidic pH and high temperature. *Archaea* 1, 411–423. doi: 10.1155/2005/543789

Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., and Durand, D. (2017). Xenolog classification. *Bioinform.* 33, 640–649. doi: 10.1093/bioinformatics/btw686

Díaz, M., Castro, M., Copaja, S., and Guiliani, N. (2018). Biofilm formation by the acidophile bacterium *Acidithiobacillus thiooxidans* involves c-di-GMP pathway and Pel exopolysaccharide. *Genes* 9:113. doi: 10.3390/genes9020113

Dopson, M., Baker-Austin, C., Koppineedi, P. R., and Bond, P. L. (2003). Growth in sulfidic mineral environments: metal resistance mechanisms in acidophilic micro-organisms. *Microbiology* 149, 1959–1970. doi: 10.1099/mic.0.26296-0

Dopson, M., and Holmes, D. S. (2014). Metal resistance in acidophilic microorganisms and its significance for biotechnologies. *Appl. Microbiol. Biotechnol.* 98, 8133–8144. doi: 10.1007/s00253-014-5982-2

Dopson, M. (2016). "Physiological and phylogenetic diversity of acidophilic bacteria," in *Acidophiles: Life in Extremely Acidic Environments*, Vol. 2016, eds R. Quatrini and D. B. Johnson (Norfolk: Caister Academic Press), 79–92. doi: 10.21775/9781910190333

Dsouza, M., Taylor, M. W., Turner, S. J., and Aislabie, J. (2014). Genome-based comparative analyses of Antarctic and temperate species of Paenibacillus. *PLoS One* 9:10. doi: 10.1371/journal.pone.0108009

Duarte, F., Araya-Secchi, R., González, W., Perez-Acle, T., González-Nilo, D., and Holmes, D. S. (2009). Protein function in extremely acidic conditions: molecular simulations of a predicted aquaporin and a potassium channel in *Acidithiobacillus ferrooxidans*. *Adv. Mat. Res.* 71, 211–214. doi: 10.4028/www.scientific.net/AMR.71-73.211

Duarte, F., Sepulveda, R., Araya, R., Flores, S., Perez-Acle, T., Gonzales, W., et al. (2011). "Mechanisms of protein stabilization at very low pH," in *Proc. 19th International Biohydrometallurgy Symposium*, (Changsha), 349–353.

Dufresne, A., Garczarek, L., and Partensky, F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Gen. Bol.* 6:R14. doi: 10.1186/gb-2005-6-2-r14

Dulmage, K. A., Darnell, C. L., Vreugdenhil, A., and Schmid, A. K. (2018). Copy number variation is associated with gene expression change in archaea. *Microb. Genom.* 4:9. doi: 10.1099/mgen.0.000210

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Fang, H. H., Zhang, T., and Li, C. (2006). Characterization of Fe-hydrogenase genes diversity and hydrogen-producing population in an acidophilic sludge. *J. Biotechnol.* 126, 357–364. doi: 10.1016/j.jbiotec.2006.04.023

Feng, S., Yang, H., and Wang, W. (2015). System-level understanding of the potential acid-tolerance components of *Acidithiobacillus thiooxidans* ZJJN-3 under extreme acid stress. *Extremophiles* 19, 1029–1039. doi: 10.1007/s00792-015-0780-z

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Foster, J. W. (2004). *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat. Rev. Microbiol.* 2, 898–907. doi: 10.1038/nrmicro1021

Fütterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., et al. (2004). Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc. Natl. Acad. Sci. USA* 101, 9091–9096. doi: 10.1073/pnas.0401356101

Galperin, M. Y. (2005). A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol.* 5, 1–19. doi: 10.1186/1471-2180-5-35

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Galperin, M. Y., Wolf, Y. I., Garushyants, S. K., Vera Alvarez, R., and Koonin, E. V. (2021). Nonessential Ribosomal Proteins in Bacteria and Archaea Identified Using Clusters of Orthologous Genes. *J. Bacteriol.* 203, e58–e21. doi: 10.1128/JB.00058-21

Gao, Z. M., Wang, Y., Tian, R. M., Wong, Y. H., Batang, Z. B., Al-Suwailem, A. M., et al. (2014). Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont "Candidatus *Synechococcus spongiarum*". *MBio* 5, e79–e14. doi: 10.1128/mBio.00079-14

Gifford, S. M., Sharma, S., Booth, M., and Moran, M. A. (2013). Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J.* 7, 281–298. doi: 10.1038/ismej.2012.96

Gillings, M. R. (2017). Lateral gene transfer, bacterial genome evolution, and the Anthropocene. *Ann. N. Y. Acad. Sci.* 1389, 20–36. doi: 10.1111/nyas.13213

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309, 1242–1245. doi: 10.1126/science.1114057

Giovannoni, S. J., Thrash, J. C., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565. doi: 10.1038/ismej.2014.60

González, A., Bellenberg, S., Mamani, S., Ruiz, L., Echeverría, A., Soulère, L., et al. (2013). AHL signaling molecules with a large acyl chain enhance biofilm formation on sulfur and metal sulfides by the bioleaching bacterium *Acidithiobacillus ferrooxidans*. *Appl. Microbiol. Biotechnol.* 97, 3729–3737. doi: 10.1007/s00253-012-4229-3

González, C., Lazcano, M., Valdés, J., and Holmes, D. S. (2016). Bioinformatic analyses of unique (orphan) core genes of the genus *Acidithiobacillus*: functional inferences and use as molecular probes for genomic and metagenomic/transcriptomic interrogation. *Front. Microbiol.* 7:2035. doi: 10.3389/fmicb.2016.02035

Goordial, J., Raymond-Bouchard, I., Zolotarov, Y., de Bethencourt, L., Ronholm, J., Shapiro, N., et al. (2016). Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile *Rhodococcus* sp. JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica. *FEMS Microbiol. Ecol.* 92:2. doi: 10.1093/femsec/fiv154

Graham, E. D., and Tully, B. J. (2021). Marine *Dadabacteria* exhibit genome streamlining and phototrophy-driven niche partitioning. *ISME J.* 15, 1248–1256. doi: 10.1038/s41396-020-00834-5

Green, E. R., and Mecsas, J. (2016). Bacterial secretion systems: an overview. *Microbiol. Spectr.* 4, 4–1. doi: 10.1128/microbiolspec.VMBF-0012-2015

Gu, J., Wang, X., Ma, X., Sun, Y., Xiao, X., and Luo, H. (2021). Unexpectedly high mutation rate of a deep-sea hyperthermophilic anaerobic archaeon. *ISME J.* 15, 1862–1869. doi: 10.1038/s41396-020-00888-5

Hedrich, S., and Schippers, A. (2021). Distribution of acidophilic microorganisms in natural and man-made acidic environments. *Curr. Issues Mol. Biol.* 40, 25–48. doi: 10.21775/cimb.040.025

Hou, S., Makarova, K. S., Saw, J. H., Senin, P., Ly, B. V., Zhou, Z., et al. (2008). Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum Verrucomicrobia. *Biol. Direct* 3:26. doi: 10.1186/1745-6150-3-26

Hu, W., Feng, S., Tong, Y., Zhang, H., and Yang, H. (2020). Adaptive defensive mechanism of bioleaching microorganisms under extremely environmental acid stress: advances and perspectives. *Biotechnol. Adv.* 42:107580. doi: 10.1016/j.biotechadv.2020.107580

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 1–8. doi: 10.1038/s41467-018-07641-9

Jan, A. T. (2017). Outer membrane vesicles (OMVs) of gram-negative bacteria: a perspective update. *Front. Microbiol.* 8:1053. doi: 10.3389/fmicb.2017.01053

Johnson, D. B. (1998). Biodiversity and ecology of acidophilic microorganisms. *FEMS Microbiol. Ecol.* 27, 307–317. doi: 10.1111/j.1574-6941.1998.tb00547.x

Johnson, D. B., and Hallberg, K. B. (2003). The microbiology of acidic mine waters. *Res. Microbiol.* 154, 466–473. doi: 10.1016/S0923-2508(03)00114-1

Johnson, D. B. (2007). Physiology and ecology of acidophilic microorganisms. *Physiol. Biochem. Extr.* 2007, 255–270. doi: 10.1128/9781555815813.ch20

Johnson, D. B., and Hallberg, K. B. (2008). Carbon, iron and sulfur metabolism in acidophilic micro-organisms. *Adv. Microb. Physiol.* 54, 201–255. doi: 10.1016/S0065-2911(08)00003-9

Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of Statistics in Behavioral Science* 2005, 501. doi: 10.1002/0470013192.bsa501

Keeling, P. J., and Slamovits, C. H. (2005). Causes and effects of nuclear genome reduction. *Curr. Opin. Genet. Dev.* 15, 601–608. doi: 10.1016/j.gde.2005.09.003

Khaleque, H. N., González, C., Kaksonen, A. H., Boxall, N. J., Holmes, D. S., and Watkin, E. L. (2019). Genome-based classification of two halotolerant extreme acidophiles, *Acidihalobacter prosperus* V6 (= DSM 14174= JCM 32253) and *Acidihalobacter ferrooxidans*' V8 (= DSM 14175= JCM 32254) as two new species, *Acidihalobacter aeolianus* sp. nov. and *Acidihalobacter ferrooxydans* sp. nov., respectively. *Int. J. Syst. Evol. Microbiol.* 69, 1557–1565. doi: 10.1099/ijsem.0.003313

Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64(Pt_2), 346–351. doi: 10.1099/ijs.0.059774-0

Kirchberger, P. C., Schmidt, M. L., and Ochman, H. (2020). The ingenuity of bacterial genomes. *Annu. Rev. Microbiol.* 74, 815–834. doi: 10.1146/annurev-micro-020518-115822

Kishimoto, N., Inagaki, K., Sugio, T., and Tano, T. (1990). Growth inhibition of *Acidiphilium* species by organic acids contained in yeast extract. *J. Biosci. Bioeng.* 70, 7–10. doi: 10.1016/0922-338X(90)90021-N

Klassen, J. L., and Currie, C. R. (2013). ORFcor: identifying and accommodating ORF prediction inconsistencies for phylogenetic analysis. *PLoS One* 8:e58387. doi: 10.1371/journal.pone.0058387

Kondratyeva, T. F., Muntyan, L. N., and Karavaiko, G. I. (1995). Zinc-and arsenic-resistant strains of *Thiobacillus ferrooxidans* have increased copy numbers of chromosomal resistance genes. *Microbiology* 141, 1157–1162. doi: 10.1099/13500872-141-5-1157

Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. USA* 101, 3160–3165. doi: 10.1073/pnas.0308653100

Korandla, D. R., Wozniak, J. M., Campeau, A., Gonzalez, D. J., and Wright, E. S. (2020). AssessORF: combining evolutionary conservation and proteomics to assess prokaryotic gene predictions. *Bioinform.* 36, 1022–1029. doi: 10.1093/bioinformatics/btz714

Kreitmeier, M., Ardern, Z., Abele, M., Ludwig, C., Scherer, S., and Neuhaus, K. (2021). Shadow ORFs illuminated: long overlapping genes in *Pseudomonas*

aeruginosa are translated and under purifying selection. *bioRxiv* doi: 10.2139/ssrn.3866842

Lear, G., Lau, K., Perchec, A. M., Buckley, H. L., Case, B. S., Neale, M., et al. (2017). Following Rapoport's Rule: the geographic range and genome size of bacterial taxa decline at warmer latitudes. *Environ. Microbiol.* 19, 3152–3162. doi: 10.1111/1462-2920.13797

Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* 30, 5382–5390. doi: 10.1093/nar/gkf693

Lehtovirta-Morley, L. E., Ge, C., Ross, J., Yao, H., Nicol, G. W., and Prosser, J. I. (2014). Characterisation of terrestrial acidophilic archaeal ammonia oxidisers and their inhibition and stimulation by organic compounds. *FEMS Microbiol. Ecol.* 89, 542–552. doi: 10.1111/1574-6941.12353

López-Pérez, M., Ghai, R., Leon, M. J., Rodríguez-Olmos, A., Copa-Patiño, J. L., Soliveri, J., et al. (2013). Genomes of "*Spiribacter*", a streamlined, successful halophilic bacterium. *BMC Genom* 14:787. doi: 10.1186/1471-2164-14-787

Lukhele, T., Selvarajan, R., Nyoni, H., Mamba, B. B., and Msagati, T. A. (2020). Acid mine drainage as habitats for distinct microbiomes: current knowledge in the era of molecular and omic technologies. *Curr. Microbiol.* 77, 657–674. doi: 10.1007/s00284-019-01771-z

Lund, P., Tramonti, A., and De Biase, D. (2014). Coping with low pH: molecular strategies in neutralophilic bacteria. *FEMS Microbiol. Rev.* 38, 1091–1125. doi: 10.1111/1574-6976.12076

Luo, H., Swan, B. K., Stepanauskas, R., Hughes, A. L., and Moran, M. A. (2014). Evolutionary analysis of a streamlined lineage of surface ocean Roseobacters. *ISME J.* 8, 1428–1439. doi: 10.1038/ismej.2013.248

Lynch, M. (2006). Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60, 327–349. doi: 10.1146/annurev.micro.60.080805.142300

Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2015). Evolution of small prokaryotic genomes. *Front. Microbiol.* 5:742. doi: 10.3389/fmicb.2014.00742

McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670

McMurdie, P. J., Behrens, S. F., Müller, J. A., Goke, J., Ritalahti, K. M., Wagner, R., et al. (2009). Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genet.* 5:11. doi: 10.1371/journal.pgen.1000714

Méndez-García, C., Mesa, V., Sprenger, R. R., Richter, M., Diez, M. S., Solano, J., et al. (2014). Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. *ISME J.* 8, 1259–1274. doi: 10.1038/ismej.2013.242

Mendler, K., Chen, H., Parks, D. H., Lobb, B., Hug, L. A., and Doxey, A. C. (2019). AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* 47, 4442–4448. doi: 10.1093/nar/gkz246

Min-Juan, X. U., Jia-Hua, W. A. N. G., Xu-Liang, B. U., He-Lin, Y. U., Peng, L. I, Hong-Yu, O. U., et al. (2016). Deciphering the streamlined genome of *Streptomyces xiamenensis* 318 as the producer of the anti-fibrotic drug candidate xiamenmycin. *Sci. Rep.* 6:18977. doi: 10.1038/srep18977

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Moinier, D., Byrne, D., Amouric, A., and Bonnefoy, V. (2017). The global redox responding RegB/RegA signal transduction system regulates the genes involved in ferrous iron and inorganic sulfur compound oxidation of the acidophilic *Acidithiobacillus ferrooxidans*. *Front. Microbiol.* 8:1277. doi: 10.3389/fmicb.2017.01277

Murray, G. G., Charlesworth, J., Miller, E. L., Casey, M. J., Lloyd, C. T., Gottschalk, M., et al. (2021). Genome reduction is associated with bacterial pathogenicity across different scales of temporal and ecological divergence. *Mol. Biol. Evol.* 38, 1570–1579. doi: 10.1093/molbev/msaa323

Mykytczuk, N. C. S., Trevors, J. T., Ferroni, G. D., and Leduc, L. G. (2010). Cytoplasmic membrane fluidity and fatty acid composition of *Acidithiobacillus ferrooxidans* in response to pH stress. *Extremophiles* 14, 427–441. doi: 10.1007/s00792-010-0319-2

Nakai, R., Fujisawa, T., Nakamura, Y., Nishide, H., Uchiyama, I., Baba, T., et al. (2016). Complete genome sequence of *Aurantimicrobium minutum* type Strain KNCT, a planktonic ultramicrobacterium isolated from river water. *Genome Announc.* 4, e616–e616. doi: 10.1128/genomeA.00616-16

Navarro, C. A., von Bernath, D., and Jerez, C. A. (2013). Heavy metal resistance strategies of acidophilic bacteria and their acquisition: importance for biomining and bioremediation. *Biol. Res.* 46, 363–371. doi: 10.4067/S0716-97602013000400008

Naz, K., Ullah, N., Zaheer, T., Shehroz, M., Naz, A., and Ali, A. (2020). *Pan-genomics of model bacteria and their outcomes. In Pan-genomics: Applications, Challenges, and Future Prospects*. Cambridge, MA: Academic Press, 189–201. doi: 10.1016/B978-0-12-817076-2.00009-3

Neira, G., Cortez, D., Jil, J., and Holmes, D. S. (2020). AciDB 1.0: a database of acidophilic organisms, their genomic information and associated metadata. *Bioinform.* 36, 4970–4971. doi: 10.1093/bioinformatics/btaa638

Nielsen, D. A., Fierer, N., Geoghegan, J. L., Gillings, M. R., Gumerov, V., Madin, J. S., et al. (2021). Aerobic bacteria and archaea tend to have larger and more versatile genomes. *Oikos* 130, 501–511. doi: 10.1111/oik.07912

Osorio, H., Mettert, E., Kiley, P., Dopson, M., Jedlicki, E., and Holmes, D. S. (2019). Identification and unusual properties of the master regulator FNR in the extreme acidophile *Acidithiobacillus ferrooxidans*. *Front. Microbiol.* 10:1642. doi: 10.3389/fmicb.2019.01642

Panja, A. S., Maiti, S., and Bandyopadhyay, B. (2020). Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. *Sci. Rep.* 10, 1–9. doi: 10.1038/s41598-020-58825-7

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

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

Pavesi, A. (2021). Origin, Evolution and Stability of Overlapping Genes in Viruses: a Systematic Review. *Genes* 12:809. doi: 10.3390/genes12060809

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.1080/13696998.2019.1666854

Quatrini, R., Appia-Ayme, C., Denis, Y., Jedlicki, E., Holmes, D. S., and Bonnefoy, V. (2009). Extending the models for iron and sulfur oxidation in the extreme acidophile *Acidithiobacillus ferrooxidans*. *BMC Genom.* 10, 1–19. doi: 10.1186/1471-2164-10-394

Remm, M., Storm, C. E., and Sonnhammer, E. L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* 314, 1041–1052. doi: 10.1006/jmbi.2000.5197

Rodríguez-Gijón, A., Nuy, J. K., Mehrshad, M., Buck, M., Schulz, F., Woyke, T., et al. (2021). A genomic perspective on genome size distribution across Earth's microbiomes reveals a tendency to gene loss. *bioRxiv* doi: 10.1101/2021.01.18.427069

Rzhepishevska, O. I, Valdés, J., Marcinkeviciene, L., Gallardo, C. A., Meskys, R., Bonnefoy, V., et al. (2007). Regulation of a novel *Acidithiobacillus caldus* gene cluster involved in metabolism of reduced inorganic sulfur compounds. *Appl. Environ. Microbiol.* 73, 7367–7372. doi: 10.1128/AEM.01497-07

Sabath, N., Ferrada, E., Barve, A., and Wagner, A. (2013). Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* 5, 966–977. doi: 10.1093/gbe/evt050

Saha, D., Panda, A., Podder, S., and Ghosh, T. C. (2015). Overlapping genes: a new strategy of thermophilic stress tolerance in prokaryotes. *Extremophiles* 19, 345–353. doi: 10.1007/s00792-014-0720-3

Sauer, D. B., and Wang, D. N. (2019). Predicting the optimal growth temperatures of prokaryotes using only genome derived features. *Bioinform.* 35, 3224–3231. doi: 10.1093/bioinformatics/btz059

Saw, J. H., Mountain, B. W., Feng, L., Omelchenko, M. V., Hou, S., Saito, J. A., et al. (2008). Encapsulated in silica: genome, proteome and physiology of the thermophilic bacterium *Anoxybacillus flavithermus* WK1. *Genome Biol.* 9:R161. doi: 10.1186/gb-2008-9-11-r161

Schneiker, S., dos Santos, V. A. M., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., et al. (2006). Genome sequence of the ubiquitous hydrocarbon-degrading

marine bacterium *Alcanivorax borkumensis*. *Nat. Biotechnol.* 24, 997–1004. doi: 10.1038/nbt1232

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., et al. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020, 1–21. doi: 10.1093/database/baaa062

Seabold, S., and Perktold, J. (2010). "Statsmodels: Econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, Vol. 57, (Scipy), 61.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinform.* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shmaryahu, A., Lefimil, C., Jedlicki, E., and Holmes, D. S. (2009). Small regulatory RNAs in *Acidithiobacillus ferrooxidans*: case studies of 6S RNA and Frr. *Adv. Mat. Res.* 71, 191–194. doi: 10.4028/www.scientific.net/AMR.71-73.191

Slonczewski, J. L., Fujisawa, M., Dopson, M., and Krulwich, T. A. (2009). Cytoplasmic pH measurement and homeostasis in bacteria and archaea. *Adv. Microb. Physiol.* 55, 1–317. doi: 10.1016/S0065-2911(09)05501-5

Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., et al. (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J.* 3, 93–105. doi: 10.1038/ismej.2008.83

Sriaporn, C., Campbell, K. A., Van Kranendonk, M. J., and Handley, K. M. (2021). Genomic adaptations enabling *Acidithiobacillus* distribution across wide-ranging hot spring temperatures and pHs. *Microbiome* 9, 1–17. doi: 10.1186/s40168-021-01090-1

Sun, Z., and Blanchard, J. L. (2014). Strong genome-wide selection early in the evolution of *Prochlorococcus* resulted in a reduced genome through the loss of a large number of small effect genes. *PLoS One* 9:3. doi: 10.1371/journal.pone.0088837

Suzuki, S., Kuenen, J. G., Schipper, K., Van Der Velde, S., Ishii, S. I, Wu, A., et al. (2014). Physiological and genomic features of highly alkaliphilic hydrogen-utilizing Betaproteobacteria from a continental serpentinizing site. *Nat. Commun.* 5, 1–12. doi: 10.1038/ncomms4900

Swan, B. K., Tupper, B., Sczyrba, A., Lauro, F. M., Martinez-Garcia, M., González, J. M., et al. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci. USA* 110, 11463–11468. doi: 10.1073/pnas.1304246110

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. USA* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Ullrich, S. R., González, C., Poehlein, A., Tischler, J. S., Daniel, R., and Schlomann, M. (2016). Gene loss and horizontal gene transfer contributed to the genome evolution of the extreme acidophile "*Ferrovum*". *Front. Microbiol.* 7:797. doi: 10.3389/fmicb.2016.00797

Ulrich, L. E., Koonin, E. V., and Zhulin, I. B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* 13, 52–56. doi: 10.1016/j.tim.2004.12.006

Vargas-Straube, M. J., Beard, S., Norambuena, R., Paradela, A., Vera, M., and Jerez, C. A. (2020). High copper concentration reduces biofilm formation in *Acidithiobacillus ferrooxidans* by decreasing production of extracellular polymeric substances and its adherence to elemental sulfur. *J. Prot.* 225:103874. doi: 10.1016/j.jprot.2020.103874

Veloso, F., Riadi, G., Aliaga, D., Lieph, R., and Holmes, D. S. (2005). Large-scale, multi-genome analysis of alternate open reading frames in bacteria and archaea. *OMICS* 9, 91–105. doi: 10.1089/omi.2005.9.91

Vergara, E., Neira, G., González, C., Cortez, D., Dopson, M., and Holmes, D. S. (2020). Evolution of Predicted Acid Resistance Mechanisms in the Extremely Acidophilic *Leptospirillum* Genus. *Genes* 11:389. doi: 10.3390/genes11040389

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

Walker, J. E. (1992). The NADH: ubiquinone oxidoreductase (complex I) of respiratory chains. *Q. Rev. Biophys.* 25, 253–324. doi: 10.1017/S003358350000425X

Westoby, M., Nielsen, D. A., Gillings, M. R., Litchman, E., Madin, J. S., Paulsen, I. T., et al. (2021). Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecol. Evol.* 11, 3956–3976. doi: 10.1002/ece3.7290

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinform* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

Zhan, Y., Yang, M., Zhang, S., Zhao, D., Duan, J., Wang, W., et al. (2019). Iron and sulfur oxidation pathways of *Acidithiobacillus ferrooxidans*. *World J. Microbiol. Biotechnol.* 35, 1–12. doi: 10.1007/s11274-019-2632-y

Zhang, X., Liu, X., Liang, Y., Fan, F., Zhang, X., and Yin, H. (2016). Metabolic diversity and adaptive mechanisms of iron-and/or sulfur-oxidizing autotrophic acidophiles in extremely acidic environments. *Environ. Microbiol. Reports* 8, 738–751. doi: 10.1111/1758-2229.12435

Zhang, X., Liu, X., Liang, Y., Guo, X., Xiao, Y., Ma, L., et al. (2017). Adaptive evolution of extreme acidophile *Sulfobacillus thermosulfidooxidans* potentially driven by horizontal gene transfer and gene loss. *Appl. Environ. Microbiol.* 83, e3098–e3016. doi: 10.1128/AEM.03098-16

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cortez, Neira, González, Vergara and Holmes. 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: 28 March 2022
doi: 10.3389/fmicb.2022.765105

Polyphasic Identification and Genomic Insights of *Leptothermofonsia sichuanensis* gen. sp. nov., a Novel Thermophilic Cyanobacteria Within Leptolyngbyaceae

Jie Tang[1], Mahfuzur R. Shah[2†], Dan Yao[1], Ying Jiang[2], Lianming Du[1], Kelei Zhao[1], Liheng Li[2], Meijin Li[2], Michal M. Waleron[3], Malgorzata Waleron[4], Krzysztof Waleron[3] and Maurycy Daroch[2]*

[1] Antibiotics Research and Re-evaluation Key Laboratory of Sichuan Province, Sichuan Industrial Institute of Antibiotics, Chengdu University, Chengdu, China, [2] School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China, [3] Department of Pharmaceutical Microbiology, Faculty of Pharmacy Medical University of Gdańsk, Gdańsk, Poland, [4] Laboratory of Plant Protection and Biotechnology, Intercollegiate Faculty of Biotechnology University of Gdańsk and Medical University of Gdańsk, University of Gdańsk, Gdańsk, Poland

OPEN ACCESS

Edited by:
Brian P. Hedlund,
University of Nevada, Las Vegas,
United States

Reviewed by:
Vartul Sangal,
Northumbria University,
United Kingdom
Gongliang Yu,
Institute of Hydrobiology (CAS), China
Janaina Rigonato,
Genoscope, France

***Correspondence:**
Maurycy Daroch
m.daroch@pkusz.edu.cn

†Present address:
Mahfuzur R. Shah,
Department of Cell Biology, Metabolism and Systems Biology, Noblegen Inc., Peterborough, ON, Canada

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

Received: 26 August 2021
Accepted: 31 January 2022
Published: 28 March 2022

Citation:
Tang J, Shah MR, Yao D, Jiang Y, Du L, Zhao K, Li L, Li M, Waleron MM, Waleron M, Waleron K and Daroch M (2022) Polyphasic Identification and Genomic Insights of Leptothermofonsia sichuanensis gen. sp. nov., a Novel Thermophilic Cyanobacteria Within Leptolyngbyaceae. Front. Microbiol. 13:765105. doi: 10.3389/fmicb.2022.765105

Thermal environments are an important reservoir of thermophiles with significant ecological and biotechnological potentials. However, thermophilic isolates remain largely unrecovered from their habitats and are rarely systematically identified. In this study, we characterized using polyphasic approaches a thermophilic strain, PKUAC-SCTAE412 (E412 hereafter), recovered from Lotus Lake hot spring based in Ganzi prefecture, China. The results of 16S rRNA/16S-23S ITS phylogenies, secondary structure, and morphology comparison strongly supported that strain E412 represent a novel genus within Leptolyngbyaceae. This delineation was further confirmed by genome-based analyses [phylogenomic inference, average nucleotide/amino-acid identity, and the percentages of conserved proteins (POCP)]. Based on the botanical code, the isolate is herein delineated as *Leptothermofonsia sichuanensis* gen. sp. nov, a genus adjacent to recently delineated *Kovacikia* and *Stenomitos*. In addition, we successfully obtained the first complete genome of this new genus. Genomic analysis revealed its adaptations to the adverse hot spring environment and extensive molecular components related to mobile genetic elements, photosynthesis, and nitrogen metabolism. Moreover, the strain was capable of modifying the composition of its light-harvesting apparatus depending on the wavelength and photoperiod, showing chromatic adaptation capacity characteristic for T1 and T2 pigmentation types. Other physiological studies showed the strain's ability to utilize sodium bicarbonate and various sulfur compounds. The strain was also shown to be diazotrophic. Interestingly, 24.6% of annotated protein-coding genes in the E412 genome were identified as putatively acquired, hypothesizing that a large number of genes acquired through HGT might contribute to the genome expansion

and habitat adaptation of those thermophilic strains. Most the HGT candidates (69.4%) were categorized as metabolic functions as suggested by the KEGG analysis. Overall, the complete genome of strain E412 provides the first insight into the genomic feature of the genus *Leptothermofonsia* and lays the foundation for future global ecogenomic and geogenomic studies.

Keywords: 16S rRNA, 16S-23S ITS, genomics, thermophilic cyanobacterium, Leptolyngbyaceae, *Kovacikia*, *Stenomitos*

INTRODUCTION

Thermophilic cyanobacteria are photoautotrophic prokaryotes that inhabit inhospitable niches, such as hot springs and other thermal environments. These thermophiles are essentially primary producers of the geothermal ecosystems with substantial ecological importance (Esteves-Ferreira et al., 2018; Li et al., 2021). Besides, thermophilic cyanobacteria and their secondary metabolites have shown some potentials for biotechnological applications, including biosynthesis of thermostable metabolites and carbon valorisation vehicles (Patel et al., 2019). Unfortunately, thermophilic isolates remain largely unrecovered from their habitats to date and lack comprehensive knowledge about their physiology, ecology, systematics, and adaptations to the adverse thermal habitats.

Recently, next-generation sequencing (NGS) has been extensively applied to explore the cyanobacterial diversity in thermal environments (Tang et al., 2018b; Alcorta et al., 2020; Chen et al., 2021). However, NGS only generates ample abstract data and its assembly often results in unspecified operational taxonomic units (OTUs) and higher taxonomic groups. Isolation of thermophilic cyanobacteria from different ecosystems is fundamental for detailed characterization of their morphology, genetics, physiology, and biochemistry and for a better understanding of thermal ecology and their adaptation to thermal habitats (Cordeiro et al., 2020; Tang et al., 2022). However, the simple morphology of cultivated cyanobacteria makes these strains ambiguous and their taxonomic allocation challenging (Komárek and Anagnostidis, 2005). To address this problem, polyphasic taxonomic classification approaches have been widely demonstrated to be effective for cyanobacterial identification, especially for understudied or unresolved polyphyletic families/genera/species and identification of novel families and genera (Raabova et al., 2019; Shalygin et al., 2020). Consequently, establishing new genera as references will benefit the reallocation of numerous taxonomically ambiguous strains. For instance, *Leptolyngbya*-like strains have been scattered across phylogenies based on 16S rRNA or multi-locus sequence analysis within or even beyond family Leptolyngbyaceae (Sciuto and Moro, 2016; Yao et al., 2021). As the taxonomic delineation of isolates through systematic identification, numerous *Leptolyngbya*-like strains were assigned to new genus or species, such as *Allonema*, *Enugrolinea,* and *Thermoleptolyngbya* (Sciuto and Moro, 2016; Walter et al., 2017).

Genomic studies on thermophilic cyanobacteria are increasingly popular, and the better availability of complete genome sequences from thermal isolates and Metagenome-Assembled Genomes (MAGs) allows the studies of taxonomic relationships and genomic adaptations to different ecological niches among thermophilic strains and provides insight into survival strategies in extreme conditions not fully comprehended up to now (Chen et al., 2021). In addition, the acquisition of a complete genome may provide novel insights into the genomic features of new thermophilic microorganisms. More importantly, under the current scenario of global warming, it is important to understand the evolution of hot spring genomes as an example of selective pressure in warmer environments (Alcorta et al., 2020). Therefore, the acquisition of genomes from thermophilic cyanobacteria as much as possible is a prerequisite for such studies.

In the current study, we collected and analyzed morphological, physiological, and molecular data for the representative of an entirely novel genus of a filamentous thermophilic cyanobacterium isolated from a Lotus Lake hot spring based in Ganzi prefecture western Sichuan province of China. After thorough taxogenomic evaluation combined with 16S rRNA/16S-23S ITS phylogenies, secondary structure, and morphology comparison, a new name *Leptothermofonsia sichuanensis* gen. sp. nov. has been proposed based on the botanical code for the strain as the first representative of genus *Leptothermofonsia*. Furthermore, genomic features of this genus were studied, and some genes were correlated with the physiological properties.

MATERIALS AND METHODS

Leptothermofonsia sichuanensis gen. sp. nov. E412: Origins, Cultivation, and Basic Physiological Assessment

The strain E412 presented in the current study was initially isolated from Lotus Lake hot springs in the Ganzi region of Sichuan Province, China. Information about the sampling site and preliminary taxonomic allocation of the strain was reported in our previous studies (Tang et al., 2018a,b). The unicyanobacterial culture of the strain was cryopreserved as 10% DMSO in BG11 stocks in −80°C. Final precultures for experiments were established as previously described (Tang et al., 2018b) and cultivated at 45°C in 150 mL BG-11 medium in 500 mL Erlenmeyer flasks agitated at 100 rpm under 12L:12D photoperiod at 45 μmol m^{-2} s^{-1} provided by fluorescent tubes unless stated otherwise. The strain initially denoted as

PKUAC-SCTE412 has been recently deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with an accession number FACHB-2490. Physiological assessment of the strain was performed using BG-11 medium free of an essential nutrient (nitrogen or sulfur) supplemented with 17 mM $NaNO_2$, 85 mM $NaNO_3$, 10 mM Na_2SO_4, and 10 mM $NaHSO_3$. The nitrogen fixation capacity during 72 h was performed using the methodology described by Li et al. (2021).

The chromatic adaptation capacity of strain E412 was assessed by culturing the cells at constant LED illumination using either white light (6,500 K) or far-red light (730 nm) at the intensity of 250 and 25 μmol m^{-2} s^{-1}, respectively. After 15 days, cyanobacterial cells were collected to measure chlorophyll a, chlorophyll b, carotenoids, allophycocyanin, phycocyanin, and phycoerythrin, according to previously published protocols (Bennett and Bogorad, 1973; Dere et al., 1998). Lipophilic and water-soluble pigments from 15 mg of lyophilized biomass were extracted with 100% methanol and phosphate-buffered saline (PBS), respectively. The absorbance of the supernatant at 470, 562, 615, 652, 653, and 666 nm were recorded against the relative blank using a UV-Vis spectrophotometer (Shimadzu UV-1,800, Japan). The concentration of pigments was calculated using respective formulae for chlorophylls and carotenoids (Dere et al., 1998) and phycobiliproteins (Bennett and Bogorad, 1973).

Genome Sequencing, Assembly, and Annotation of Strain E412

Genomic DNA was extracted and assessed following the previously described method (Tang et al., 2018a). The whole genome of E412 was sequenced using a combination of Oxford Nanopore Technologies (ONT) and Illumina short-read approaches. Illumina sequencing of E412 genome generated 6,810,074 filtered paired-end reads (clean data), providing approximately 185-fold coverage of the genome. The clean data were assembled into contigs using MaSuRCA v. 3.4.1 with default parameters (Zimin et al., 2013) and Flye for final assembly, generating a single contig after manual curation. The draft circular genome was error-corrected with Illumina NovaSeq reads using Burrows-Wheeler Aligner, BWA v0.7.17 (Li and Durbin, 2009), and then Pilon v1.23 (Walker et al., 2014). The complete genome has been deposited in GenBank with an accession number CP072600.

The annotation of the E412 genome was performed using a pipeline described by Tang et al. (2019). In short, initial gene prediction and annotation were completed using the automatic NCBI prokaryotic genome annotation pipeline (Tatiana et al., 2016). Subsequently, poor calls were corrected using the RAST annotation system. The insertion sequence (IS) was detected and annotated by ISsaga (Varani et al., 2011). Prophage regions and CRISPR/Cas loci were detected by PHASTER (David et al., 2016) and by CRISPRCasFinder server (Couvin et al., 2018), respectively.

For functional classification, the predicted protein sequences were searched against the NCBI non-redundant database using BLASTP with an *E*-value cutoff of 1e-5. The search outputs were imported into BLAST2GO V5.2.5 (Conesa et al., 2005) for GO term mapping. GO classification was subsequently conducted by WEGO (Ye et al., 2006) using the following ontologies: biological process, molecular function, and cellular component. Additionally, KEGG orthology (KO) identifiers were assigned to predicted protein sequences by KofamKOALA (Aramaki et al., 2019), and then BRITE mapping process was performed online.[1]

Phylogenetic Reconstruction of 16S rRNA and 16S-23S ITS

The 16S rRNA gene and 16S-23S ITS regions were extracted from the E412 for phylogenetic analysis. Cyanobacterial sequences were retrieved from GenBank through BLAST search as references to construct datasets for phylogenetic analyses of 16S rRNA gene and 16S-23S ITS, respectively. Multiple alignments of sequences were performed by Muscle implemented in Mega7 (Kumar et al., 2016) and manually edited where necessary. Maximum-Likelihood (ML) phylogenetic analyses were carried out using PhyML v3.0 (Guindon et al., 2010), and the substitution models were selected by the Model Selection function implemented in PhyML (Vincent et al., 2017) under Akaike information criterion (AIC). A non-parametric bootstrap test (1,000 replications) was performed to assess the robustness of tree topologies.

Prediction of Secondary Structures

The conserved domains of the 16S-23S ITS region: D1-D1′, D2, D3, boxA, and D4; and its variable regions (V2, boxB, and V3) were identified as previously described (Iteman et al., 2000). The tRNAs presented in the spacer were identified by tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997). The secondary structures of the identified fragments were individually determined by RNAstructure web server using default settings (Mathews, 2014).

Microscopic Analysis

Strain E412 was investigated at 400 × magnification using light microscopy (LM, DP72, OLYMPUS, Japan), equipped with an image acquisition system (U-TV0.63XC, OLYMPUS, Japan). Microscopic observations were also performed using scanning electron microscopy (SEM) (SU8100, HITACHI, Japan), and using transmission electron microscopy (TEM) (HT7800, HITACHI, Japan), essentially as described before (Tang et al., 2022).

Genome-Based Analysis

To compare divergence in genomes of focal taxa (E412 and previously established genera within Leptolyngbyaceae), a high-quality dataset comprising genomes with near completeness ($\geq$ 90%) and low contamination (<5%) was constructed to calculate whole-genome average nucleotide identity (ANI) and average amino acid identity (AAI) using the ANI/AAI calculator with default settings.[2] In addition, the percentages of conserved proteins (POCP) between the E412 genome and focal taxa were pairwise calculated to estimate their evolutionary and phenotypic distance. Finally, the POCP was determined for prokaryotic

[1] https://www.genome.jp/kegg/ko.html

[2] http://enve-omics.ce.gatech.edu/g-matrix/

genus delineation according to the method described previously (Qin et al., 2014).

The concatenated sequences from single-copy genes shared by all the genomes were used to elucidate the phylogenomic relationship between E412 and focal taxa. Single-copy genes shared by all genomes were refined from the homologous gene clusters identified by OrthoMCL (Li et al., 2003) and concatenated using custom Perl script. MAFFT v7.453 (Standley, 2013) was used to generate multisequence alignment. The supergene alignment was subjected to phylogenomic inference using IQ-TREE v2.1.3 (Minh et al., 2020). ModelFinder implemented in IQ-TREE was employed to select the optimal substitution model for phylogenomic analysis from 546 protein models. Bootstrap tests (1,000 replicates) were performed to assess tree topologies using UltraFast Bootstrap (Hoang et al., 2018). A strain from the family Oculatellaceae, *Thermoleptolyngbya sichuanensis,* was rooted as an outgroup.

Taxonomic Evaluation

For a valid description, the classification system applied was based on Komárek et al. (2014), and the taxon description follows the prescriptions of the Botanical Code, International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen code) (Turland et al., 2018).

Identification of Horizontal Gene Transfer Candidates

A BLASTP-based approach was employed to identify potential genes acquired through horizontal gene transfer (HGT). BLASTP searches were performed against the NCBI non-redundant protein database (last accessed January 20, 2018). The pairs that accounted for at least 90% of the query length and amino acid sequence similarity of at least 40% were used for the HGT detection. Taxonomic classification was assigned to each hit with the taxonomy files downloaded from the NCBI database. Protein sequences were identified as HGT candidates if all the top-five hits were not from the family Leptolyngbyaceae.

RESULTS AND DISCUSSION

Phylogeny of 16S rRNA Gene

The ML phylogram (**Figure 1**) inferred from 16S rRNA gene sequences recognized 17 well-supported clades of previously established genera: *Alkalinema* (Vaz et al., 2015), *Arthronema* (Komárek and Lukavský, 1988), *Chroakolemma* (Becerra-Absaln et al., 2018), *Kovacikia* (Miscoe et al., 2016), *Leptodesmis* (Raabova et al., 2019), *Leptolyngbya sensu stricto* (Taton et al., 2010), *Limnolyngbya* (Li and Li, 2016), *Myxacorys* (Soares et al., 2019), *Neosynechococcus* (Dvorak et al., 2014), *Onodrimia* (Jahodářová et al., 2017), *Pantanalinema* (Vaz et al., 2015), *Phormidesmis* (Raabova et al., 2019), *Pinocchia* (Dvorak et al., 2015), *Planktolyngbya* (Thomazeau et al., 2010), *Plectolyngbya* (Taton et al., 2011), *Scytolyngbya* (Song et al., 2015), *Stenomitos* (Miscoe et al., 2016), and *Gloeobacter* as an outgroup of the phylogram. Strain E412, together with *Leptolyngbya* sp. Greenland 10, formed a well-defined clade that was phylogenetically novel to the other described taxa (**Figure 1**), suggesting a new genus within the family Leptolyngbyaceae.

The sequence identity (**Supplementary Table 1**) indicated that the nearest neighbor of E412 was *Leptolyngbya* sp. Greenland 10 (96.2% identity), a poorly described strain isolated from the hot spring (56–61°C), Rømerfjord, Greenland (Roeselers et al., 2007), followed by *Kovacikia* (94.9–95.0% identity) and *Pantanalinema* (94.0% identity). The other focal taxa exhibited 89.3–93.9% identities of 16S rRNA gene to strain E412 (**Supplementary Table 1**). According to the recommended threshold of 16S rRNA gene identity for bacterial species (98–99%) or genera (94.5–95%) demarcation (Rodriguez-R et al., 2018), strain E412 was proposed to be categorized into a new genus within the family Leptolyngbyaceae. And *Leptolyngbya* sp. Greenland 10 was also assigned to this new genus but a different species based on the 16S rRNA gene identity. This was also supported by the ML topology of the 16S rRNA gene that Greenland 10 formed a distinct sister branch alongside strain E412 (**Figure 1**). Interestingly, the two strains within the proposed genus were both isolated from thermal habitats, but the isolation sources were geographically divergent and exhibited distinct geochemical characteristics (Roeselers et al., 2007; Tang et al., 2018a). Notably, the newly established genus offers a solid basis for further investigations or comparisons on numerous aspects, such as whether the proposed genus has a cosmopolitan distribution and all the strains within this genus are specific to thermal habitat.

Phylogeny of 16S-23S ITS

Additional to the 16S rRNA gene, the 16S-23S ITS region has also been commonly recommended to establish cyanobacterial ecotypes or species (Brito et al., 2017). Several genera included in the 16S rRNA phylogram were excluded in the ITS phylogenetic analysis due to sequence unavailability. As shown in **Figure 2**, strain E412 was placed into a well-separated branch in the ITS phylogram. Previously described genera were also recognized in the 16S-23S ITS phylogram and supported by robust bootstrap values. However, the ITS phylogram (**Figure 2**) showed inconsistent topology to that of the 16S rRNA gene (**Figure 1**). Previous reports have manifested that the sequences of 16S-23S ITS were extraordinarily divergent (Johansen et al., 2011; Tang et al., 2021). Furthermore, regions/domains within ITS were occasionally absent in some cyanobacterial strains, e.g., in the present study, conserved tRNAs were absent in *Pantanalinema rosaneae* (Vaz et al., 2015). Taken together, phylogenetic inference of ITS could result in erroneous taxonomic classification. Therefore, secondary structure analyses of domains in 16S-23S ITS were also carried out as an essential complement to the final taxonomy determination.

16S-23S ITS Secondary Structures

Strain E412 and representative strains of 13 focal genera within the family Leptolyngbyaceae were used for 16S-23S ITS secondary structure analysis. After removing the two highly conserved tRNAs (Ile and Ala) from ITS sequences,

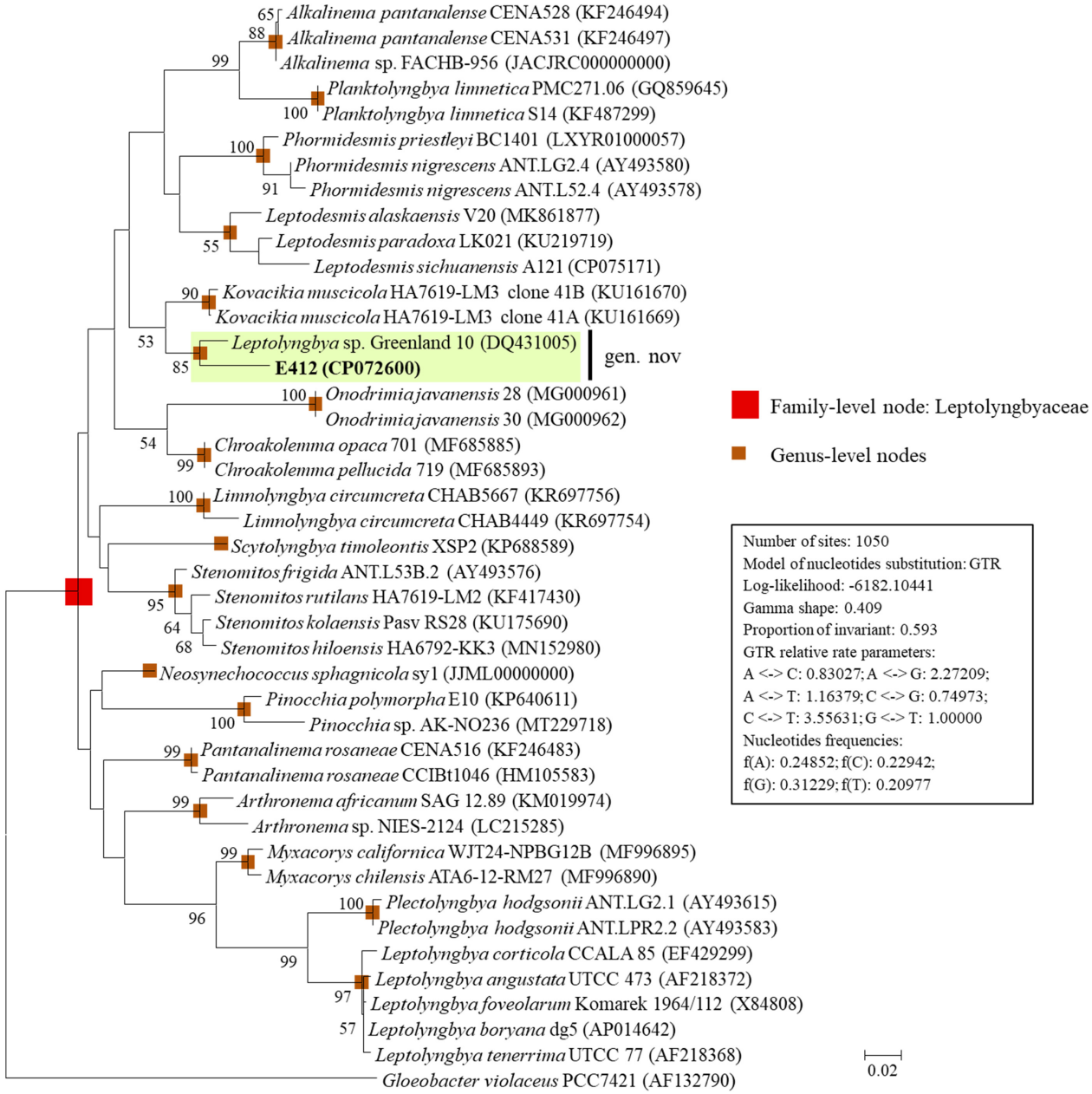


FIGURE 1 | ML phylogenetic tree of 16S rRNA gene sequences. Strain no. in bold represent the strains identified in this study. Only bootstrap values > 50% (1,000 non-parametric replications) are indicated at nodes. Scale bar = 2% substitutions per site.

the remaining ITS sequences varied in length from 230 to 388 bp (**Table 1**). Such variation was primarily ascribed to the length differences in D1-D1′ (51–121 bp), V2 (7–90 bp), boxB (33–63 bp), and V3 (19–98 bp (**Table 1**). Consequently, the nucleotide differences of these domains among strains resulted in divergences of the secondary structures.

The inferred D1-D1′ helix of strain E412 (**Supplementary Figure 1**) was distinct from the other inferred structures. The most similar structure to D1-D1′ helix of strain E412 was that of *O. javanensis*. Both D1-D1′ helix of the two strains were mainly composed of a stem fragmented by asymmetrical/symmetrical loop, right/left bulge, and terminated with hairpin loop. But the structures of the two strains varied in the residue size and/or the number of loops, bulges, and terminal hairpins. In addition, the fragmented stems also differed in number and length between the two strains.

The hypothetical V2 helix of strain E412 (**Supplementary Figure 2**) was divergent from the other analyzed strains. The tremendous length variations were responsible for the divergence in secondary structures. The V2 helix of strain E412 was

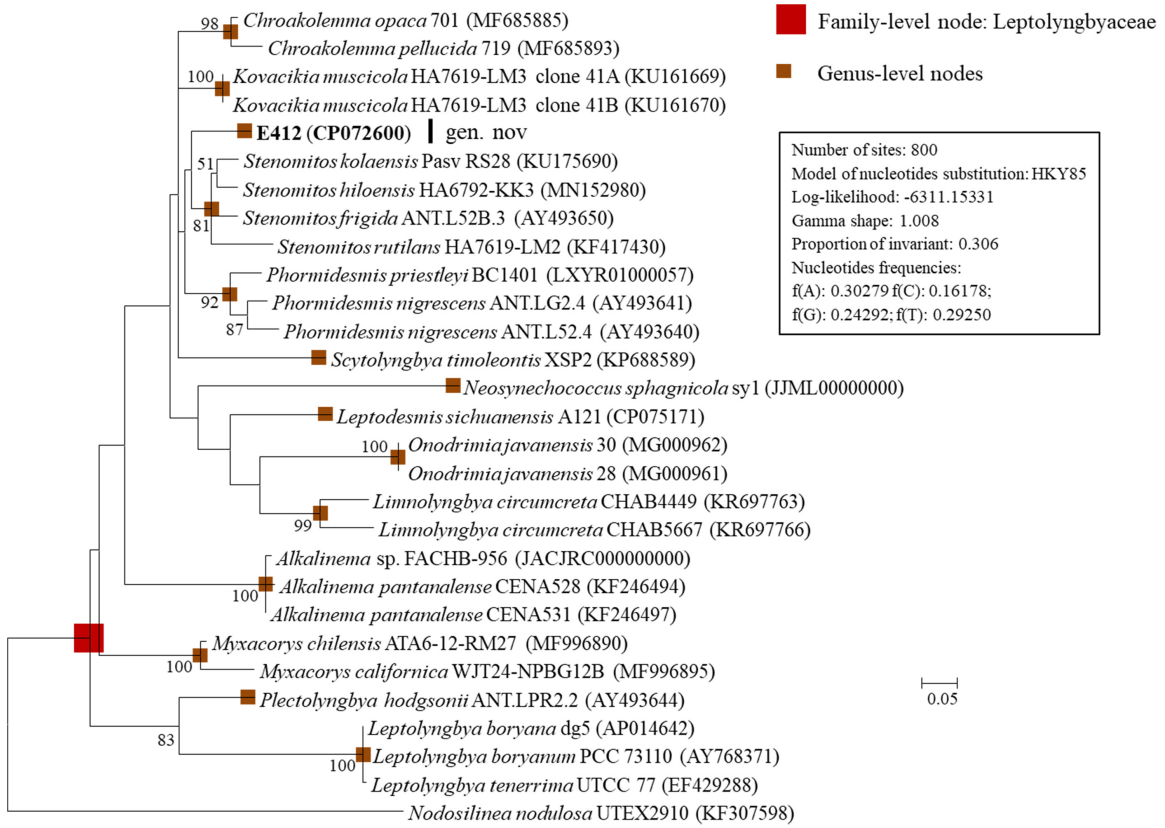


FIGURE 2 | ML phylogenetic tree of 16S-23S ITS sequences. Strain no. in bold represent the strains identified in this study. Only bootstrap values > 50% (1,000 non-parametric replications) are indicated at nodes. Scale bar = 5% substitutions per site.

TABLE 1 | The length (bp) summary of regions within16S-23S ITS of Leptolyngbyaceae strains studied.

Strain	ITS length (tRNA removed)	D1-D1′ helix	D2	D3	$tRNA^{Ile}$	boxA	D4	V2 helix	$tRNA^{Ala}$	boxB helix	V3 helix
E412	380	121	12	5	74	12	7	76	73	45	98
Alkalinema pantanalense CENA528	296	64	12	5	74	12	7	24	73	48	54
Chroakolemma pellucida 719	268	61	12	5	74	12	7	16	73	41	53
Kovacikia muscicola HA7619-LM3 clone 41A	345	63	12	5	74	12	7	90	73	41	95
Leptodesmis sichuanensis A121	325	63	12	5	74	12	7	81	73	33	98
Leptolyngbya boryanum PCC 73110	275	51	12	5	74	12	7	10	73	33	21
Limnolyngbya circumcreta CHAB5667	388	98	12	5	74	12	7	83	73	63	76
Myxacorys californica WJT24-NPBG12B	258	86	12	5	74	12	7	9	73	33	71
Neosynechococcus sphagnicola sy1	230	63	12	5	74	12	7	11	73	39	95
Onodrimia javanensis 28	280	105	12	5	74	12	7	7	73	44	47
Phormidesmis priestleyi ANT.L52.4	329	113	12	5	74	12	7	12	73	56	77
Plectolyngbya hodgsonii ANT.LPR2.2	306	55	12	5	74	12	7	29	73	44	19
Scytolyngbya timoleontis XSP2	276	64	12	5	74	12	7	14	73	40	94
Stenomitos rutilans HA7619-LM2	258	65	12	5	74	12	7	7	73	34	92

straightforward and comprised 5 stems, 4 symmetrical loops, and an 8-residue hairpin loop. Although *K. muscicola*, *Leptodesmis* sp., and *L. circumcreta* possessed similar sequence lengths with respect to the strain E412, the helices differed in the stem, loop, bulge, and hairpin.

A basal stem structure (AGCA-UGCU) was shared by boxB helices of all strains except for *A. pantanalense* (**Supplementary Figure 3**). Strain E412 showed similar residue length to several strains, but no boxB helix structures were analogous to that of strain E412. The boxB helix of strain E412 was mainly composed

of a stem orderly fragmented by single base right bulge, single base left bulge, 10-residue symmetrical loop, and terminated with a 5-residue hairpin loop.

Strain E412 exhibited the longest V3 helix (98 residues) (**Table 1**), comprising 3 asymmetrical loops, 3 symmetrical loops, a single base left bulge, an 8-residue hairpin loop, and fragmented stems (**Supplementary Figure 4**). The V3 helix of strain E412 was distinct from those of the other strains, whereas a basal stem structure (GUC-GAC) was shared by all the strains. Although the helix length of *K. muscicola* (95 residues), *N. sphagnicola* (95 residues), *S. timoleontis* (94 residues), and *S. rutilans* (92 residues) was similar to that of strain E412, the structures differed from each other in terms of bulge, loop, hairpin, and stem.

Conclusively, the result of 16S-23S ITS secondary structure analysis and the phylogenetic inferences of 16S rRNA and 16S-23S ITS verified strain E412 as a new genus within the family Leptolyngbyaceae. Our results (**Supplementary Figures 1–4**) also demonstrated that the secondary structure analysis of D1-D1′, V2, boxB, and V3 helix was an effective tool for genus-level identification within the family Leptolyngbyaceae.

Genome-Based Analyses

Based on the genomes available on a public database, comparison analyses at the genomic level were performed among genera within Leptolyngbyaceae. Considerable divergences in genomes were observed among different genera as revealed by the ANI and AAI values (**Table 2**). Particularly, the ANI and AAI values between strain E412 and the other eight focal taxa were less than 79 and 69%, respectively. The results of genome-wide ANI and AAI conformed to the suggested values for genus (ANI $<$ 83%, AAI $\leq$ 70%) delimitation (Walter et al., 2017; Jain and Rodriguez, 2018), further confirming the taxonomy delineation of a novel genus within the family Leptolyngbyaceae. However, it was reported that the classification of the prokaryotic genus using ANI or AAI might cause misleading results in some cases (Konstantinidis and Tiedje, 2005; Pannekoek et al., 2016). Therefore, the POCP specific for genus delineation was calculated to further verify the genus demarcation of strain E412. The POCP values (**Table 3**) between the E412 genome and focal taxa ranged from 35.7 to 49.3%, all within the threshold (<50%) for the definition of a prokaryotic genus (Qin et al., 2014). Additionally, the ML genomic phylogram (**Figure 3**) generated from the concatenated alignment of 845 single-copy genes showed a consistent topology to that of 16S rRNA, and again verified the conclusion that strain E412 belonged to a novel genus within Leptolyngbyaceae. Unfortunately, only some of the genera related to E412 had their genomes sequenced, resulting in a partial snapshot of genomic divergences among these organisms.

TABLE 3 | POCP values between strain E412 and representative species from Leptolyngbyaceae genera.

Species	E412
Alkalinema sp. FACHB-956	43.8%
Leptodesmis sichuanensis A121	48.5%
Leptolyngbya boryana dg5	44.6%
Myxacorys almedinensis A	45.2%
Neosynechococcus sphagnicola sy1	35.7%
Pantanalinema sp. GBBB05	49.3%
Phormidesmis priestleyi BC1401	44.1%
Stenomitos frigidus ULC18	47.6%

Morphological Characteristics of Strain E412

Analysis of strain E412 with light microscopy revealed that trichomes were brown and exhibited coiled and tangled morphology (**Figure 4A**). The SEM and TEM analysis indicated unbranched trichomes composed of elongated cylindrical shaped cells, 1.2–1.8 μm in length and 0.8–1.1 μm in width. Constrictions were detected at the cross-walls of cells (**Figures 4B,C**). Centripetal invaginations of the cell wall separated individual cells of the filaments, but the intracellular connections between individual vegetative cells were not present (**Figure 4C**). The analysis of TEM micrographs indicated that five to six thylakoids in the parietal arrangement were present at the inner periphery of cells. Additionally, typical components of filamentous cyanobacteria, i.e., sheath, septum, phycobilisomes, and carboxysomes, were identified. The strain was also identified as a cyanophycin producer, and granules of this biopolymer were observed in the cytoplasm. Polyphosphate bodies and small lipid droplets were also identified (**Figures 4C,D**). Morphological comparison of strain E412 against other focus taxa from the

TABLE 2 | Values of ANI (Average Nucleotide Identity) and AAI (Average Amino acid Identity) among Leptolyngbyaceae genomes studied.

Strain	E412	FACHB-956	A121	dg5	A	sy1	GBBB05	BC1401	ULC18
E412	*100.00*	60.87	68.49	62.45	63.54	61.83	67.45	63.93	67.86
Alkalinema sp. FACHB-956	75.64	*100.00*	60.97	61.62	62.35	57.98	60.93	62.31	59.56
Leptodesmis sichuanensis A121	78.79	75.72	*100.00*	62.61	63.55	62.09	68.21	64.29	66.56
Leptolyngbya boryana dg5	77.29	77.65	77.02	*100.00*	68.15	59.06	62.46	66.71	61.70
Myxacorys almedinensis A	74.40	78.58	72.91	74.88	*100.00*	59.58	63.12	67.93	62.91
Neosynechococcus sphagnicola sy1	75.81	75.63	76.83	76.60	71.80	*100.00*	61.92	60.79	62.14
Pantanalinema sp. GBBB05	74.40	74.49	75.41	77.67	74.84	73.88	*100.00*	64.33	66.26
Phormidesmis priestleyi BC1401	75.41	74.34	75.42	75.05	74.46	76.76	74.53	*100.00*	65.20
Stenomitos frigidus ULC18	74.46	73.53	73.79	74.59	73.60	75.81	74.91	77.97	*100.00*

The numbers above and below the diagonal indicate the AAI and ANI values (%), respectively.

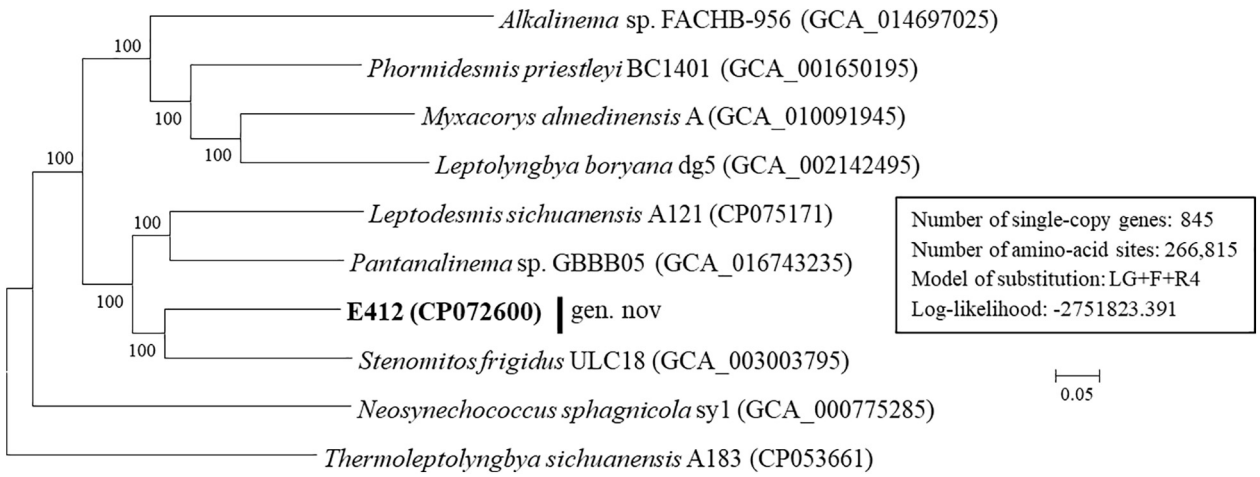


FIGURE 3 | ML phylogenomic tree of concatenated protein alignment of single-copy genes shared by all genomes. Strain no. in bold represent the strain identified in this study. Bootstrap values (1,000 replications) are indicated at nodes. Scale bar = 5% substitutions per site.

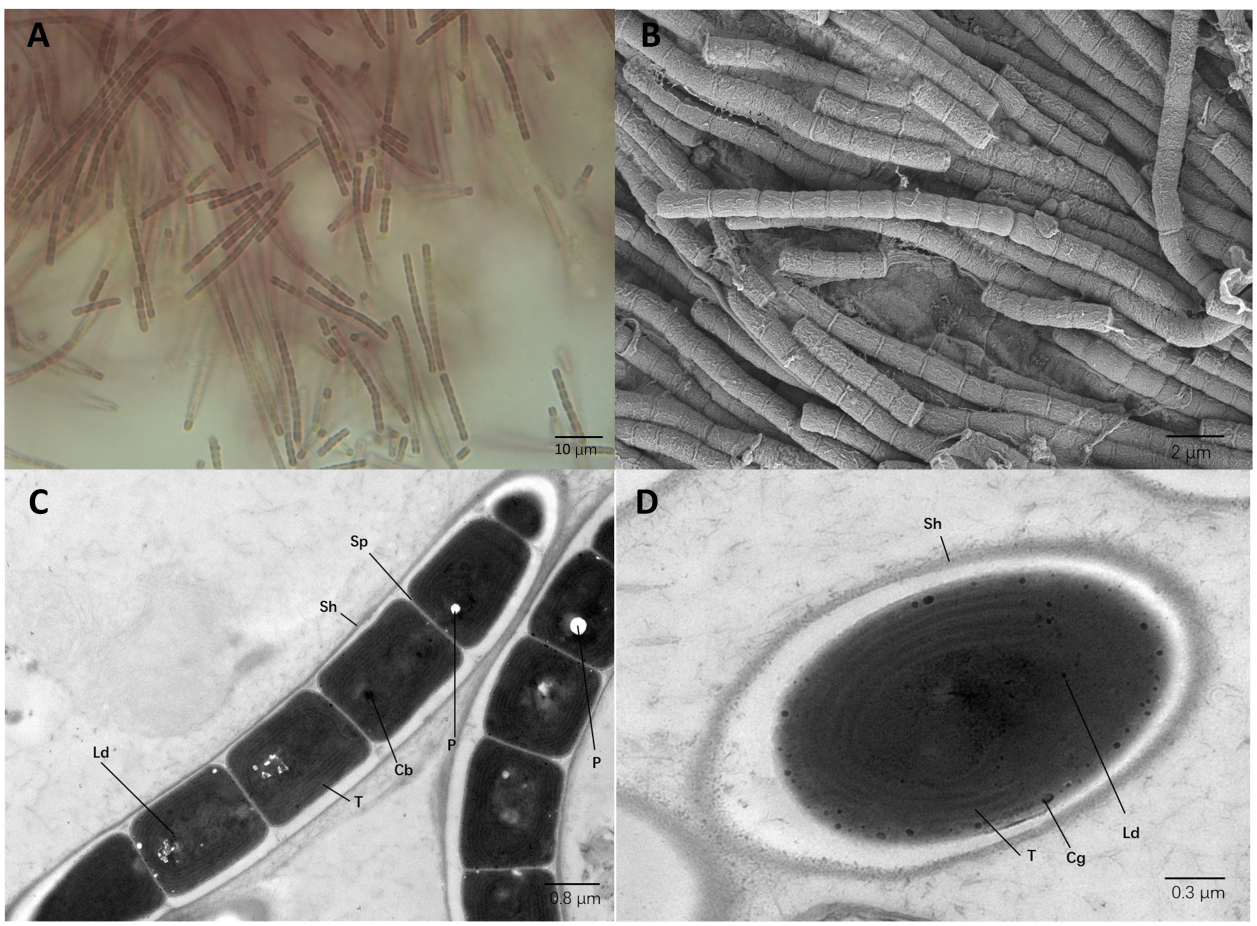


FIGURE 4 | Micrographs of strain E412. **(A)** Light microscopy image. **(B)** SEM image. **(C,D)** TEM images. Cb, carboxysome; Cg, cyanophycin granule; Ld, lipid droplet; P, polyphosphate body; Sh, sheath; Sp, septum; T, thylakoid membrane. Magnifications were 1,000× **(A)**, 5,000× **(B)**, 8,000× **(C)**, and 12,000× **(D)**.

TABLE 4 | Comparison of morphological features of Leptolyngbyaceae strains.

Species	Morphology	Cell width (μm)	Cell length (μm)	Sheaths	Thylakoids No.	Color	References
E412	Coiled, tangled, closely packed	0.8–1.1	1.2–1.8	Colorless	5–6	Brown or green depending on conditions	This study
Alkalinema pantanalense CENA528	Entangled, flexuous	1.7–2.2	2.0–4.1	-	NA	Reddish or brownish	Vaz et al., 2015
Kovacikia muscicola HA7619-LM3 clone 41A	Straight, unbranched	1.0–1.4	1.0–1.4	Thin, colorless	NA	Brownish	Miscoe et al., 2016
Leptodesmis paradoxa LK021	Straight, curved, flexuous or wavy, solitary	2.5–3.5	1.0–1.5	Thick, colorless	NA	Pale blue-green	Raabova et al., 2019
Phormidesmis nigrescens LK013	Straight, curved, solitary	1.5–2.50	1.0–2.0	Thick, blackish	+	Blackish	Raabova et al., 2019
Stenomitos rutilans HA7619-LM2	Bent, entangled	0.9–1.1	2.5–5.0	-	NA	Red brownish	Miscoe et al., 2016

NA, not available; +/−, presence/absence.

family Leptolyngbyaceae (**Table 4**) revealed the strain's closest morphological resemblance to *Stenomitos* and *Kovacikia*, in agreement with the phylogenetic allocation. All the three strains had single unbranched filaments, often entangled. Individual cells were approximately 1 μm in width, cylindrical, and slightly elongated. The cells of *Leptothermofonsia* and *Kovacikia* were typically not longer than 1.8 μm, as opposed to longer cells of *Stenomitos*. All the three strains were brown in the natural environment. Strain E412 can adapt to the green phenotype with the modification of light intensity and photoperiod. No such information was available for other strains. Among other focus taxa, only *Leptodesmis* was distinctly green and shown to be incapable of chromatic adaptation. Available morphological characteristics of focus taxa was collected and summarized in **Table 4**. Although strain E412 was also morphologically similar to *Pseudanabaena* spp. (Komárek and Anagnostidis, 2005), they were phylogenetically divergent from each other (**Supplementary Figure 5**).

Physiological Characteristics of Strain E412

Strain E412 was physiologically characterized using different variants of the standard growth medium, BG-11. The strain exhibited active growth when sodium bicarbonate at the concentration ranging from 0.1 to 0.5 M was applied, suggesting that this form of inorganic carbon can be effectively utilized by the strain. Analysis of the impact of sulfur compounds showed that the strain was capable of utilizing 10 mM sulfates and incapable of utilizing equivalent concentrations of sulphites. The strain exhibited flexibility in the utilization of nitrogen sources. Both nitrate (3 mM) and nitrite (6 mM) resulted in the active growth of the strain. Strain E412 was also diazotrophic. The acetylene reduction assay indicated a functional nitrogenase capable of a steady increase of ethylene production during 72 h of the assay (**Supplementary Figure 6**).

Analysis of the pigment composition of strain E412 under different illumination conditions suggested a significant difference in their composition at white light and far-red light illumination and across the different photoperiods (**Figure 5**). The strain cultivated at 12L:12D photoperiod in low light exhibited a brown phenotype and pigmentation typical of type T2 (Six et al., 2007) due to increased phycoerythrin accumulation (**Figures 5A,D**). Meanwhile, cultivation in high light and far-red light under constant illumination resulted in the recomposition of the photosynthetic apparatus and pigmentation type T1 (**Figures 5B,D**). Therefore, the strain can adjust its photosynthetic apparatus through chromatic adaptation and was capable of far-red light (730 nm) utilization like other thermophilic filamentous strains (Gan and Bryant, 2015). Meanwhile, the concentration of chlorophyll b increased in the cells grown in constant white light illumination (**Figure 5C**), whereas all the other pigments were relatively constant across the tested cultivation conditions (**Figures 5C,D**).

General Genomic Features of Strain E412

The combined utilization of ONT and Illumina sequencing systems resulted in the complete genome of strain E412. This genome consisted of a single circular chromosome with a size of 6,426,061 bp (GC content, 50.8%). Two ribosomal RNA (*rrn*) operons, 74 tRNA genes and 7,904 protein-coding sequences (CDS), were predicted in the E412 chromosome (**Supplementary Table 2**). Notably, 4,015 out of 7,904 (50.8%) protein-coding genes were annotated as hypothetical proteins. Identifying such a high percentage of hypothetical protein in the E412 genome was not surprising and common to the genomes of thermophilic cyanobacteria (Cheng et al., 2020; Tang et al., 2021, 2022).

The GO analysis indicated that the CDS identified in the E412 genome were assigned to a wide range of functional categories (**Supplementary Figure 7**). The majority of GO terms were assigned to the biological process. Membrane, catalytic activity, and metabolic process were the most abundant

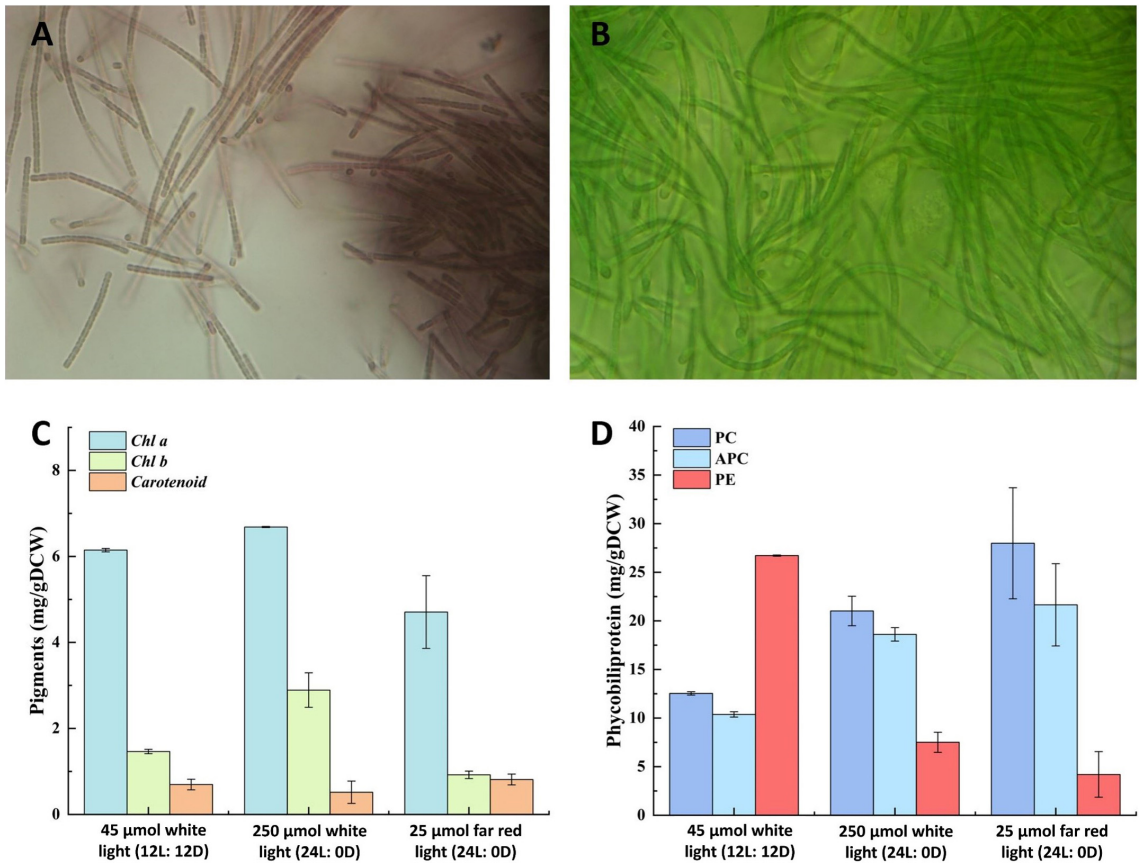


FIGURE 5 | Chromatic adaptation of strain E412. **(A)** Light microscopy image (1,000×) of strain E412 grown in white fluorescent light, 12L: 12D photoperiod at 45 μmol m^{-2} s^{-1}, **(B)** light microscopy image (1,000×) of strain E412 grown in far-red (730 nm) LED light, 24L: 0D photoperiod at 25 μmol m^{-2} s^{-1}, **(C)** composition of photosynthetic lipophilic pigments of strain E412 grown at different illumination conditions, the figure represents a mean of three biological replicates, **(D)** composition of photosynthetic water-soluble pigments of strain E412 grown at different illumination conditions, the figure represents a mean of three biological replicates.

GO term for cellular component, molecular function, and biological process, respectively. The most abundant GO terms distribution pattern was also noticed in another thermophilic strain, *T. sichuanensis* A183 isolated from a hot spring of Ganzi prefecture, China (Tang et al., 2021). However, the GO terms of the A183 genome were mostly concentrated only in eight functional categories. Meanwhile, their distribution in E412 was across many categories. Further KEGG pathway analysis showed that most genes (66.2%) in the E412 genome were distributed in the sub-category of metabolism (**Supplementary Figure 8**), suggesting that these metabolism-related genes may be crucial for this strain to survive in the oligotrophic aquatic environments of its origin.

Horizontal Gene Transfer

Based on a stringent screening of BLASP results, 1,944 out of 7,904 (24.6%) annotated protein-coding genes were identified as putatively acquired (**Supplementary Table 3**). Many putatively acquired genes were also notable in other thermophilic cyanobacteria, e.g., *Thermosynechococcus* sp. CL-1 (19%) (Cheng et al., 2020). The putatively acquired genes through HGT may contribute to the genome expansion and acquisition of new functions and conceivably acclimation to variable environments of these thermophilic strains. For instance, 22 genes annotated as heat shock proteins were identified as putatively acquired (**Supplementary Table 3**), which might play a crucial role in adapting to the thermal environment. The putatively acquired gene, flavodoxin gene (*fldA*), may substitute the function of ferredoxin in the photosynthetic electron transport chain under iron-deficient conditions (Cao et al., 2020). Twelve out of 29 genes annotated as circadian input kinase A (*cikA*) were identified as acquired genes from diverse donors (**Supplementary Table 3**). Among the acquired *cikA* genes, three showed very low protein similarities (38.4–40.5%) to the kinase of *Synechococcus elongatus* PCC 7942, which was experimentally demonstrated as a critical factor for entraining the clock in the cyanobacterium (Ivleva et al., 2006). The distinct proteins of acquired *cikA* genes in the E412 genome implied that alternative circadian rhythms could be used to time metabolic and behavioral events with the external environment through entrainment. Concerning function, the HGT candidates in the E412 genome were assigned to a wide range of functional categories as revealed by both GO and KEGG

analysis (**Supplementary Figures 7**, **8**). The most represented GO term was catalytic activity, followed by metabolic process and cellular process. Interestingly, the majority of HGT candidates (69.4%) suggested by the KEGG analysis were categorized as being related to metabolic functions. This result was consistent with the complexity hypothesis that fundamental genes are less likely than peripheral and operational genes to be horizontally transferred (Thomas and Nielsen, 2005).

We noticed that the criteria for HGT detection were distinct among reported studies. For example, genus-level (Cheng et al., 2020) or cyanobacteria-level (Chen et al., 2021) were separately customized to identify acquired genes in light of research objectives. Although modified approaches have been employed in this study or other studies, taxon-sampling biases may exist in the high-throughput BLASP-based approach. In addition, data availability was also an important factor affecting the establishment of the analysis pipeline, since only one genome sequence was obtained for this genus. More genome sequences of this genus are required for further HGT analysis, such as identification of recently acquired genes.

Mobile Genetic Elements

A total of 431 ISs (insertion sequences) corresponding to 67 different ISs were identified in the E412 genome. The IS630 family (67.29%) was dominant among the observed IS families, followed by the IS1 family (9.74%) and the IS4 family (7.66%). Besides, genes encoding transposase (86) were also noticed in abundance (**Supplementary Table 2**), indicating that intragenomic rearrangements might contribute much to the genetic plasticity of the strain.

One incomplete prophage loci was predicted in the chromosome, phiE412 (9.6 kb; positions 1,993,638–2,003,299). Thirteen phage-related genes were identified in the phiE412 region (**Supplementary Table 4**). No genes corresponding to DNA synthesis were found in the prophage loci, suggesting that this region was replication-defective. Repeats that constituted the core regions of phage attachment were not found to flank phiE412. Thus, this partial prophage loci probably was not functional. Clustered regularly interspaced short palindromic repeats (CRISPRs) were reported to function in the interference pathway to maintain genome integrity (Gasiunas et al., 2014). In the E412 chromosome, we detected three CRISPRs with high evidence levels and three Cas clusters assigned to type IA and III-D (**Supplementary Table 5**). However, the E412 genome had only one CRISPR-Cas array (type III-D), which was consistent with the finding in thermophilic strain *T. sichuanensis* A183 (Tang et al., 2021). Besides, several other genes typically associated with this system, e.g., *cmr*2, were also found to flank the array. The results indicated that this CRISPR-Cas interference system might function to limit HGT (Marraffini and Sontheimer, 2008).

Thermotolerance

As a microorganism living in a hot spring (temperature: 67.2°C), strain E412 must possess survival strategies to adapt to the thermal environment. The heat shock proteins (Hsps) play a crucial role in managing protein concentration, conformation, and subcellular location, especially when numerous stresses such as high temperature are applied. Expectedly, the E412 genome had numerous homologs of genes coding for Hsps (**Supplementary Table 2**). The homologs of the *clp* family (*clpB, -C, -P, -S, -X*) belonged to the Hsp100 family and helped maintain protein homeostasis (Labreck et al., 2017). The homologs of *htpG* protein of the Hsp90 family were likely to be primarily involved in protecting the photosynthetic machinery from heat stress (Takeshi et al., 2010). The homologs belonging to the Hsp70 family, mainly including *dnaK* and *dnaJ*, were abundant in the E412 genome. However, *dnaK* and *dnaJ* proteins might have different functions, and only part of them was responsible for thermotolerance (Duppre et al., 2011). Besides, a homolog of *grpE*, as a cofactor of the Hsp70 family, may help to prevent the aggregation of heat-denatured proteins (Schneider, 2011). As for the Hsp60 family, two distinct homologs of *groEL* genes, also referred to as the *groE* chaperone machinery, were identified in the E412 genome, and only one of them formed *groESL* operon with *groES*. This composition was similar to that of *T. sichuanensis* A183, but distinguished from that of *Gloeobacter* PCC 7421, the genome of which contains two *groESL* operons. Thus, we speculated that one *groESL* operon of the two strains lost the *groES* during the evolutionary process.

Photosynthesis

The E412 genome comprised the homologs of complete sets of genes coding for both photosystem I (14 genes; some with additional homologs) and photosystem II (22 genes; some with additional homologs) (**Supplementary Table 2**). The genome also had homologs of genes coding for phycobilisome proteins, allophycocyanin, phycocyanin, and phycoerythrin. The genes related to phycoerythrin (blue-light absorbing) might favor the strain to acclimate to low light intensity and oligotrophic environment (Tang et al., 2019). These genomic features strongly support the described experimental evidence for the chromatic adaptation capacity of this strain.

Surprisingly, the E412 genome harbored no homologs of genes encoding photoprotective proteins, such as flavodiiron proteins and orange carotenoid proteins. This result suggested that strain E412 probably copes with photodamage by alternative acclimation mechanisms in light of high altitude and high-light exposure in the environment of its origin. Interestingly, flavodoxin gene (*fldA*) and iron stress-inducible proteins (*isiA*) were found in the E412 genome. The presence of the two genes indicated that under iron-deficient conditions, strain E412 might use flavodoxins to substitute ferredoxin in electron transfer and to increase the light-absorbing efficiency of PSI in the form of PSI-*isiA*-flavodoxin supercomplex (Cao et al., 2020).

Essential genes required for the Carbon-dioxide Concentrating Mechanism (CCM) were present in the E412 genome (**Supplementary Table 2**). The uptake of gaseous CO_2 systems in cyanobacteria relied on NADPH dehydrogenase (NDH-1) complexes. The E412 genome included two NDH-1 complexes: a low-CO_2 inducible high-affinity NDH-1_3 complex encoded by *ndhD3*, *ndhF3,* and *cupA* (*chpY*) genes, and a constitutive low-affinity NDH-1_4 complex encoded by *ndhD4*, *ndhF4*, and *cupB* (*chpX*) genes. In addition, our previous study indicated that strain E412 can utilize bicarbonate as a one of the sources of inorganic carbon (Tang et al., 2018a). The genome analysis verified the experimental results through the

presence of homologs of low affinity, high flux, Na^+-dependent bicarbonate transporters (*bicA1* and *bicA2*) and several ABC-type bicarbonate transporters. Moreover, the genome had 21 *Hat*/*HatR* gene homologs encoding High-affinity carbon uptake protein. A similar abundance of *Hat*/*HatR* gene homologs was noticed in many strains, e.g., *Acaryochloris marina* and *Lyngbya aestuarii* BL J. On the contrary, strains like *Synechococcus* sp. WH8102 and *Synechocystis* sp. PCC 6,803 contain none. The prosperity of the *Hat*/*HatR* gene in E412 may be related to the mat habit of its origin, where diffusion becomes the primary transport mechanism for substrates and products of metabolism and can cause diffusion limitations to photosynthesis (Kothari et al., 2013).

Nitrogen Metabolism

Cyanobacteria can utilize various organic and inorganic nitrogen sources, a feature vital for survival in oligotrophic environments (Esteves-Ferreira et al., 2018). The E412 genome had the homologs of genes encoding nitrogenases, including *nifB*, *nifE*, *nifH*, *nifN*, *nifW,* and *nifX* (**Supplementary Table S2**). The genomic composition suggested that strain E412 was a nitrogen-fixing cyanobacterium, further verified by nitrogenase activity test (**Supplementary Figure 6**). Interestingly, strain E412 hosted a homolog of the gene *hetR*, the master regulatory gene involved in heterocyst formation, although it was not found to develop heterocysts under conditions tested.

The experimental result showed the utilization of both nitrate (3 mM) and nitrite (6 mM) by strain E412 for active growth. Tracing back to the genetic basis, the E412 genome possessed the homologs of genes encoding ABC-type nitrate transport system, which were clustered and oriented in the same direction with two essential nitrogen-related genes encoding ferredoxin-nitrite reductase (*nir*) and ferredoxin-nitrate reductase (*nar*), respectively (**Supplementary Table 2**). This result was in accordance with many freshwater cyanobacterial strains (Maeda et al., 2015). In addition, this strain had the homologs of an ammonium transporter and homologs of both glutamine synthetase and glutamine amidotransferase, which played the primary role of ammonium ion assimilation (Muro-Pastor et al., 2005). The homolog of the gene coding for *ntcB* was also present in this strain, which may function as a nitrogen control system to efficiently utilize intracellular resources in adaptation to changing nitrogen availability in the natural environment (Aichi and Omata, 1997).

CONCLUSION

In the present study, we have morphologically, physiologically, phylogenetically, and taxogenomically characterized a novel thermophilic cyanobacterium, strain E412 isolated from Lotus Lake hot spring situated in Ganzi prefecture, China. Results of polyphasic analysis suggested that strain E412 was a novel genus within the family Leptolyngbyaceae. Consequently, we have proposed a new genus *Leptothermofonsia sichuanensis Daroch, Tang and Shah et al gen. sp. nov*. as a best described to date representative of this taxon and proposed its delineation. Phylogenomic inference, average nucleotide/amino-acid identity, and the POCP between genomes supported the delineation of genus *Leptothermofonsia*. Furthermore, the obtained complete genome of *Leptothermofonsia* E412 facilitated the elucidation of genetic basis regarding genes related to thermotolerance, photosynthesis, and nitrogen metabolism. Furthermore, experiments regarding chromatic adaptation capacity and utilization of sodium bicarbonate, sulfur, and nitrogen compounds confirmed these physiological characteristics as indicated by the genome sequence. Additionally, approximately a quarter of the annotated protein-coding genes in the E412 genome could be acquired through HGT and may impact genome expansion and habitat adaptation. Overall, the complete genome of strain E412 provides the first insight into the genomic feature of the genus *Leptothermofonsia* and lays the foundations for future global ecogenomic and geogenomic studies.

Taxonomic Treatment and Description of *Leptothermofonsia sichuanensis* Daroch, Tang, and Shah et al. gen. nov

Phylum: Cyanobacteria

Order: Synechococcales

Family: Leptolyngbyaceae

Description: Cells brownish colored; filamentous; coiled or tangled filaments; sheath colorless, very thin, not lamellate; trichome slightly constricted and composed of cylindrical, elongated cells; narrow trichome; apical cells are round; cells longer than broad; no false branching (**Figure 4A**). Trichomes unbranched and composed of cylindrical, elongated cells, 1.2–1.8 μm in length and 0.8–1.1 μm in width. Cross-walls of the cells contained constrictions (**Figures 4B,C**). Centripetal invagination of the cell wall divided individual cells, and intracellular connections between vegetative cells were not present (**Figure 4C**). Five to six thylakoids in the parietal arrangement were located parallel at the cells' inner periphery. Sheath, septum, phycobilisome, carboxysomes, cyanophycin granule, lipid droplets, and polyphosphate bodies were all present in the cytoplasm (**Figures 4C,D**).

Type strain: is E412 (= FACHB-2490).

Type species: *Leptothermofonsia sichuanensis Daroch, Tang, and Shah et al. gen. nov*. (see below).

Etymology: "Lepto" exhibiting morphology typical to the members of Leptolynbyaceae family, "thermo" similar to thermophilic (high temperature tolerant), "fonsia," genus epithet derived from the Latin word *fons* meaning spring, since both representative of the genus to date, i.e., E412 and Greenland 10, are hot spring isolates; "sichuanensis" species epithet derives from the name of collection province.

Type locality: Thermal spring, Lotus Lake in Ganzi Prefecture of Sichuan Province, China.

Ecology of type locality: the sample occurred as a macroscopic brown-green submerged in the pond. Sample collection was done on 12.05.2016, with the humidity close to 71%. At the time of collection, the air temperature was 15°C, and the light intensity was around 1,000 lux. The pH of the spring was 6.32, and the concentration of total dissolved solids was 447 mmol L^{-1}.

Habitat: Thermal springs in Ganzi Prefecture of Sichuan Province, China (30°05′14″ N, 101°56′55″ E). This species (strain E412) is capable of bicarbonate assimilation (Tang et al., 2018a). The genome has the homologs of genes encoding nitrogenases, including *nifB*, *nifE*, *nifH*, *nifN*, *nifW,* and

nifX (**Supplementary Table 2**). The strain was experimentally confirmed as nitrogen-fixing. Interestingly, this strain hosted a homolog of the gene *hetR*, the master regulatory gene involved in the formation of heterocyst, although it does not develop heterocyst. Strain shows the capacity of chromatic adaptation by modifying the content of phycobilisomes depending on light conditions.

Holotype here designated: the culture of *Leptothermofonsia sichuanensis Daroch, Tang and Shah et al.* gen. sp. *nov*. was initially denoted and deposited in Peking University Algae Collection as PKUAC-SCTE412 has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession number FACHB-2490 as *Leptolyngbya* species after identification and authentication based on the full-length sequencing of the 16S rRNA gene along with folding of the secondary structures of the 16S-23S ITS region. After proper identification and authentication, the culture is maintained in the FACHB under the accession number FACHB-2490.

DATA AVAILABILITY STATEMENT

The complete genome sequence reported in this study has been deposited in GenBank with an accession number CP072600.

AUTHOR CONTRIBUTIONS

JT: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization, supervision, project administration, and funding acquisition. MS: investigation, methodology, validation, and writing—original draft. DY: formal analysis, software, and data curation. YJ: formal analysis, software, data curation, and writing—review and editing. LD, MW, and MMW: methodology, software, and data analysis. KZ: software and data curation. LL: formal analysis, investigation, data curation, and writing—original draft. ML: formal analysis, investigation, and data curation. KW: conceptualization, methodology, writing—review and editing. MD: conceptualization, methodology, resources, data curation, writing—original draft, writing-review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (31970092 and 32071480), the Antibiotics Research and Re-evaluation Key Laboratory of Sichuan Province (ARRLKF21-03), the Shenzhen Fundamental Research Program (GXWD20201231165807007-20200806170221001), and Tenure-Track Fund to MD. Funding bodies had no influence over the design and execution of this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.765105/full#supplementary-material

REFERENCES

Aichi, M., and Omata, T. (1997). Involvement of NtcB, a LysR family transcription factor, in nitrite activation of the nitrate assimilation operon in the cyanobacterium synechococcus sp. strain PCC 7942. *J. Bacteriol.* 179, 4671–4675. doi: 10.1128/jb.179.15.4671-4675.1997

Alcorta, J., Alarcón-Schumacher, T., Salgado, O., and Díez, B. (2020). Taxonomic novelty and distinctive genomic features of hot spring cyanobacteria. *Front. Genet.* 11:568223. doi: 10.3389/fgene.2020.568223

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2019). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859

Becerra-Absaln, I., Johansen, J. R., Muoz-Martn, M. A., and Montejano, G. (2018). Chroakolemma gen. nov. (*Leptolyngbyaceae*, cyanobacteria) from soil biocrusts in the semi-desert central region of Mexico. *Phytotaxa* 367, 201–218.

Bennett, A., and Bogorad, L. (1973). Complementary chromatic adaptation in a filamentous blue-green alga. *J. Cell Biol.* 58, 419–435. doi: 10.1083/jcb.58.2.419

Brito, Â, Ramos, V., Mota, R., Lima, S., Santos, A., Vieira, J., et al. (2017). Description of new genera and species of marine cyanobacteria from the portuguese Atlantic coast. *Mol. Phylogenet. Evolu.* 111, 18–34. doi: 10.1016/j.ympev.2017.03.006

Cao, P., Cao, D., Si, L., Su, X., Tian, L., Chang, W., et al. (2020). Structural basis for energy and electron transfer of the photosystem I–IsiA–flavodoxin supercomplex. *Nat. Plants* 6, 167–176. doi: 10.1038/s41477-020-0593-7

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., et al. (2021). Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J.* 15, 211–227. doi: 10.1038/s41396-020-00775-z

Cheng, Y. I., Lin, C., Chiu, Y. F., Hsueh, H. T., and Chu, H. A. (2020). Comparative genomic analysis of a novel strain of taiwan hot-spring cyanobacterium *Thermosynechococcus* sp. CL-1. *Front. Microbiol.* 11:82. doi: 10.3389/fmicb.2020.00082

Conesa, A., Götz, S., Garcíagómez, 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

Cordeiro, R. I. P., Luz, R., Vasconcelos, V., Gonalves, V., and Fonseca, A. (2020). Cyanobacteria phylogenetic studies reveal evidence for polyphyletic genera from thermal and freshwater habitats. *Diversity* 12:298.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46, W246–W251. doi: 10.1093/nar/gky425

David, A., Grant, J. R., Ana, M., Tanvir, S., Allison, P., Yongjie, L., 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

Dere, S., güneş, T., and Sivaci, R. (1998). Spectrophotometric determination of chlorophyll-a, b and total carotenoid contents of some algae species using different solvents. *Turkish J. Bot.* 22, 13–17.

Duppre, E., Rupprecht, E., and Schneider, D. (2011). Specific and promiscuous functions of multiple DnaJ proteins in *Synechocystis* sp. PCC 6803. *Microbiology* 157, 1269–1278. doi: 10.1099/mic.0.045542-0

Dvorak, P., Hindak, F., Hasler, P., Hindakova, A., and Poulickova, A. (2014). Morphological and molecular studies of *Neosynechococcus sphagnicola*, gen. et sp. nov. (*Cyanobacteria, Synechococcales*). *Phytotaxa* 170, 24–34.

Dvorak, P., Jahodarova, E., Hašler, P., Gusev, E., and Poulíčková, A. (2015). A new tropical cyanobacterium *Pinocchia polymorpha* gen. et sp. nov. derived from the genus *Pseudanabaena*. *J. Czech Phycol. Soc.* 15, 113–120.

Esteves-Ferreira, A. A., Inaba, M., Fort, A., Araújo, W. L., and Sulpice, R. (2018). Nitrogen metabolism in cyanobacteria: metabolic and molecular control, growth consequences and biotechnological applications. *Crit. Rev. Microbiol.* 44, 541–560. doi: 10.1080/1040841X.2018.1446902

Gan, F., and Bryant, D. A. (2015). Adaptive and acclimative responses of cyanobacteria to far-red light. *Environ. Microbiol.* 17, 3450–3465. doi: 10.1111/1462-2920.12992

Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell. Mol. Life Sci.* 71, 449–465. doi: 10.1007/s00018-013-1438-6

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

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evolu.* 35, 518–522. doi: 10.1093/molbev/msx281

Iteman, I., Rippka, R., Tandeau, D. M. N., and Herdman, M. (2000). Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. *Microbiology* 146, 1275–1286. doi: 10.1099/00221287-146-6-1275

Ivleva, N., Gao, T., Liwang, A., and Golden, S. (2006). Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17468–17473. doi: 10.1073/pnas.0606639103

Jahodářová, E., Dvorak, P., Hasler, P., and PoulícKová, A. (2017). Revealing hidden diversity among tropical cyanobacteria : the new genus *Onodrimia* (*Synechococcales, Cyanobacteria*) described using the polyphasic approach. *Phytotaxa* 326, 28–40.

Jain, C., and Rodriguez, R. L. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Johansen, J. R., Kovacik, L., Casamatta, D. A., Iková, K. F., and Kaštovský, J. (2011). Utility of 16S-23S ITS sequence and secondary structure for recognition of intrageneric and intergeneric limits within cyanobacterial taxa: *Leptolyngbya corticola* sp. nov. (*Pseudanabaenaceae. Cyanobacteria*). *Nova Hedwigia* 92, 283–302.

Komárek, J., and Anagnostidis, K. (2005). *Cyanoprokaryota 2. Teil/2ndPart : Oscillatoriales*. München: SpringerSpektrum.

Komárek, J., Kaštovský, J., Mares, J., and Johansen, J. (2014). Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86, 295–335.

Komárek, J., and Lukavský, J. (1988). Arthronema, a new cyanophyte genus from Afro-Asian deserts. *Arch. Fur Hydrobiol.* 50, 249–267.

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

Kothari, A., Vaughn, M., and Garcia-Pichel, F. (2013). Comparative genomic analyses of the cyanobacterium, lyngbya aestuarii BL J, a powerful hydrogen producer. *Front. Microbiol.* 4:363. doi: 10.3389/fmicb.2013.00363

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

Labreck, C. J., Shannon, M., Viola, M. G., Joseph, C., and Camberg, J. L. (2017). The protein chaperone ClpX targets native and non-native aggregated substrates for remodeling, disassembly, and degradation with ClpP. *Front. Mol. Biosci.* 4:26. doi: 10.3389/fmolb.2017.00026

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, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Li, M., Cheng, L., Tang, J., and Daroch, M. (2021). Molecular components of nitrogen fixation gene cluster and associated enzymatic activities of non-heterocystous thermophilic cyanobacterium *Thermoleptolyngbya* sp. *Life* 11:640. doi: 10.3390/life11070640

Li, X., and Li, R. (2016). Limnolyngbya circumcreta gen. & comb. nov. (*Synechococcales, Cyanobacteria*) with three geographical (provincial) genotypes in China. *Phycologia* 55, 478–491.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Maeda, S.-i, Murakami, A., Ito, H., Tanaka, A., and Omata, T. (2015). Functional characterization of the FNT family nitrite transporter of marine picocyanobacteria. *Life* 5, 432–446. doi: 10.3390/life5010432

Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845. doi: 10.1126/science.1165771

Mathews, D. H. (2014). RNA secondary structure analysis using RNAstructure. *Curr. Protocols Bioinform.* 46, 1–25.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evolu.* 37, 1530–1534.

Miscoe, L. H., Johansen, J. R., Kociolek, J. P., Lowe, R. L., Vaccarino, M. A., Pietrasiak, N., et al. (2016). The diatom flora and cyanobacteria from caves on kauai, hawaii. II. novel cyanobacteria from caves on kauai, hawaii. *Bibliotheca Phycologica* 58, 3–4.

Muro-Pastor, M. I., Reyes, J. C., and Florencio, F. J. (2005). Ammonium assimilation in cyanobacteria. *Photosynthesis Res.* 83, 135–150.

Pannekoek, Y., Qi-Long, Q., Zhang, Y.-Z., and van der Ende, A. (2016). Genus delineation of chlamydiales by analysis of the percentage of conserved proteins justifies the reunifying of the genera chlamydia and chlamydophila into one single genus chlamydia. *Pathogens Dis.* 74:ftw071. doi: 10.1093/femspd/ftw071

Patel, A., Matsakas, L., Rova, U., and Christakopoulos, P. (2019). A perspective on biotechnological applications of thermophilic microalgae and cyanobacteria. *Bio. Technol.* 278, 424–434. doi: 10.1016/j.biortech.2019.01.063

Qin, Q., Xie, B., Zhang, X., Chen, X., Zhou, B., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196, 2210–2215.

Raabova, L., Kovacik, L., Elster, J., and Strunecky, O. (2019). Review of the genus phormidesmis (cyanobacteria) based on environmental, morphological, and molecular data with description of a new genus leptodesmis. *Phytotaxa* 395, 1–16.

Rodriguez-R, L. M., Santosh, G., Harvey, W. T., Ramon, R.-M., Tiedje, J. M., Cole, J. R., et al. (2018). The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of archaea and bacteria at the whole genome level. *Nuclc Acids Res.* 46, W282–W288. doi: 10.1093/nar/gky467

Roeselers, G., Norris, T. B., Castenholz, R. W., Rysgaard, S., Glud, R. N., Kühl, M., et al. (2007). Diversity of phototrophic bacteria in microbial mats from arctic hot springs (greenland). *Environ. Microbiol.* 9, 26–38. doi: 10.1111/j.1462-2920.2006.01103.x

Schneider, D. (2011). Thermostability of two cyanobacterial GrpE thermosensors. *Plant Cell Physiol.* 52, 1776–1785. doi: 10.1093/pcp/pcr116

Sciuto, K., and Moro, I. (2016). Detection of the new cosmopolitan genus thermoleptolyngbya (*Cyanobacteria, Leptolyngbyaceae*) using the 16S rRNA gene and 16S–23S ITS region. *Mol. Phylogenet. Evolu.* 105, 15–35. doi: 10.1016/j.ympev.2016.08.010

Shalygin, S., Shalygina, R., Redkina, V., Gargas, C., and Johansen, J. (2020). Description of stenomitos kolaenensis and S. hiloensis sp. nov. (*Leptolyngbyaceae, Cyanobacteria*) with an emendation of the genus. *Phytotaxa* 440, 108–128.

Six, C., Thomas, J.-C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., et al. (2007). Diversity and evolution of phycobilisomes in marine synechococcus spp.: a comparative genomics study. *Genome Biol.* 8:R259. doi: 10.1186/gb-2007-8-12-r259

Soares, F., Tiago, I., Trovo, J., Coelho, C., and Portugal, A. (2019). Description of myxacorys almedinensis sp. nov. (*Synechococcales, Cyanobacteria*) isolated from the limestone walls of the old cathedral of coimbra, portugal (UNESCO world heritage site). *Phytotaxa* 419, 77–90.

Song, G., Jiang, Y., and Li, R. (2015). Scytolyngbya timoleontis, gen . et sp . nov . (*Leptolyngbyaceae, Cyanobacteria*): a novel false branching cyanobacteria from China. *Phytotaxa* 224, 72–84.

Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evolu.* 30, 772–780. doi: 10.1093/molbev/mst010

Takeshi, S., Shun, M., and Erika, K. (2010). HtpG, the prokaryotic homologue of Hsp90, stabilizes a phycobilisome protein in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Mol. Microbiol.* 76, 576–589. doi: 10.1111/j.1365-2958.2010.07139.x

Tang, J., Du, L., Li, M., Yao, D., Waleron, M., Waleron, K. F., et al. (2022). Characterization of a novel hot-spring cyanobacterium leptodesmis sichuanensis sp. nov. and genomic insights of molecular adaptations into its habitat. *Front. Microbiol.* 12:739625. doi: 10.3389/fmicb.2021.739625

Tang, J., Du, L.-M., Liang, Y.-M., and Daroch, M. (2019). Complete genome sequence and comparative analysis of *Synechococcus* sp. CS-601 (SynAce01), a cold-adapted cyanobacterium from an oligotrophic antarctic habitat. *Int. J. Mol. Sci.* 20:152. doi: 10.3390/ijms20010152

Tang, J., Liang, Y., Jiang, D., Li, L., Luo, Y., Shah, M. M. R., et al. (2018b). Temperature-controlled thermophilic bacterial communities in hot springs of western sichuan, china. *BMC Microbiol.* 18:134. doi: 10.1186/s12866-018-1271-z

Tang, J., Jiang, D., Luo, Y., Liang, Y., Li, L., Shah, M. M. R., et al. (2018a). Potential new genera of cyanobacterial strains isolated from thermal springs of western sichuan. *China. Algal Res.* 31, 14–20.

Tang, J., Li, L., Li, M., Du, L., Shah, M. R., Waleron, M., et al. (2021). Description, taxonomy, and comparative genomics of a novel species, *Thermoleptolyngbya sichuanensis* sp. nov., isolated from hot springs of ganzi, sichuan, China. *Front. Microbiol.* 12:696102. doi: 10.3389/fmicb.2021.696102

Tatiana, T., Michael, D. C., Azat, B., Vyacheslav, C., Nawrocki, E. P., Leonid, Z., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624.

Taton, A., Grubisic, S., Ertz, D., Hodgson, D. A., Piccardi, R., Biondi, N., et al. (2010). Polyphasic study of Antarctic cyanobacterial strains. *J. Phycol.* 42, 1257–1270.

Taton, A., Wilmotte, A., Smarda, J., Elster, J., and Komarek, J. (2011). Plectolyngbya hodgsonii: a novel filamentous cyanobacterium from Antarctic lakes. *Polar Biol.* 34, 181–191.

Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between Bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi: 10.1038/nrmicro1234

Thomazeau, S., Houdan-Fourmont, A., Couté, A., Duval, C., Couloux, A., Rousseau, F., et al. (2010). The contribution of sub-saharan african strains to the phylogeny of cyanobacteria: focusing on the nostocaceae (nostocales, cyanobacteria). *J. Phycol.* 46, 564–579.

Turland, N., Wiersema, J., Barrie, F. R., Greuter, W., and Smith, G. F. (2018). *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress Shenzhen*. Glashütten: Koeltz Botanical Books.

Varani, A. M., Siguier, P., Gourbeyre, E., and Charneau, V. (2011). ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* 12:R30. doi: 10.1186/gb-2011-12-3-r30

Vaz, M. G. M. V., Genuário, D. B., Andreote, A. P. D., Malone, C. F. S., Sant'Anna, C. L., Barbiero, L., et al. (2015). Pantanalinema gen. nov. and alkalinema gen. nov.: novel pseudanabaenacean genera (cyanobacteria) isolated from saline–alkaline lakes. *Int. J. Syst. Evolu. Microbiol.* 65, 298–308. doi: 10.1099/ijs.0.070110-0

Vincent, L., Jean-Emmanuel, L., and Olivier, G. (2017). SMS: smart model selection in PhyML. *Mol. Biol. Evolu.* 34, 2422–2424. doi: 10.1093/molbev/msx149

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

Walter, J. M., Coutinho, F. H., Dutilh, B. E., Swings, J., Thompson, F. L., and Thompson, C. C. (2017). Ecogenomics and taxonomy of cyanobacteria phylum. *Front. Microbiol.* 8:2132. doi: 10.3389/fmicb.2017.02132

Yao, D., Cheng, L., Du, L., Li, M., Daroch, M., and Tang, J. (2021). Genome-wide investigation and analysis of microsatellites and compound microsatellites in leptolyngbya-like species, cyanobacteria. *Life* 11:1258. doi: 10.3390/life11111258

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, 293–297.

Zimin, A. V., Marcais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677. doi: 10.1093/bioinformatics/btt476

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tang, Shah, Yao, Jiang, Du, Zhao, Li, Li, Waleron, Waleron, Waleron and Daroch. 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 | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fmicb.2022.777133

The Isolation and Characterization of Rare Mycobiome Associated With Spacecraft Assembly Cleanrooms

Adriana Blachowicz[1†], Snehit Mhatre[1†], Nitin Kumar Singh[1], Jason M. Wood[1], Ceth W. Parker[1], Cynthia Ly[1], Daniel Butler[2], Christopher E. Mason[2,3] and Kasthuri Venkateswaran[1]*

[1] Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, [2] Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, United States, [3] The WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, New York, NY, United States

Ensuring biological cleanliness while assembling and launching spacecraft is critical for robotic exploration of the solar system. To date, when preventing forward contamination of other celestial bodies, NASA Planetary Protection policies have focused on endospore-forming bacteria while fungi were neglected. In this study, for the first time the mycobiome of two spacecraft assembly facilities at Jet Propulsion Laboratory (JPL) and Kennedy Space Center (KSC) was assessed using both cultivation and sequencing techniques. To facilitate enumeration of viable fungal populations and downstream molecular analyses, collected samples were first treated with chloramphenicol for 24 h and then with propidium monoazide (PMA). Among cultivable fungi, 28 distinct species were observed, 16 at JPL and 16 at KSC facilities, while 13 isolates were potentially novel species. Only four isolated species *Aureobasidium melanogenum*, *Penicillium fuscoglaucum*, *Penicillium decumbens,* and *Zalaria obscura* were present in both cleanroom facilities, which suggests that mycobiomes differ significantly between distant locations. To better visualize the biogeography of all isolated strains the network analysis was undertaken and confirmed higher abundance of *Malassezia globosa* and *Cyberlindnera jadinii*. When amplicon sequencing was performed, JPL-SAF and KSC-PHSF showed differing mycobiomes. Metagenomic fungal reads were dominated by Ascomycota (91%) and Basidiomycota (7.15%). Similar to amplicon sequencing, the number of fungal reads changed following antibiotic treatment in both cleanrooms; however, the opposite trends were observed. Alas, treatment with the antibiotic did not allow for definitive ascribing changes observed in fungal populations between treated and untreated samples in both cleanrooms. Rather, these substantial differences in fungal abundance might be attributed to several factors, including the geographical location, climate and the in-house cleaning procedures used to maintain the cleanrooms. This study is a first step in characterizing cultivable and viable fungal populations in cleanrooms to assess fungal potential as biocontaminants during interplanetary explorations. The outcomes of this and future studies could be implemented in other cleanrooms that require to reduce microbial burden, like intensive care units, operating rooms, or cleanrooms in the semiconducting and pharmaceutical industries.

Keywords: mycobiome, cleanroom, chloramphenicol, metagenome, amplicon sequencing

OPEN ACCESS

Edited by:
André Antunes, Macau University of Science and Technology, Macao SAR, China

Reviewed by:
Parag Vaishampayan, National Aeronautics and Space Administration (NASA), United States
Marta Filipa Simões, Macau University of Science and Technology, Macao SAR, China

****Correspondence:***
Kasthuri Venkateswaran
kjvenkat@jpl.nasa.gov

† These authors have contributed equally to this work

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

Received: *14 September 2021*
Accepted: *04 March 2022*
Published: *26 April 2022*

Citation:
Blachowicz A, Mhatre S, Singh NK, Wood JM, Parker CW, Ly C, Butler D, Mason CE and Venkateswaran K (2022) The Isolation and Characterization of Rare Mycobiome Associated With Spacecraft Assembly Cleanrooms. Front. Microbiol. 13:777133. doi: 10.3389/fmicb.2022.777133

INTRODUCTION

The primary aim of the National Aeronautical Space Administration (NASA) Planetary Protection (PP) program is to prevent forward and back contamination between Earth and other celestial bodies (Rummel et al., 2002). To comply with these requirements, rigorous maintenance procedures have been implemented in the cleanrooms where spacecrafts are assembled, including the spacecraft assembly facility (SAF) at Jet Propulsion Laboratory (JPL) and Payload Hazardous Servicing Facility (PHSF) at Kennedy Space Center (KSC) (La Duc et al., 2003, 2004). Nevertheless, despite stringent "housekeeping" rules, reports have shown that viable microbial populations can be retrieved (La Duc et al., 2004, 2012; Weinmaier et al., 2015).

Humans may be considered walking incubators for microorganisms (Meadow et al., 2015), shedding commensal microorganisms to any environment they inhabit. These cast-off microorganisms influence persisting populations in closed habitats, including Lunar/Mars Analog Habitat (ILMAH) (Mayer et al., 2016; Blachowicz et al., 2017), Mars500 facility (Mora et al., 2016) and SAF cleanroom (Weinmaier et al., 2015). Since bacteria are predominant in the human (Meadow et al., 2015; Sender et al., 2016) closed habitats' (La Duc et al., 2003, 2004, 2012), ILMAH (Mayer et al., 2016), Mars500 (Mora et al., 2016) and the International Space Station (ISS) microbiomes (Checinska et al., 2015; Singh et al., 2018; Checinska Sielaff et al., 2019), the majority of conducted environmental studies have focused on investigating their resilience and pathogenicity (Bashir et al., 2016; Singh et al., 2018; Urbaniak et al., 2018). However, despite lower abundance in JPL cleanrooms (∼2–3% of all isolated and identified microorganisms) (Hendrickson et al., 2017), fungi should not be disregarded in the context of future outer space exploration. Fungal survival in space environments can be fostered by their hardy nature and capacity to produce a myriad of bioactive compounds (Blachowicz et al., 2019a; Keller, 2019).

Enormous phenotypic plasticity and adaptability allow fungi to persist and grow in extreme (Zhdanova et al., 2004; Sterflinger et al., 2012) and built environments (Adams et al., 2013; Blachowicz et al., 2017). Traditional assays employed by NASA PP program target specifically the detection of endospore-forming bacteria (Benardini et al., 2014). Their extreme hardiness allows for tolerating inhospitable conditions for long periods of time, which makes them suitable candidates for surviving the journey to other planetary bodies. However, while fungal species also produce protective structures (spores, conidia, or cysts), often in response to environmental stressors, few studies have examined their survival under simulated space conditions (Gorbushina, 2003; Blachowicz et al., 2019a). A few reports discussed the fungal distribution in assembly facilities (Herring et al., 1974; La Duc et al., 2012), but no systematic characterization utilizing state-of-the-art molecular techniques has been attempted. Since the presence of fungi has been documented in JPL cleanrooms (La Duc et al., 2012; Weinmaier et al., 2015), there is an unmet need to study the complexity of fungal population and its fluctuation over time. Furthermore, in light of emerging research, such analyses are imperative to discern the potential for fungi to be contamination agents while exploring space (Knox et al., 2016; Romsdahl et al., 2018a; Blachowicz et al., 2019b).

Several studies have tested using antibiotics to reduce bacterial populations in environmental or food samples to facilitate the estimation of the fungal components (Cooke, 1954; Koburger and Rodgers, 1978; Tournas et al., 2020). The impact of antibiotics on the growth of selected fungi was tested, revealing that chloramphenicol and gentamicin were effective in reducing bacterial populations without impairing fungal growth (Dolan, 1971). In principle, chloramphenicol reversibly binds to the bacterial 50S ribosomal subunit and interferes with protein synthesis and cell proliferation (Oong and Tadi, 2021). Given that, unlike gentamicin, chloramphenicol is bacteriostatic to both Gram-positive and Gram-negative bacteria, short incubations with it were expected to enhance retrieval of fungi from SAF samples.

This study is the first report focusing mainly on the mycobiomes of two NASA SAF cleanrooms implementing an approach not undertaken before. Collected environmental samples were treated with chloramphenicol for 24 h to suppress bacterial activity. Subsequently, samples were treated with propidium monoazide (PMA) to intercalate DNA released from *dead/damaged* microbial cells. As a result of these treatments, characterization of *viable/intact* fungal populations was possible. Culture- and molecular-based techniques were implemented to profile the composition of SAF mycobiomes. Additionally, fungal diversity was determined by both amplicon and metagenome sequencing to analyze identified communities in terms of their taxonomy, virulence and stress-related functional properties.

MATERIALS AND METHODS

Sampling Locations and Environmental Conditions

Two geographically distinct NASA cleanrooms, JPL-SAF and the KSC-PHSF, were sampled between April and September of 2018. The specific sampling locations and the corresponding environmental data are given in **Table 1** and **Supplementary Figure 1**. Both facilities were maintained with cleaning regimens appropriate to the current level of on-going activities. During samplings at JPL-SAF, assembly activities of critical spacecraft hardware were happening; therefore, SAF was cleaned daily, including vacuuming and mopping with a cleaning solution (Kleenol 30, a highly concentrated, industrial strength heavy-duty cleaner/degreaser, which contains12.5% ethylene glycol monobutyl ether, 1–5% nonylphenol ethoxylate, 1% dodecylbenzenesulfonate, and 1–4% silicic acid, disodium salt). When, KSC-PHSF was sampled, no assembly activity or spacecraft hardware were present, thus daily cleaning included vacuuming and mopping with plain water. Additionally, daily cleaning regimens in both cleanrooms include replacing tacky mats. All personnel who enter these cleanrooms must follow good manufacturing practice procedures to minimize the influx of particulates. Specific entry procedures depend on the certification level of the cleanroom and the presence of mission hardware.

TABLE 1 | Metadata of samples collected from JPL-SAF and KSC-PHSF.

Facility/Location/ Dimensions [m]	Characteristics[a]	Significance	Sample ID	Abbrev. name	Description	Date	Sampling method	Sampling area	Extraction volume (ml)
Jet Propulsion Laboratory Spacecraft Assembly Facility (JPL-SAF) Pasadena, CA West coast; dry desert like 25 × 36 × 15	ISO 7 (Class 10,000) HEPA Room Temperature: 20.4 ± 0.3°C Room Humidity: 45.5 ± 2%	Pathfinder (1997); Mars Exploration Rovers (2004); Mars Science Laboratory (2012); Perseverance (2020)	FD-01-001	L1	SW, near desk area and walkway to desk area	4/17/2018	BiSKit	1 m^2	45
			FD-01-002	L2	SW, corner between desk and front of unused door				
			FD-01-003	L3	SW, desk area & walking area to Cruise Stage				
			FD-01-004	L4	S, between cruise stage and Descent stage				
			FD-01-005	L5	Center, between desk and Descent stage				
			FD-01-006	L6	Center, walking area to desk behind monitor				
			FD-01-007	L7	W Center, next to center walkway, crane, and desk				
			FD-01-008	L8	SE, area in front of high bay spacecraft exit doors				
			FD-01-009	L9	Center, next to storage boxes in center of room				
			FD-01-010	L10	E, directly in front of room entrance, after tacky mat				
			FD-01-HC-01	HC	Handling control (exposed to facility air)			—	
			FD-01-DC-01	DC	Device control (moistened BiSKit, not used)			—	
Kennedy Space Center Payload Hazardous Servicing Facility (KSC-PHSF) Cape Canaveral, FL East coast; swamp-like 33 × 18 × 29	ISO 7 (Class 10,000) HEPA Room Temperature: 21.7 ± 3.3°C Room Humidity: =55%	Mars Exploration Rovers (2004); MarsReconn-AissanceOrbiter (2006)Mars Science Laboratory (2012); Perseverance (2020)	FD-03-001	L1	SW, near airlock door for equipment storage moving	6/12/2018	BiSKit	1 m^2	45
			FD-03-002	L2	S center, between airlock entrance and exit into cleanroom				
			FD-03-003	L3	SE, near wall of high bay, grounding plate included				
			FD-03-004	L4	E center, left of grounding plate				
			FD-03-005	L5	Center, high bay center right				
			FD-03-006	L6	Center, high bay center left				
			FD-03-007	L7	W center, near west wall by fuel drain				
			FD-03-008	L8	NW, corner near west exit door 19 to outside				
			FD-03-009	L9	N center, front of door 18				
			FD-03-010	L10	NE, front of door 17				
			FD-03-HC-01	HC	Handling control (exposed to facility air)			—	
			FD-03-DC-01	DC	Device control (moistened BiSKit, not used)			—	
			FD-03-CM-01	C1	SW, alongside length of grate		ClipperMop	10 m^2	200
			FD-03-CM-02	C2	Center, high bay center				
			FD-03-CM-03	C3	NE, between door 18 and 17				
			FD-03-CM-HC	CHC	Handling control (exposed to facility air)			—	
			FD-03-CM-DC	CDC	Device control (moistened CM wipe, not used)			—	

(Continued)

TABLE 1 | (Continued)

Facility/Location/ Dimensions [m]	Characteristics[a]	Significance	Sample ID	Abbrev. name	Description	Date	Sampling method	Sampling area	Extraction volume (ml)
Kennedy Space Center Payload Hazardous Servicing Facility (KSC-PHSF) Cape Canaveral, FL East coast; swamp-like	ISO 7 (Class 10,000) HEPA Room Temperature: 21.7 ± 3.3°C Room Humidity: =56%	Mars Exploration Rovers (2004); MarsReconn-AissanceOrbiter (2006)Mars Science Laboratory (2012); Perseverance (2020)	FD-04-001	L1	SW, near airlock door for equipment storage moving	7/24/2018	BiSKit	1 m^2	45
			FD-04-002	L2	S center, between airlock entrance and exit into cleanroom				
			FD-04-003	L3	SE, near wall of high bay, grounding plate included				
			FD-04-004	L4	E center, left of grounding plate				
			FD-04-005	L5	Center, high bay center right				
			FD-04-006	L6	Center, high bay center left				
			FD-04-007	L7	W center, near west wall by fuel drain				
			FD-04-008	L8	NW, corner near west exit door 19 to outside				
33 × 18 × 29			FD-04-009	L9	N center, front of door 18				
			FD-04-010	L10	NE, front of door 17				
			FD-04-HC-01	HC	Handling Control (exposed to facility air)			—	
			FD-04-DC-01	DC	Device control (moistened BiSKit, not used)			—	
			FD-04-CM-01	C1	SW, alongside length of grate		ClipperMop	10 m^2	200
			FD-04-CM-02	C2	Center, high bay center				
			FD-04-CM-03	C3	NE, between door 18 and 17				
			FD-04-CM-HC	CHC	Handling control (exposed to facility air)			—	
			FD-04-CM-DC	CDC	Device control (moistened CM wipe, not used)			—	
Jet Propulsion Laboratory Spacecraft Assembly Facility (JPL-SAF) Pasadena, CA West coast; dry desert like	ISO 7 (Class 10,000) HEPA Room Temperature: 20.3 ± 0°C Room Humidity: 45.8 ± 0.1%	Pathfinder (1997); Mars Exploration Rovers (2004); Mars Science Laboratory (2012); Perseverance (2020)	FD-05-010	L1	SW, near desk area and walkway to desk area	9/25/2018	Wipe	1 m^2	45
			FD-05-005	L7	W Center, next to center walkway, crane, and desk				
			FD-05-003	L9	Center, next to storage boxes in center of room				
			FD-05-001	L10	E, directly in front of room entrance, after tacky mat				
			FD-05-002	L11	NE, desk area near entrance				
			FD-05-004	L12	SE, directly front of door to other airlock (unused)				
			FD-05-006	L13	N center, in front of cabinets				
			FD-05-007	L14	Center, between wheels of lift				
			FD-05-008	L15	NW, between GSE				
25× 36 × 15			FD-05-009	L16	NW, between tool boxes				
			FD-05-HC-01	HC	Handling control (exposed to facility air)			—	
			FD-05-DC-01	DC	Device control (moistened wipe not used)			—	
			FD-05-CM-01	C1	E center, near entrance in front of solar panels		Wipe	10 m^2	200
			FD-05-CM-02	C2	Center, near walkway				
			FD-05-CM-03	C3	E, alongside high bay spacecraft exit doors				
			FD-05-CM-HC	CHC	Handling control (exposed to facility air)			—	
			FD-05-CM-DC	CDC	Device control (moistened CM wipe, not used)			—	

[a]*Characteristic of the cleanroom ISO level is a certification defined by the maximum number of particles of the size >0.5 μm in one cubic foot of air. The cleanrooms' temperature, humidity, and filter (HEPA, high-efficiency particle filter) are given as its average with± consideration to its min/maximum.*

General precautions include donning of cleanroom-certified garments to minimize exposure of skin, hair and the regular clothing of personnel. In addition, general precautions prohibit the use of cosmetics, fragrances, body spray and hair gels before entry into the cleanroom. Lastly, the air of both facilities was filtered through high-efficiency particle air filters.

Sampling Procedure and Initial Sample Processing

A schematic overview of the methodology implemented to carry out this study is presented in **Supplementary Figure 2**. Cleanroom floor samplings were performed using three sampling devices: (i) BiSKit; (ii) ClipperMop and (iii) polyester wipe, each of which were premoistened with sterile molecular grade water. The first sampling at JPL-SAF on April 17, 2018 was carried out with BiSKits (Quicksilver Analytics Inc., Abingdon, MD) prepared following the previously described procedure (Blachowicz et al., 2017). Each BiSKit was used to sample a 1 m^2 area from 10 selected locations (**Supplementary Figure 1**). Due to the presence of the sensitive payload in JPL-SAF the sampling on September 25, 2018 required changes in the sampling locations and procedure, as the BiSKit is not a flight certified sampling device. Likewise, six sampled locations (L11–L16), where adjusted during the second sampling. Therefore, 10 samples (1 m^2 each) were collected with 12" × 12" premoistened, flight-certified polyester wipes (Sterile TexTra10 TX3225; Texwipe, Kernersville, NC) and additional three samples (10 m^2 each) were collected with ClipperMop (TexWipe) using 12" × 12" premoistened, flight certified polyester wipes attached to the mop. Both sampling events at KSC-PHSF on June 12, 2018 and July 17, 2018 were conducted using BiSKits and ClipperMops, collecting samples from 10 to 3 different locations, respectively.

After sampling, BiSKits and corresponding handling and device controls were processed following previously described steps (Blachowicz et al., 2017). Briefly, each sampled BiSKit was extracted three times with 15 mL phosphate buffer saline (PBS) (MoBio Laboratories Inc., Carlsbad, CA) amounting to 45 mL. Polyester and ClipperMop wipes were placed in 500 mL bottle containing 200 mL of sterile PBS, vigorously shaken for 1 min to dislodge microbial cells. These environmental samples were then concentrated using CP-150 InnovaPrep concentrating pipette (Innova Prep LLC, Drexel, MO) to a final volume of ∼6 mL (Kwan et al., 2011). Concentrated BiSKit, polyester wipe and ClipperMop samples were divided into two aliquots (∼2 mL each).

Chloramphenicol Treatment

Chloramphenicol is a bacteriostatic antibiotic (Oong and Tadi, 2021) that most fungal species resist; therefore, it can be used to suppress bacterial proliferation allowing for isolation of fungi. Aliquots of the concentrated samples were treated with chloramphenicol at final concentration 100 μg/mL and incubated overnight at 25°C along with untreated counterparts. After 18–24 h of incubation, both chloramphenicol treated and untreated samples were processed for various analyses as detailed below.

Sample Treatment and Assays

Cultivable Fungal Burden and Diversity

To assess cultivable burden, samples collected during the first sampling at JPL-SAF (JPL-1) were 10-fold diluted and 100 μL was added in duplicate to potato dextrose agar (PDA) containing chloramphenicol (25 mg/L) and dichloran rose bengal chloramphenicol agar (DRBC) and grown at room temperature (RT) (∼25°C). After 7 days, colony-forming units (CFUs) were enumerated and used to calculate the cultivable fungal population per meter square. All colonies that grew were collected and stored as stab cultures and in glycerol stocks. Based on the observed results for JPL-1, the processing method for the following three sampling events (JPL-2, KSC-1, and KSC-2) was altered as follows. Upon concentration, samples were divided into two equal aliquots, with one treated with chloramphenicol (100 μg/mL) for 24 h while the other was left untreated. Antibiotic-treated and untreated samples were then 10-fold diluted and 100 μL was plated in duplicates on PDA and DRBC containing chloramphenicol. After 7 days of incubation at room temperature, colonies were counted and collected for identification with the internal transcribed spacers (ITS) region using primers ITS 1F (5′-CTT GGT CAT TTA GAG GAA GTA A-3′) (Lai et al., 2007) and Tw13 (5′-GGT CCG TGT TTC AAG ACG-3′) (Taylor and Bruns, 1999). DNA was extracted using the Maxwell-16 MDx automated system following the manufacturer's instructions (Promega, Madison, WI). PCR conditions and sample preparation steps for sequencing were exactly as described elsewhere (Blachowicz et al., 2017). Sanger sequencing was performed and ITS sequences were identified through the Basic Local Alignment Search Tool (BLAST) algorithm (Altschul et al., 1990) using the National Center for Biotechnology Information (NCBI) database to find the type strains with the closest percent similarity to each isolate.

Assessment of UV-C Resistance

To evaluate ultraviolet radiation with wavelengths between 200 and 290 nm (UV-C) sensitivity, approximately 10^2 conidia/plate were exposed to UV-C. Conidia were suspended in PDA top agar (PDA, containing half of the agar concentration, 7.5 g/L) and 5 mL of this mix was added to plates containing 20 mL of PDA (total agar amount, 15 g/L). Duplicates of each isolate were exposed to 1,000 J/m^2 using a low-pressure handheld mercury arc UV lamp (model UVG-11; UVP Inc., Upland, CA). Following exposure, treated and untreated (control) plates were incubated at room temperature. After 7 days, colonies were enumerated and relative survival was assessed and calculated using the following formula: no. of CFU exposed to any given dose (N)/CFU in the control plate (N_0).

Validation of Propidium Monoazide Assay to Detect Viable Microbial Cells

DNA of *Aureobasidium pullulans* MO-28v1 was used to confirm that PMA was intercalating into the fungal DNA and preventing qPCR amplification. Heat sterilization was used to create fungal suspensions containing dead cells. Fungal suspensions (10^6 cells or conidia per mL) were either subjected or not to heat sterilization for 15 min at 121°C. Subsequently, both sets of

samples were treated with PMA (see below) prior to DNA extraction and then extracted DNA from viable/intact cells was quantified with ITS-region qPCR (Checinska Sielaff et al., 2019). Simultaneously, the effect of heat treatment and fungal viability were assessed by cultivation of heat-treated samples and subsequent CFU counts when compared to controls. As expected, PMA treatment resulted in no amplification of the ITS region following heat-sterilization and consequent differentiation between living and dead cells. Negative controls, such as sterile buffers and water were included in all experiments to avoid false-positive results.

Initial Sample Processing and DNA Extraction

Both antibiotic-treated and -untreated samples were divided into equal aliquots. One half of each set was treated with PMA dye (final concentration 25 μM) (Biotum, Inc., Hayward, CA) and the other half was left untreated following exactly the steps described elsewhere (Blachowicz et al., 2017). In brief, PMA-treated and untreated samples were kept in darkness for 5 min, followed by 15 min exposure to light in PHaST Blue-Photo activation system for tubes (GenIUL, S.L, Terrassa, Spain). Following dye intercalation, each sample was split in half. One aliquot was subjected to bead-beating for 60 s at 5 m/s on the Fastprep-24 bead-beating instrument (MP Biomedicals, Santa Ana, CA). Subsequently, both bead-beaten and not bead-beaten aliquots were combined and used for DNA extraction using Maxwell 16 Tissue LEV Total RNA purification kit following manufacturer's instructions (Promega, Madison, WI). As previously reported, the DNA purification was based on magnetic beads capturing and washing off non-DNA materials (Venkateswaran et al., 2012). Extracted and purified DNA was divided into aliquots and stored at −80°C until further analysis.

Fungal Community Analysis Using Amplicon Sequencing

Target amplification and sample index PCR reaction: Amplicons were prepared by adding 2 μL of DNA (less than 1 ng/μL) to a PCR mix containing 4X UCP Multiplex Master Mix and 1 μL of each panel pool primer. To cover the six 16S rRNA gene regions and ITS region, three-panel pool primers were used as instructed in the QIAseq16S-ITS kit (QIAGEN, Frederick, MD). The cycling conditions were set to an initial activation step of 95°C for 2 min, followed by 20 cycles of denaturation (95°C, 30 s), annealing (50°C, 30 s), and extension (72°C, 2 min), with a final extension step at 72°C for 7 min. After amplification, the three PCR reactions were combined into a single LoBind tube. PCR products were purified using the QIAGEN QIAseq beads at RT to remove all free unused primers.

Barcoded libraries were generated using QIAseq 16S/ITS Screening Panel kit, which utilizes phased primer technology. To index amplicons using sample indices and Illumina adaptors, 32.5 μL of the amplicons were added to a mix containing 12.5 μL of UCP master Mix and 4 μL of indices. The cycling conditions were set to an initial activation step at 95°C for 2 min followed by 14 cycles of denaturation (95°C, 30 s), annealing (60°C, 30 s), and extension (72°C, 2 min) with a final extension step at 72°C for 7 min. The libraries were purified using two rounds of QIAseq bead cleanup to remove excess adaptors.

Library QC and Quantification

The quality of barcoded libraries was determined by electrophoresis using TapeStation Bioanalyzer (Agilent, Santa Clara, CA). Libraries were quantified using QIAGEN QuantArray kit and adjusted to 2 nM, then pooled in equal volume. The pooled amplicons were diluted to 12 pM and sequenced using Illumina MiSeq and the protocol for paired-end (2 × 276 bp) sequencing V3 kit (Illumina, San Diego, CA).

Since the main focus of this study was to characterize the mycobiome, amplicon 16S rRNA sequences were not discussed. ITS sequences obtained from each PMA treated sample were compared with the UNITE database (version 8.1, released 2019-02-02) (Nilsson et al., 2019b) using NCBI BLASTn (version 2.6.0+, parameters: max_target_seqs 5, word_size 5, *e*-value 1e-5) (Altschul et al., 1990). The top hit to the database was retained for each sequence and a table containing the abundance of each ITS amplicon sequence types (STs) in each sample was generated. Canonical correspondence analysis (CCA) (Braak, 1986; Legendre and Legendre, 2020) from the Vegan R package (Oksanen et al., 2013) was used to analyze the distribution of samples with respect to their ST constituents. The variation in distribution among STs was analyzed using the sampling event as a linear predictor. The effect of variation present in control samples was removed prior to analysis.

Fungal Diversity Assessed With Shotgun Metagenomics

Samples for metagenome shotgun sequencing were prepared following the steps described previously (Singh et al., 2018). In brief, DNA libraries for shotgun metagenome sequencing were prepared and adapters were added using Nextera Flex DNA Library Preparation Kit from Illumina. The Agilent Fragment Analyzer (Agilent, Santa Clara, CA) was used to check the quality of prepared libraries. Libraries were normalized to 2 nM, pooled, denatured and diluted to 1.8 pM following recommendations by Illumina. The HiSeq 4000 platform (Illumina) was used for sequencing, resulting in 100-bp paired-end reads.

Received reads were trimmed with Trimmomatic (version 0.32) (Bolger et al., 2014) to remove adapters and low-quality ends. Reads shorter than 80 bp were discarded. Remaining reads were normalized (Nayfach and Pollard, 2016) and binned (Singh et al., 2018) to taxonomic levels (domains through species) using the lowest common ancestor (LCA) algorithm by MEGAN 6 (Huson et al., 2016). The samples were analyzed for temporal distribution (sampling 1 and 2) and spatial distribution (locations 1–10 in each cleanroom and differences between geographical locations) to assess succession and persistence using the following statistical approach. Microbial diversity analyses were carried out on normalized reads ($\sim 3.1 \times 10^8$) and analyses were set to keep at least one unique read to minimize the loss of diversity in low depth samples or for unique reads. BLAST hits of ≥ 20 amino acids and $\geq 90\%$ similarity were collected and used for taxonomic and functional assignments. MEGAN 6 (69) was used

for downstream processing and data visualization. The NCBI taxonomy base (Sayers et al., 2009), NCBI-NR protein database were used to assign taxonomic features using DIAMOND (Buchfink et al., 2015) and the weighted LCA algorithm of MEGAN 6 (Huson et al., 2016).

Analysis of similarities (ANOSIM) was carried out using the ANOSIM function from the R package vegan. Mann-Whitney-Wilcoxon (MWW) analyses were performed using the R function wilcox.test and a custom script. Venn diagrams were produced using the R package venneuler and a custom script.

Open access platform Cytoscape (Shannon et al., 2003) was used to perform network analysis and visualize the biogeographical interactions of viable metagenomic reads, while EggNog database (Powell et al., 2012; Huerta-Cepas et al., 2018) was utilized to perform functional analysis of metagenomic reads.

FUNGuild version 1.2 (Nguyen et al., 2016) was used to identify ecological guilds for the most abundant fungal taxa identified in the study. The abundance of members of each guild were summed and used to generate pie charts using GraphPad Prism 9.

Controls

To rule out the contamination, all handling and device controls were processed as the collected environmental samples. These controls include sampling devices, sterile PBS used to resuspend collected materials, reagent controls and nucleic acids extraction reagents. Since four different sampling events were carried out, controls were included during each sampling event. Additionally, all samples were processed in the biohood with laminar flow following aseptic techniques. Care was taken to wear proper personal protective equipments and change gloves frequently while working with the samples.

Data Availability

Amplicon and metagenome data are available under BioProject PRJNA641079, while ITS reads of cultivable isolates are available in GeneBank under MT704867–MT704931 accession numbers.

RESULTS

Cultivable Fungal Burden and Diversity

Cultivable fungal populations associated with various locations in two NASA cleanrooms were assessed during two sampling events of each SAF (**Supplementary Tables 1, 2**). The first sampling at JPL showed fungal counts ranging from 10^2 to 10^3 CFU/m^2; however, cultivable fungi were retrieved from only four of 10 sampled locations. During the second sampling, cultivable fungal counts ranged from 10^1 to 10^2 to CFU/m^2, but fungal populations were observed in all sampled locations (**Supplementary Table 1**). The ubiquitous isolation of fungi in JPL second sampling might be associated with different activities performed in the JPL-SAF cleanroom between performed samplings. At the time of the first sampling (April 17, 2018), the JPL-SAF was being prepared for Mars 2020 mission (M2020) with minimal human traffic, while more human activities were documented during the second JPL-SAF sampling (September 25, 2018) with rigorous cleaning regime. Fungal burden observed at KSC-PHSF was at levels $\sim 10^2$ CFU/m^2 and cultivable populations were retrieved from six and five out of 10 sampled locations (**Supplementary Table 2**).

Interestingly, using Clipper Mop (CM) to sample ten times more surface area (10 m^2) yielded similar fungal burden per meter square like for areas sampled with BiSKits or wipes (1 m^2). Additionally, when extracted samples were enriched with chloramphenicol (100 μg/mL for 24 h.) in the majority of the samples cultivable fungal counts were at the same level as in untreated ones. Only, two locations one from JPL (CM01—ClipperMop sample from location 1) and one from KSC (CM03—ClipperMop sample from location 3) SAF exhibited a onefold increase in fungal burden after antibiotic enrichment (**Supplementary Tables 1, 2**).

A total of 71 isolates were collected from both SAF and 65 were identified via sequencing the ITS region using NCBI and UNITE databases (Abarenkov et al., 2010), whereas ITS regions from the remaining six isolates were not successfully amplified. In total, 28 distinct species were observed, 16 in the JPL-SAF and 16 in the KSC-PHSF, while 13 isolates were potentially novel species due to low ITS sequence similarity to any known fungal species (**Supplementary Table 3**). These potentially novel species are subjects of follow up analyses to ascribe their phylogeny. Noteworthy, species identification, based on the ITS region, yielded some discrepancies in the identified species among both databases. Therefore, only when the whole genome is sequenced, annotated, and phenotype is determined, it is possible to definitely ascribe a species. As expected, given the geographic locations and differences in SAF maintenance, only four isolated species *Aureobasidium melanogenum*, *Penicillium fuscoglaucum*, *Penicillium decumbens,* and *Zalaria obscura* were present in both cleanroom facilities. The most prevalent species in the JPL-SAF were *Cladosporium austrohemisphaericum* (5 isolates) and *Arthrocladium tropicale* (3 isolates), while *Aspergillus pseudodeflectus* (7 isolates) and *Talaromyces veerkampii* (3 isolates) were predominant in KSC-PHSF. Forty-two isolates were collected from the samples not treated with antibiotic, while 23 were isolated after antibiotic treatment. Furthermore, when tested for UV-C resistance, 23 of the 65 isolates survived exposure to UV-C dose of 1,000 J/m^2 (**Supplementary Table 3**) and showed that UV-C resistance was strain specific. Lastly, neither field nor sampling controls, which underwent the same experimental and processing steps as investigated samples, showed any fungal growth.

Fungal Diversity Estimated by the Internal Transcribed Spacers Amplicon Sequencing

Samples collected from both cleanrooms were assessed for viable/intact (PMA treated) fungal populations in antibiotic-treated and untreated samples (**Supplementary Data Set 1**). The observed ITS amplicon reads and STs are shown in **Figure 1A** and **Supplementary Figures 3A,B**. Overall, higher diversity of viable fungal STs was observed in antibiotic treated samples in both sampled cleanrooms. The number of viable STs in KSC-PHSF was 4 and 6 in untreated samples

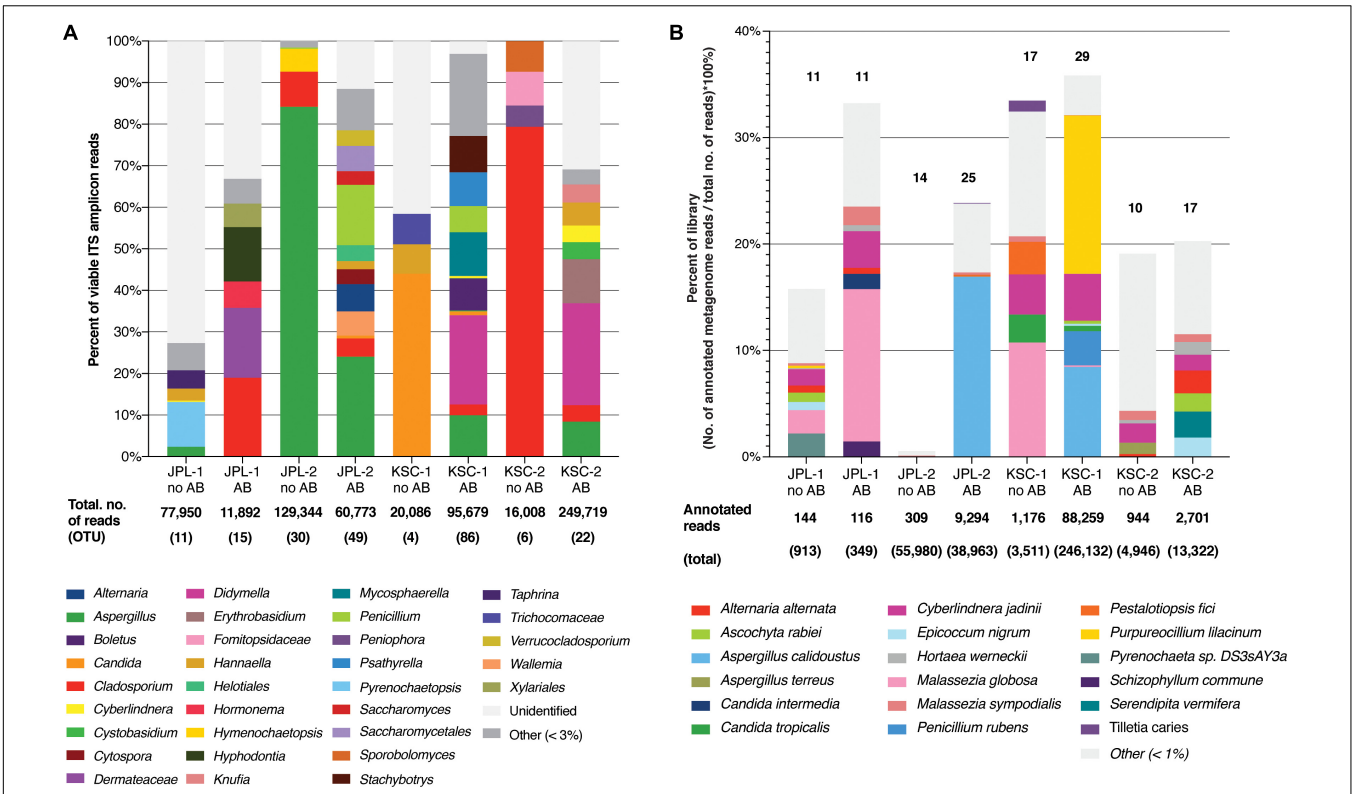


FIGURE 1 | Viable fungal community profiles before and after antibiotic treatment. **(A)** Fungal genera assessed by amplicon sequencing. Numbers of total ITS reads and OTUs for each cleanroom and sampling event are indicated below each bar. **(B)** Fungal species assessed by shotgun metagenomic sequencing. Numbers of total and annotated metagenome reads are listed below each bar. JPL-1 stands for sampling at JPL-SAF on 17 April, 2018; JPL-2 corresponds to the second sampling of JPL-SAF on 25 September, 2018; KSC-1 refers to KSC-PHSF sampling on 12 June, 2018 and KSC-2 corresponds to KSC-PHSF sampling on 24 July, 2018; AB/noAB indicates presence or absence of the antibiotic treatment.

and 86 and 22 in treated samples during the consecutive samplings. In JPL-SAF STs counts were 11 and 30 in untreated samples and 15 and 49 in treated samples during respective samplings (**Supplementary Figure 3A**). As expected, the total number of viable reads identified via ITS amplicon sequencing differed among sampled cleanrooms and consecutive sampling events (**Supplementary Figure 3B**). Noteworthy, the antibiotic treatment showed an impact on observed fungal diversity. The ITS region amplicon sequencing revealed the presence of 32 different fungal genera (**Figure 1A** and **Supplementary Table 4**) belonging mostly to phyla Ascomycota (65%) and Basidiomycota (31%). Enrichment with chloramphenicol resulted in retrieval of 5 and 8 STs not observed prior to the treatment at JPL and 11 and 7 STs at KSC during consecutive samplings. Conversely, several STs found prior to antibiotic treatment were not present following chloramphenicol-enrichment (**Figure 1A** and **Supplementary Table 4**). Importantly, the STs (with at least 100 reads per ST) extracted from handling and sampling device controls were compared to STs (with at least 100 reads per ST) from surface samples. This analysis showed that the "kitome," fungal STs innate to sampling devices, did not interfere with the mycobiome analysis of the samples. The STs observed in controls were different kind than the STs extracted from the environment, which was confirmed by GenBank accession numbers (**Supplementary Data Set 2**). Lastly, to assess the ecology of the identified genera FUNGuild analysis was performed (**Figure 2A**) revealing that ∼33% of the identified genera were not successfully ascribed to any guild category. However, among the successfully ascribed genera saprotrophs were the most dominant (∼17%) followed by endophytes (∼11), animal pathogens (∼8%), and plant pathogens (∼8).

The exact location-wise incidence of observed viable fungal taxa (PMA-treated) in antibiotic enriched samples is presented in **Figure 3** for JPL-SAF and **Figure 4** for KSC-PHSF. At JPL-SAF viable fungi were observed in only three locations (L7, L8, and L10) during the first sampling (April 17, 2018). However, all tested locations revealed fungal populations during the second sampling (September 25, 2018) as high mission activities were documented. Interestingly, both samplings at KSC-PHSF revealed high abundance of fungal populations at each sampled location (**Figure 4**).

Surface samples were also analyzed based on their ST constituents using CCA (**Figure 5**), which ordinates samples based on their ST composition similarity. Samples that measured alike ST in similar proportions cluster together, while samples that measured different communities occurred separately in the ordination. The analysis revealed that samples collected during one sampling event at either JPL-SAF or KSC-PHSF cleanrooms

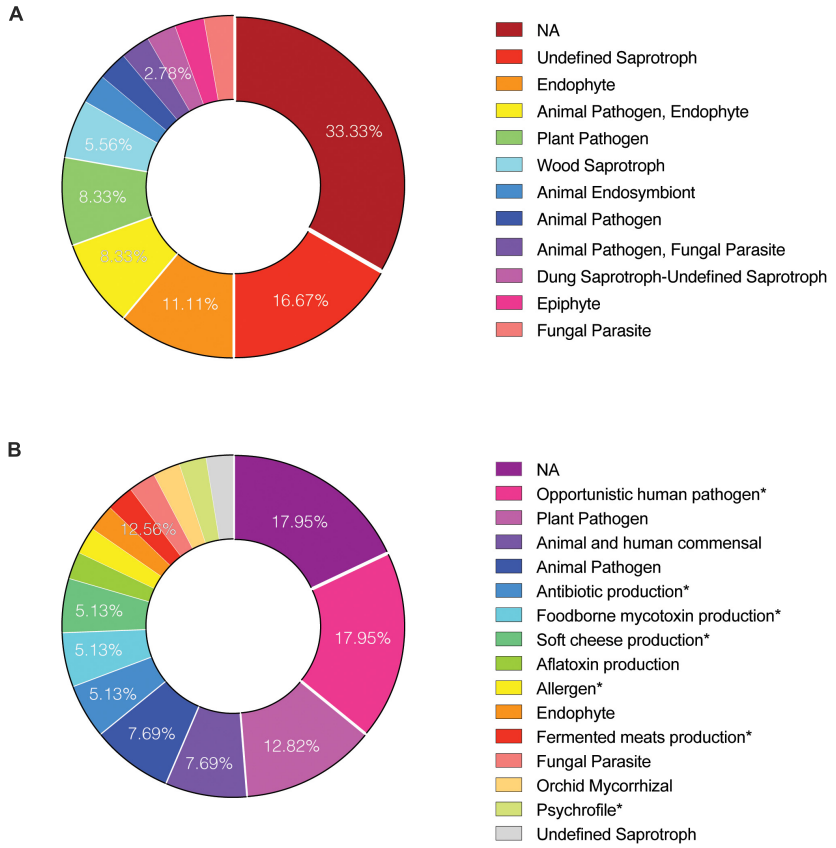


FIGURE 2 | Functional guilds of identified viable fungi. To determine the "ecology" of reported fungal genera and species FUNGuild analysis was performed on **(A)** amplicon and **(B)** metagenome reads, respectively. Identified guild categories are presented. NA, not applicable.

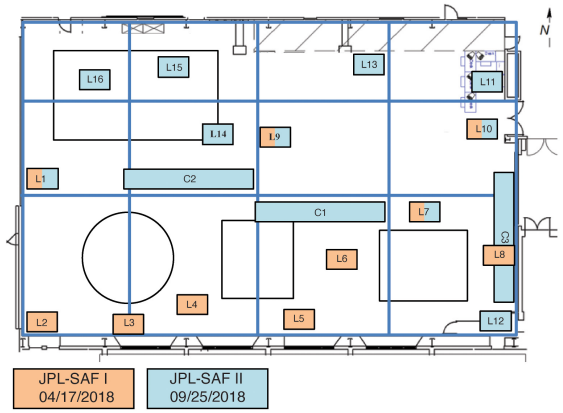
tend to cluster together in the CCA ordination and clustered separately from samples collected during different sampling events. Samples collected from KSC (blue and yellow glyphs) cluster independently, but more closely to each other than to the two groups of samples collected from JPL (red and green). Samples collected from the two sampling events at JPL are more disjointed from each other than samples collected from KSC, perhaps representing different community succession patterns in these two cleanrooms.

Lastly, the viable fungal diversity was compared and assessed using ANOSIM and MWW statistics (**Table 2**). When viable fungal populations were grouped based on the antibiotic treatment, there was a significant difference observed with MWW analysis ($p = 0.0386$), but not ANOSIM (**Table 2A**). The analysis of viable and antibiotic treated mycobiome constituents of each cleanroom and sampling event revealed some significant differences between JPL-1 vs. JPL-2 ($p = 0.007$) or JPl-1 vs. KSC-1 ($p = 0.004$); however, these may not be definitely ascribed to succession patterns, but rather to an exploratory nature of the study (**Table 1B**).

All PMA treated viable/intact samples, both with and without chloramphenicol treatment, were assessed for alpha and beta diversity using various microbial diversity indices, including Chao1 (**Figure 6Ai**), Shannon (**Figure 6Aii**), Simpson (**Figure 6Aiii**), and Venn diagrams showing species overlap (**Figure 6Aiv**). The abundance-based Chao1 estimator of alpha diversity showed slight increase in the fungal diversity in chloramphenicol treated samples. The Shannon estimator, which measures species richness and the Simpson estimator, which measures species evenness, revealed no differences upon chloramphenicol treatment. Lastly, the Venn diagram, which reveals the presence of unique species within each location and sampling event showed that only six different species were found in chloramphenicol-treated samples when compared to the chloramphenicol-untreated samples. In addition, the alpha and beta diversity were assessed location-wise for samples treated with antibiotic (**Figure 6B**). The observed differences in the indices confirmed no clear pattern in fungal succession within each sampled SAF (**Figure 6B**). Venn Diagram showed variations among sampled locations, which most likely stem from the exploratory nature of the study and geographic separation of the cleanrooms.

Fungal Diversity Estimated by Metagenome Shotgun Sequencing

After normalization across all samples, chloramphenicol- and PMA-treated or untreated ($n = 192$), the metagenomic analysis



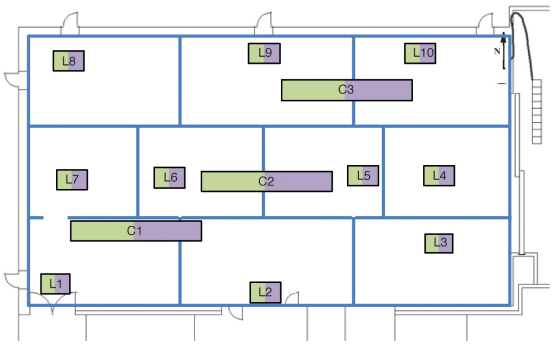
Viable fungal taxa with chloramphenicol (JPL-1)	Accession no.	No. of reads	Location 1	2	3	4	5	6	7	8	9	10
Ascomycota sp	KT264561	2,445								■		
Basidiomycota sp	KT819328	127							■			
Cladosporium delicatulum	HM148081	609										■
Cladosporium sp	KX784262	1,644										■
Dermateaceae sp	AM901772	2,000							■			
Dothiora sp	LT993343	208							■			
Hormonema sp	AF182378	760								■		
Hyphodontia sp	KJ140659	1,551										■
Lecanoromycetes sp	FJ553844	261							■			
Oedocephalum sp	KF227802	120							■			
Sakaguchia lamellibrachiae	KU171061	120							■			
Xylariales sp	MF359641	672										■
Unidentified fungi	AY843153	1,375							■	■		

Viable fungal taxa with chloramphenicol (JPL-2)	Accession no.	No. of reads	Location 1	7	9	10	11	12	13	14	15	16
Acremonium stromaticum	DQ825969	367		■						■		
Agaricomycetes sp	AB971703	358								■		
Alternaria sp	KT895930	3,963	■			■	■		■	■	■	■
Ascomycota sp	FN562038	164	■	■					■	■	■	■
Aspergillus calidoustus	HE616558	113				■				■		■
Aspergillus cibarius	JQ918177	729							■			
Aspergillus penicillioides	KY087743	1,812							■		■	
Aspergillus proliferans	EF652064	731		■					■			
Aspergillus pseudodeflectus	EF591743	3,245	■			■				■	■	■
Aspergillus subversicolor	JQ301894	4,189		■	■					■	■	■
Aspergillus sydowii	AB267812	3,647			■	■				■	■	■
Aspergillus versicolor	KR611591	150								■		■
Candida parapsilosis	KY418063	414	■	■	■	■			■	■	■	■
Capronia pilosella	DQ826737	248									■	
Chaetothyriales sp	KF675713	689									■	
Cladosporium delicatulum	HM148081	816							■			
Cladosporium dominicanum	DQ780353	1,829	■				■					
Cytospora eucalyptina	AY347375	2,191				■				■		
Dothideomycetes sp	DQ421226	1,600	■			■						■
Hannaella siamensis	AB922854	1,176	■									
Helotiales sp	JF449564	2,324	■				■					■
Malassezia restricta	AY743636	107				■						
Orbiliaceae sp	AY965757	221										■
Penicillium chrysogenum	AY213669	5,625	■		■	■	■		■	■	■	■
Penicillium concentricum	KC411763	135	■								■	■
Penicillium gladioli	AF033480	107	■			■					■	■
Penicillium goetzii	JX997091	140	■								■	■
Penicillium rubens	JX997057	859			■	■						
Penicillium sp	GQ120990	1,091	■		■	■					■	■
Penicillium vanluykii	JX997025	889	■			■					■	■
Platygloeales sp	UDB014788	356			■					■		
Saccharomyces cerevisiae	AB018043	1,982	■						■			
Saccharomycetales sp	HG935799	3,702		■		■	■	■		■		
Sarocladium bactrocephalum	HG965006	414		■					■			
Sarocladium strictum	AY214439	1,101							■			
Sordariomycetes sp	LC374624	602							■	■		
Verrucocladosporium dirinae	KF800501	2,265	■									
Wallemia tropicalis	KJ409878	3,565							■			
Unidentified fungi	KP893218	6,857	■	■		■	■		■	■	■	■

FIGURE 3 | Spatial distribution of viable fungi in samples treated with chloramphenicol at JPL-SAF. JPL-SAF sampling on 17 April, 2018 is marked in orange and on 25 September, 2018 in blue. The locations, marked L1-L16 and C1-C3, indicate areas sampled during the first and second sampling. Locations sampled in the second sampling had to be adjusted to account for on-going M2020 activities. Fungal taxa, GenBank accession #, sampling location, and ITS reads are indicated in the tables. L indicates locations sampled with BiSKit or wipe, while C indicates locations sampled with ClipperMop.

yielded ∼4.8 × 10^8 reads associated with microorganisms. Out of these reads, only the reads from viable and intact cells (reads from PMA treated samples) were the focus of this study (**Supplementary Data Set 3**). At the domain level 92.52% of the reads were bacterial, 7.43% eukaryotic, while both viral (0.02%) and archaeal (0.04%) reads were negligible. Such a distribution of metagenomic reads remains in sharp contrast to other systems, including the ISS, homes and offices where about 30–50% of the reads belong to eukaryotes (Singh et al., 2018). Among the 3.6 × 10^7 eukaryotic reads detected, only 2.1 × 10^6 belonged to fungi and were further characterized and are presented in this manuscript. Furthermore, about 13–35% of all reads obtained from both cleanrooms were successfully annotated, with the exception of samples collected from JPL during the second sampling and not treated with chloramphenicol, where only 1% of the reads were annotated (**Figure 1B** and **Supplementary Figure 3C**).

Metagenomic fungal reads were dominated by Ascomycota (91%) and Basidiomycota (7.15%); however, representative reads of Blastocladiomycota, Chytridiomycota, Cryptomycota, Mucoromycota, and Zoopagomycota were also retrieved. Similar to amplicon sequencing (**Figure 1A** and **Supplementary Figure 3B**), the total number of fungal reads differed among analyzed samples. In total, 32 fungal species were identified using shotgun metagenome sequencing. The number of viable fungal species differed among antibiotic untreated and treated samples. At KSC-PHSF it was, respectively, 10 and 17 identified species during the first sampling and 10 and 17 during the second sampling. At the JPL-SAF, the difference was observed during the second sampling only, when the number of species changed from 14 to 25, while the number of species remained the same during the first sampling (**Figure 1B** and **Supplementary Table 5**). Among the most dominant species at both cleanrooms were *Malassezia globosa* (human skin and hair-associated fungus) and *Cyberlindnera jadinii* (teleomorph of *Candida utlilis*). Noteworthy, the antibiotic treatment resulted in retrieving *Aspergillus calidoustus* and other *Aspergillus* species at JPL-SAF, and *Penicillium* species, *Purpureocillium lilacinum*, *Alternaria alternata, Epicoccum nigrum, and Serendipita vermifera* at KSC-PHSF when compared to untreated samples (**Supplementary Table 5**). To visualize the biogeography of isolated strains the network analysis was computed for chloramphenicol-treated samples (**Figure 7A**). This analysis confirmed higher abundances of *M. globosa* and *C. jadinii* marked with thicker edges as oppose to less abundant isolates represented with finer edges. Detected reads were also



Viable fungal taxa with chloramphenicol (KSC-1)	Accession no.	No. of reads	Location 1	2	3	4	5	6	7	8	9	10
Alternaria porri	KJ174426	136				■			■	■	■	
Ascomycota sp	FN562038	674			■	■	■		■	■	■	■
Aspergillus calidoustus	HE616558	323	■								■	■
Aspergillus fischeri	EF669936	430					■					
Aspergillus fumigatus	KT358863	764					■					
Aspergillus insuetus	EU076356	130									■	■
Aspergillus keveii	EU076354	136									■	■
Aspergillus porphyreostipitatus	KJ775564	168									■	
Aspergillus pseudodeflectus	EF591743	7,560	■	■				■		■	■	■
Aureobasidium pullulans	AJ244232	1,558							■		■	■
Boletus campestris	GQ166879	7,300				■						
Candida parapsilosis	KY418063	841	■	■	■	■	■		■	■	■	■
Cladosporium cladosporioides	EU714397	355		■					■	■	■	■
Cladosporium delicatulum	HM148081	940		■	■				■	■	■	■
Cladosporium dominicanum	DQ780353	496										■
Cladosporium sp	KP401885	669							■	■	■	■
Cordycipitaceae sp	LC092894	2,515				■						
Cosmospora gigas	EF121863	520										■
Cyberlindnera jadinii	DQ249199	501	■		■							
Cystobasidium slooffiae	KX079906	242								■		
Didymella gardeniae	FJ427003	13,757				■	■		■			
Didymella glomerata	FJ427013	342							■			
Didymella nigricans	GU237867	6,493				■	■		■			■
Dothideomycetes sp	DQ421226	172								■		
Epicoccum brasiliense	GU237760	177				■			■			■
Filobasidium sp	EU266556	258							■			
Fusarium sp	GU721385	107							■		■	
Fusicolla violacea	JN198450	519										■
Gibberella intricans	GQ505688	100							■		■	
Gibbosporina bifrons	KM887878	126									■	
Hortaea werneckii	AJ238468	230								■	■	
Lecanoromycetes sp	KM624596	298							■			

Viable fungal taxa with chloramphenicol (KSC-1 cd.)	Accession no.	No. of reads	Location 1	2	3	4	5	6	7	8	9	10
Mycosphaerella tassiana	EF679363	10,042		■				■	■		■	■
Myrmecridium schulzeri	KU574706	253									■	
Myrmecridium sp	JQ081655	208									■	
Naganishia diffluens	AF145330	2,091		■								
Nectriaceae sp	MF398836	2,695		■	■				■			■
Nigrospora oryzae	DQ219433	272							■			
Paraconiothyrium brasiliense	KJ767102	184									■	
Paraphaeosphaeria sp	KF227869	144							■			
Penicillium bialowiezense	EU587315	142							■			
Penicillium chrysogenum	JQ693404	3,524	■	■				■	■	■	■	■
Penicillium concentricum	KC411763	132										
Penicillium gladioli	AF033480	111	■						■	■	■	■
Penicillium rubens	JX997057	671	■						■	■	■	■
Penicillium sp	GQ120990	822	■						■	■	■	■
Penicillium vanluykii	JX997025	649	■						■	■	■	■
Phaeosphaeriaceae sp	HQ631018	1,001									■	
Pleosporales sp	GU581261	651							■			■
Polyporales sp	HM136871	132										■
Psathyrella sp	KT186173	6,025				■					■	
Psathyrella trinitatensis	KC992882	157				■						
Psathyrella tuberculata	MH497604	1,603				■						
Rhodosporidiobolus odoratus	EU002899	667									■	
Rhodotorula dairenensis	KJ706797	396										■
Rhodotorula mucilaginosa	KJ706479	513										■
Sordariomycetes sp	LC374624	390				■			■		■	■
Stachybotryaceae sp	KX515901	194							■			
Stachybotrys echinatus	MG430270	7,306				■	■		■	■		
Stachybotrys microspora	LC228682	1,021				■				■		
Symmetrospora vermiculata	AB030335	228										■
Talaromyces ruber	JX315662	344									■	
Teratosphaeriaceae sp	KT833169	626							■			
Toxicocladosporium cacti	KY752809	1,134							■			
Vishniacozyma dimennae	EU852364	245										■
Unidentified fungi	KF800326	2,269	■		■	■		■	■	■	■	■

Viable fungal taxa with chloramphenicol (KSC-2)	Accession no.	No. of reads	Location 1	2	3	4	5	6	7	8	9	10
Allophoma piperis	GU237816	104		■								■
Aspergillus cibarius	JQ918177	10,346		■		■		■			■	■
Aspergillus proliferans	EF652064	10,386		■		■		■			■	
Aspergillus ruber	EF652066	165						■				
Clavispora lusitaniae	KP674846	10,054		■	■	■		■		■	■	
Cyberlindnera jadinii	DQ249199	10,065	■	■	■	■	■	■			■	
Cystobasidium slooffiae	KX079906	10,086		■					■			■
Debaryomyces hansenii	AB053101	442			■							
Didymella gardeniae	FJ427003	44,525		■		■		■		■	■	
Didymella nigricans	GU237867	16,608		■				■		■	■	
Erythrobasidium hasegawianum	AB030352	26,604					■					
Hannaella oryzae	AF314986	13,786	■	■	■	■		■	■	■	■	
Knufia perforans	AJ244275	10,795		■				■	■		■	
Pseudotaeniolina globosa	AY128700	7,128		■		■			■			
Resupinatus sp	KP026232	1,330		■								■
Serendipita sp	HQ154304	112							■			
Unidentified fungi	KX515713	77,183	■	■	■	■	■	■	■	■	■	■

FIGURE 4 | Spatial distribution of viable fungi in samples treated with chloramphenicol at KSC-PHSF. KSC-PHSF sampling on 12 June, 2018 is marked in green and on 24 July, 2018 in violet. A schematic diagram depicts from where samples were collected during both sampling events. Locations are marked as L1–L10 and C1–C3. Fungal taxa, GenBank accession #, sampling location, and ITS reads are indicated in the tables. L indicates locations sampled with BiSKit, while C indicates locations sampled with ClipperMop.

assigned to specific sampling locations within the JPL-SAF and KSC-PHSF (**Figure 7B**). Overall metagenomic reads retrieved from handling and sampling device controls were scarce. However, in case of observing more abundant reads in controls, the matching ST reads in the samples were excluded from downstream analyses (*M. globosa* JPL-2 and KSC-2 and *C. jadinii* JPL-2) (**Supplementary Data Set 4**). The insight into "ecology" of reported species assessed either via FUNGuild or literature search (**Figure 2B**) revealed that ∼18% of identified species were not successfully ascribed to any guild category. Among the ascribed categories opportunistic pathogens (∼18), plant pathogens (∼13%), animal and human commensals (∼8%), and animal pathogens (∼8%) were predominant.

Similar to ITS amplicon reads, viable fungal diversity at the species level estimated by the shotgun metagenomic was analyzed using ANOSIM and MWW statistics (**Table 3**). When viable fungal populations were grouped based solely on the antibiotic treatment, there was a significant difference observed with ANOSIM analysis, but not in MWW (**Table 2A**). When PMA and antibiotic treated mycobiome constituents of each cleanroom and sampling event were pooled, some statistical differences were observed amongst the locations, but again these not revealed any succession patterns just the expected distinction related to geographical separation (**Table 2B**).

All viable samples (PMA treated), with and without chloramphenicol treatment, were assessed for alpha and beta diversity (**Figure 8A**). The Chao1 estimator showed an increase in the fungal diversity in chloramphenicol treated samples (**Figure 8Ai**). In contrast, the Shannon and Simpson indicators revealed a decrease in fungal diversity upon chloramphenicol treatment (**Figures 8Aii,iii**). Lastly, Venn diagrams showed that only nine species were unique for chloramphenicol-treated samples when compared to the untreated ones (**Figure 8Aiv**). The fungal diversity assessed for locations using antibiotic-treated samples showed no definite patterns allowing for assessing fungal diversity in consecutive samplings, as the observed trends were opposite in sampled cleanrooms (**Figure 8B**). Lastly, Venn diagrams showed that eight species—namely *Alternaria alternata, Aspergillus sydowii, Cyberlindnera jadinii,*

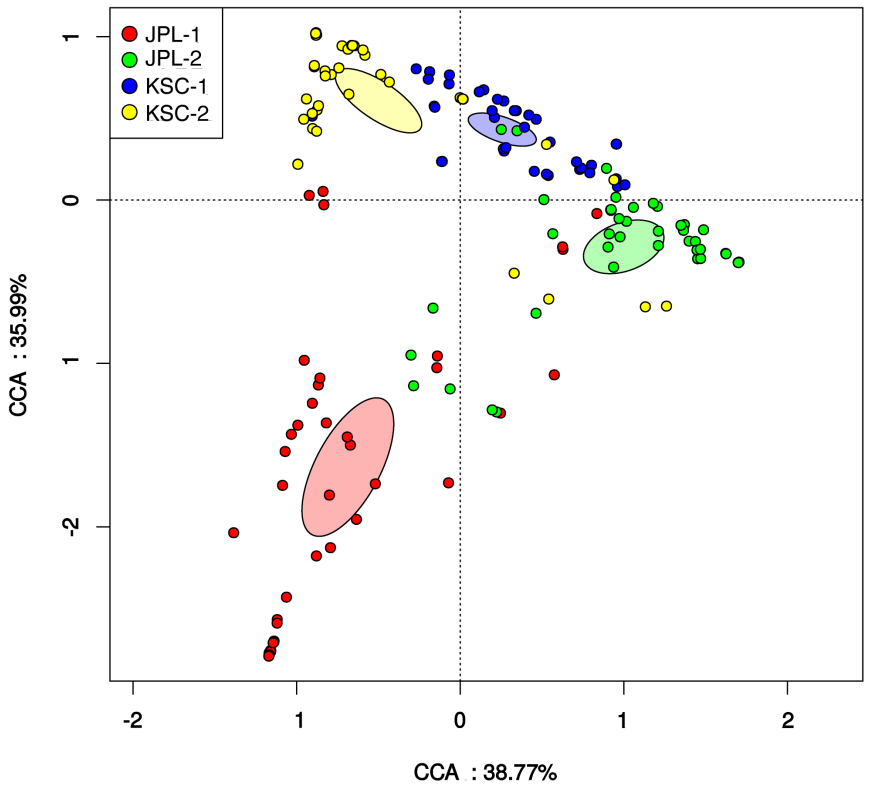


FIGURE 5 | Canonical correspondence analysis (CCA) revealing relationships among fungal populations and associated cleanroom environments. Presented is the distribution of viable fungal taxa in both cleanrooms during consecutive sampling events assessed by amplicon sequencing. Glyphs are colored based on the sampling event. Colored ellipses are included for each sampling event to represent the standard error of the mean of each cluster (99% confidence). CCA axes include the percent variation described by the axis. JPL-1 stands for sampling at JPL-SAF on 17 April, 2018; JPL-2 corresponds to the second sampling of JPL-SAF on 25 September, 2018; KSC-1 refers to KSC-PHSF sampling on 12 June, 2018 and KSC-2 corresponds to KSC-PHSF sampling on 24 July, 2018.

TABLE 2 | ANOSIM and Mann-Whitney-Wilcoxon analyses of viable amplicon reads at genus level of all antibiotic treated and not treated samples **(A)** and samples separated by location and sampling event with chloramphenicol **(B)**.

(A) Antibiotic vs. no antibiotic.

MWW	**ANOSIM**
W = 830.5	*R*: 0.007428
p = 3.86E-02	*p* = 0.274

(B) Location and sampling event in samples with chloramphenicol.

Location	**JPL-1**	**JPL-2**	**KSC-1**	**KSC-2**	
JPL-1		*W* = 663 *p* = 7.35E-03	*W* = 720 *p* = 2.79E-02	*W* = 692 *p* = 1.57E-02	
JPL-2	*R* = 0.2594 *p* = 0.007		*W* = 1038 *p* = 5.48E-01	*W* = 1023.5 *p* = 6.37E-01	**MWW**
KSC-1	*R* = 0.2668 *p* = 0.004	*R* = 0.012111 *p* = 0.316		*W* = 1043 *p* = 5.21E-01	
KSC-2	*R* = 0.04844 *p* = 0.2	*R* = 0.02089 *p* = 0.323	*R* = 0.07022 *p* = 0.104		
		ANOSIM			

Hortaea werneckii, Malassezia globosa, M. pachydermatis, M. sympodialis, and *Penicillium steckii*—were common to all sampling events and the locations in antibiotic treated samples (**Figure 8Biv**).

Noteworthy, characterization of metagenomic reads revealed the prevalence of genes involved in processes related to chromatin structure and dynamics, replication, recombination and repair and signal transduction mechanisms

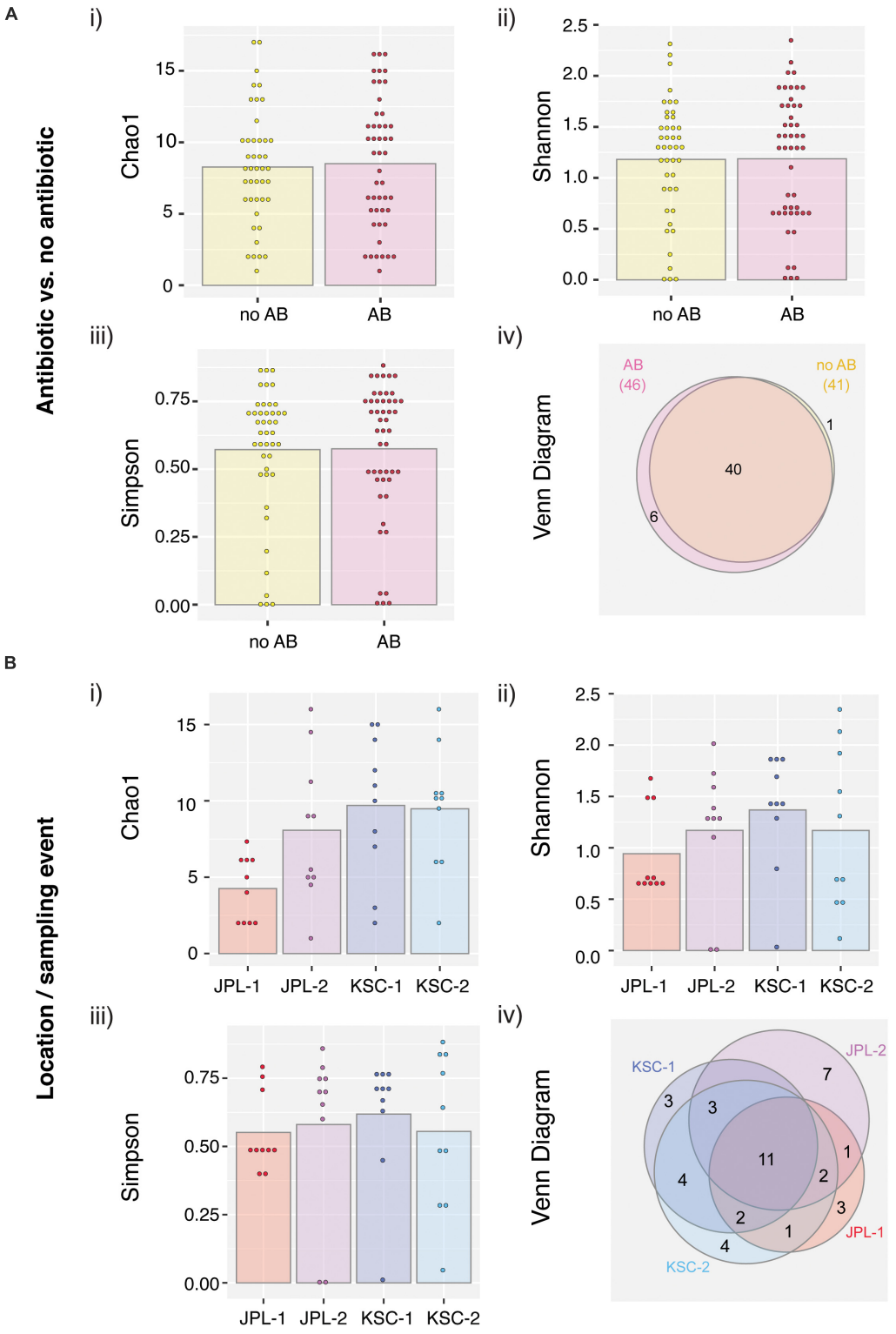


FIGURE 6 | Viable fungal taxa ordinate analysis of samples from JPL-SAF and KSC-PHSF. **(A)** To determine the impact of the chloramphenicol treatment on the fungal diversity, normalized amplicon reads of antibiotic treated and untreated samples were analyzed. Tested indices include **(i)** Chao1 diversity analysis **(ii)** Shannon index **(iii)** Simpson index, and **(iv)** Venn diagram. **(B)** Location/sampling event-based diversity was assessed using amplicon reads of antibiotic treated samples only. Tested indices include **(i)** Chao1 diversity analysis, **(ii)** Shannon index, **(iii)** Simpson index, and **(iv)** Venn diagram showing the distribution of viable fungal taxa. JPL-1 stands for sampling at JPL-SAF on 17 April, 2018; JPL-2 corresponds to the second sampling of JPL-SAF on 25 September, 2018; KSC-1 refers to KSC-PHSF sampling on 12 June, 2018 and KSC-2 corresponds to KSC-PHSF sampling on 24 July, 2018; AB/noAB indicates presence or absence of the antibiotic treatment.

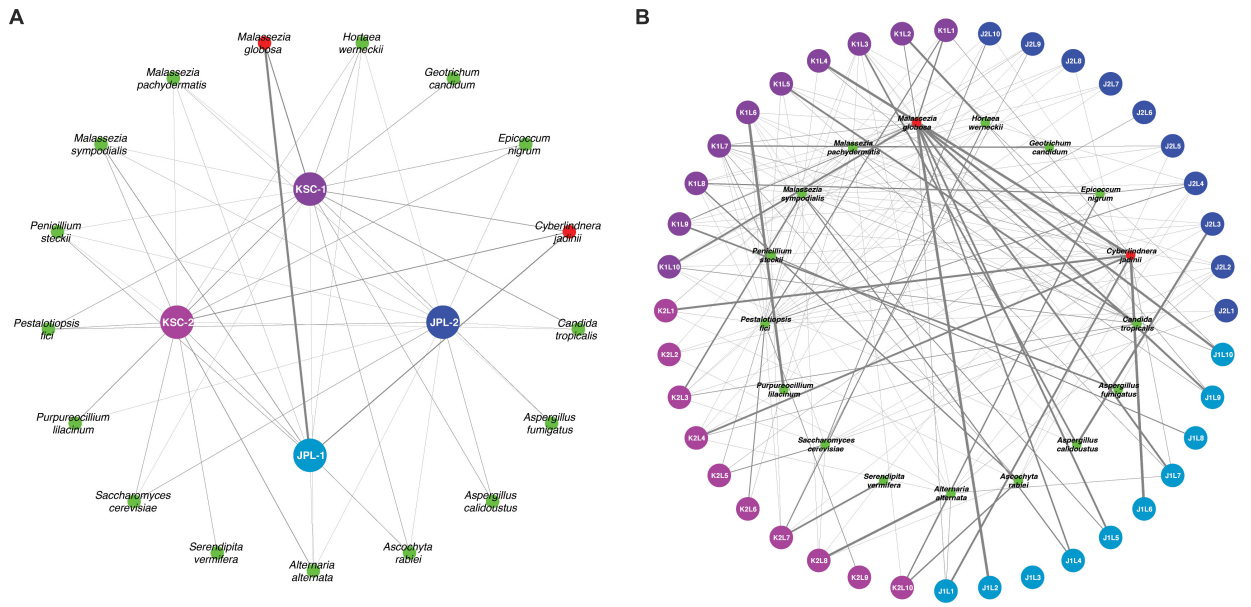


FIGURE 7 | Network analysis presenting biogeographical distribution of viable fungal populations after antibiotic treatment. **(A)** Presented are the most abundant (abundance > 0.01% of all reads) fungal species in JPL-SAF and KSC-PHSF. **(B)** The most abundant fungal reads are assigned to sampled locations. The thickness and transparency of the edges (lines) follows the calculated e-weights. JPL-1 stands for sampling at JPL-SAF on 17 April, 2018; JPL-2 corresponds to the second sampling of JPL-SAF on 25 September, 2018; KSC-1 refers to KSC-PHSF sampling on 12 June, 2018 and KSC-2 corresponds to KSC-PHSF sampling on 24 July, 2018. In **(B)**, sampling locations are marked using the system that indicates sampled cleanroom at JPL or KSC marked as J or K, first or second sampling event indicated by 1 or 2, followed by sampled location labeled as L1–L10, for example, J1L2 stands for JPL-SAF first sampling at location 2. The red colored taxa were more abundant in control samples.

TABLE 3 | ANOSIM and Mann-Whitney-Wilcoxon analyses of viable metagenome reads at species level of all antibiotic treated and not treated samples **(A)** and samples separated by location and sampling event with chloramphenicol **(B)**.

(A) Antibiotic vs. no antibiotic.

MWW	**ANOSIM**
W = 124	R: −0.01645
p = 1.989E-10	p = 0.881

(B) Location and sampling event in samples with chloramphenicol.

Location	**JPL-1**	**JPL-2**	**KSC-1**	**KSC-2**	
JPL-1		W = 229 p = 2.57E-08	W = 36.5 p = 9.495E-14	W = 466 p = 1.16E-03	**MWW**
JPL-2	R = 0.2734 p = 0.004		W = 323 p = 1.245E-05	W = 995.5 p = 1.77E-02	
KSC-1	R = 0.08628 p = 0.061	R = −0.08178 p = 0.842		W = 1302 p = 5.56E-08	
KSC-2	R = 0.1259 p = 0.032	R = 0.1002 p = 0.124	R = 0.1557 p = 0.052		
		ANOSIM			

in all samples except JPL-1 treated with chloramphenicol (**Figure 9** and **Table 4**). When chloramphenicol-treated samples were compared to untreated ones the change in transporter proteins (ENOG410XNQK, COG1131, COG1132) was observed (**Table 4**). However, metatranscriptomic analysis is required to confirm dynamic changes in gene expression in response to the treatment with antibiotic.

DISCUSSION

Successful isolation of fungi is necessary for determining and characterizing their unique molecular features and biotechnological value, which might enable development of countermeasures (Knox et al., 2016; Romsdahl et al., 2018b, 2020; Urbaniak et al., 2019). Therefore, seeking for more efficient isolation methods of fungi from the cleanroom environments

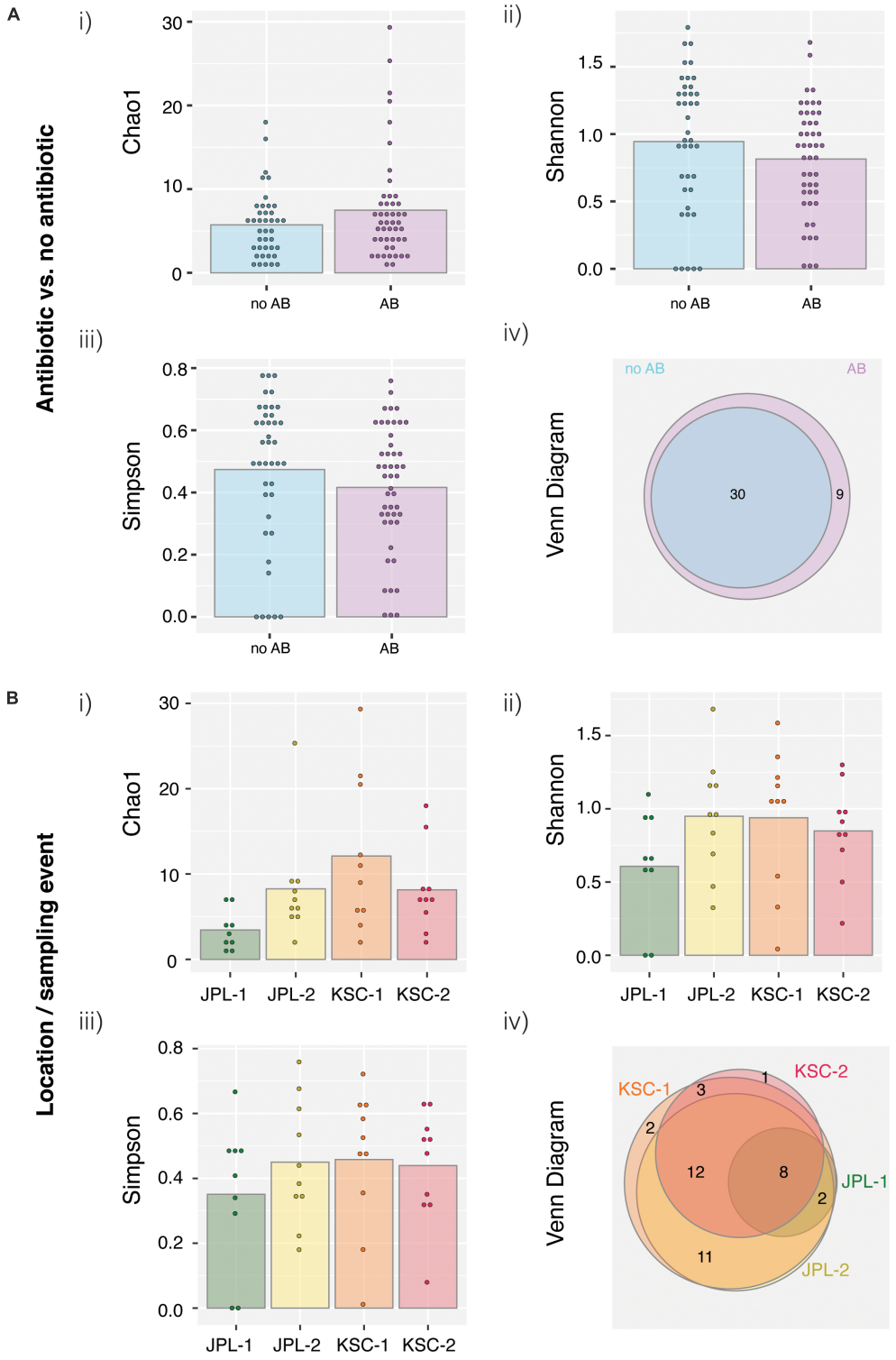


FIGURE 8 | Species-level ordinate analysis of samples from JPL-SAF and KSC-PHSF cleanrooms. **(A)** To determine the impact of the chloramphenicol treatment on the fungal diversity, normalized metagenomic reads of antibiotic treated and untreated samples were analyzed. Tested indices include **(i)** Chao1 diversity analysis **(ii)** Shannon index **(iii)** Simpson index, and **(iv)** Venn diagram. **(B)** Location/sampling event-based diversity was assessed using metagenomic reads of antibiotic treated samples only. Tested indices include **(i)** Chao1 diversity analysis, **(ii)** Shannon index, **(iii)** Simpson index, **(iv)** Venn diagram showing the distribution of viable fungal taxa. JPL-1 stands for sampling at JPL-SAF on 17 April, 2018; JPL-2 corresponds to the second sampling of JPL-SAF on 25 September, 2018; KSC-1 refers to KSC-PHSF sampling on 12 June, 2018 and KSC-2 corresponds to KSC-PHSF sampling on 24 July, 2018; AB/noAB indicates presence or absence of the antibiotic treatment.

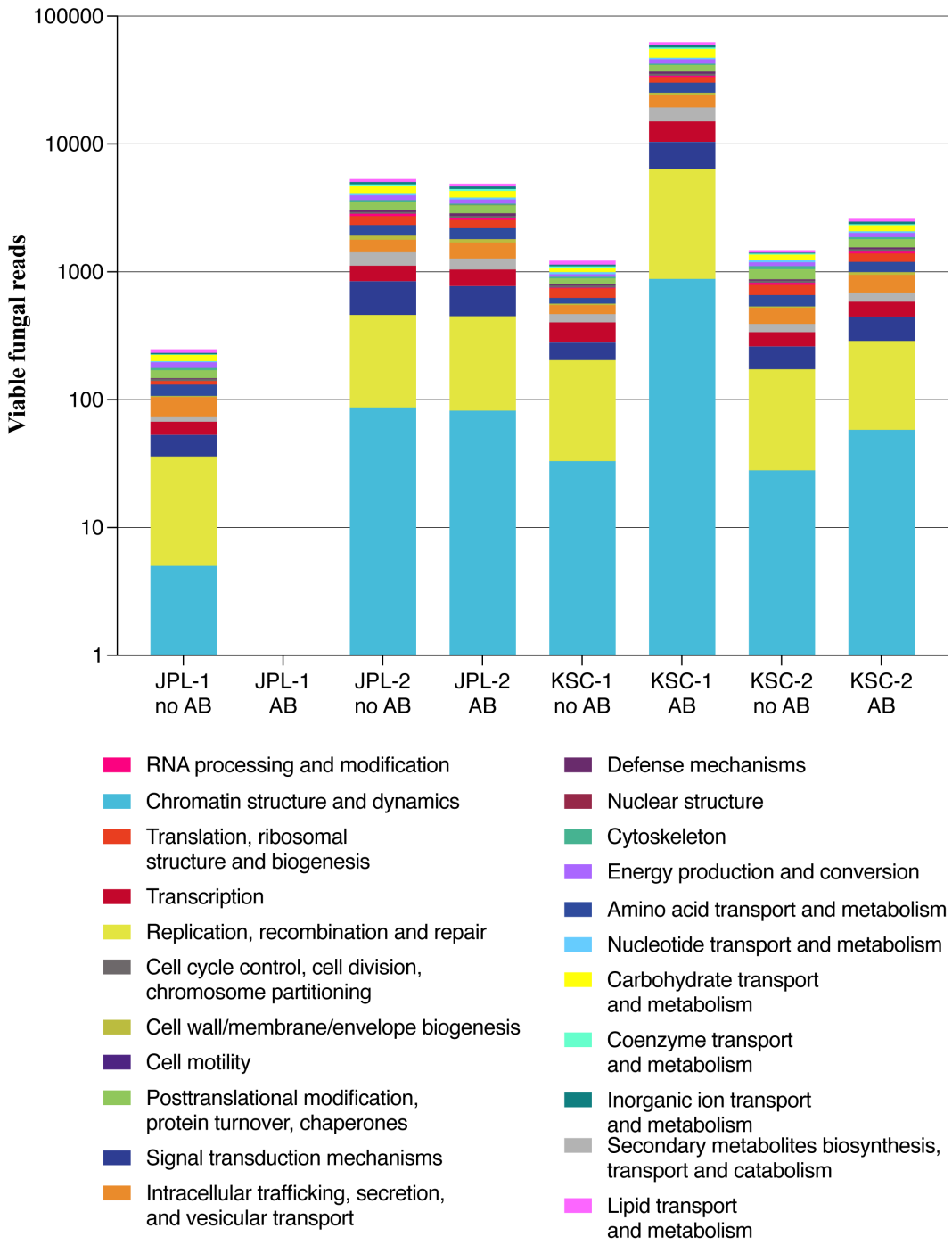


FIGURE 9 | Analysis of metagenomic reads from JPL-SAF and KSC-PHSF cleanrooms. Presented are the most abundant genes/proteins in both cleanrooms during consecutive sampling events assessed by shotgun metagenomics. JPL-1 stands for sampling at JPL-SAF on 17 April, 2018; JPL-2 corresponds to the second sampling of JPL-SAF on 25 September, 2018; KSC-1 refers to KSC-PHSF sampling on 12 June, 2018 and KSC-2 corresponds to KSC-PHSF sampling on 24 July, 2018; AB/noAB indicates presence or absence of the antibiotic treatment.

is critical. In this study chloramphenicol was paired-with PMA treatment to assess the amplifying effect on fungal isolation. In the course of the study 13 potentially novel fungal species have been isolated, but only six were unique to the chloramphenicol-treated samples, showing moderate benefit of antibiotic use. More importantly, characterization of these potentially novel species might lead to the understanding of how fungi thrive in harsh, oligotrophic environments.

TABLE 4 | Functional analysis of viable metagenome reads of all antibiotic treated and not treated samples.

EggNog	Putative function/activity	JPL-1 no AB	JPL-1 AB	JPL-2 no AB	JPL-2 AB	KSC-1 no AB	KSC-1 AB	KSC-2 no AB	KSC-2 AB
ENOG410XNQK	Transporter	4	0	77	87	5	1,647	32	31
COG0477	Major facilitator Superfamily	6	0	65	57	10	1,369	11	29
COG2801	Retrotransposon protein	0	0	5	8	52	1,377	5	4
COG2124	Cytochrome p450	0	0	99	68	12	994	13	30
COG3321	Synthase	0	0	27	29	13	944	4	15
COG1020	Non-ribosomal peptide synthetase	0	0	31	20	4	774	2	4
COG0531	Amino acid	6	0	37	23	0	561	15	13
COG1131	(ABC) transporter	0	0	37	37	9	532	0	17
COG1132	(ABC) transporter	0	0	39	50	6	445	11	20
COG0474	P-type ATPase	2	0	20	17	12	399	7	18
COG5032	Phosphatidylinositol kinase	3	0	25	30	9	376	8	22
COG1472	Hydrolase family 3	0	0	35	17	14	334	4	2
COG1012	Dehydrogenase	0	0	11	20	1	355	2	15
COG2303	Oxidoreductase	7	0	16	17	0	333	8	17
COG0654	Xidoreductase	2	0	19	19	6	317	2	10
COG2072	Monooxygenase	0	0	9	10	6	338	1	2
COG0366	Alpha amylase, catalytic	2	0	17	12	0	306	5	9
COG0464	AAA ATPase	1	0	25	40	0	269	8	5
COG0604	Alcohol dehydrogenase	0	0	18	11	0	284	2	10
COG1215	Glycosyl transferase, family	0	0	42	30	6	211	1	14
COG1112	Helicase	3	0	14	30	0	247	1	8
COG0277	FAD linked oxidase	0	0	6	9	4	258	2	11
COG1643	Helicase	1	0	43	36	4	189	4	10
COG0318	Amp-dependent synthetase and ligase	2	0	31	22	0	214	5	7
COG5059	Kinesin family member	0	0	8	12	9	235	7	6
ENOG410XRBH	SRSF protein kinase	0	0	27	11	0	235	0	3
COG3320	Domain protein	1	0	16	4	8	230	3	0
COG1501	Hydrolase, family 31	0	0	6	8	1	238	2	6
COG2272	Carboxylesterase	0	0	6	5	0	243	0	1
ENOG410XNPJ	Polyketide synthase	0	0	1	2	7	240	0	2

Interestingly, despite using growth media containing varying dose of the chloramphenicol (DRBC vs. PDA) no differences in isolated fungal species nor abundance were observed. Among cultured fungi, *Penicillium, Aspergillus,* and *Cladosporium* species were predominant, which were also reported on the Skylab (Brockett et al., 1978), Mir Space Station (Alekhova et al., 2005), Japanese ISS module "KIBO" (Satoh et al., 2016), and ISS high-efficiency particle air filters (HEPA) debris (Checinska et al., 2015). In contrast, ISS environmental surfaces near where humans dwell on the spacecraft, were dominated by *Rhodotorula* and other members of *Sporiodiobolaceae* (Checinska Sielaff et al., 2019). Fungal burden assessed by ITS amplicon sequencing revealed variations in fungal abundance at JPL-SAF during increased mission activity, suggesting humans as vectors and the source of the contamination. Such correlation between human presence and changes in fungal populations in confined spaces has been previously reported (Checinska et al., 2015; Blachowicz et al., 2017; Checinska Sielaff et al., 2019).

One of the main goals of the study was to overcome the overwhelming abundance of bacterial populations, which overshadow the fungal presence, while performing molecular analyses. Pairing the antibiotic and PMA treatment with targeted amplification revealed presence of rare mycobiome (Dolan, 1971; Nilsson et al., 2019a). However, the positive impact of the antibiotic itself on retrieved fungal diversity was not definitive, which may be related to the fact that chloramphenicol is considered bacteriostatic rather than bactericidal (Oong and Tadi, 2022). Even though no mission-related activities or human traffic were documented at KSC-PHSF, the cleanroom was maintained to the ISO-7 certification requirements. The ISO-7 cleanroom classification must have less than 352,000 particles > 0.5 micron per cubic meter and 60 HEPA filtered air changes per hour. Interestingly, irrespective of human traffic and stringent maintenance, there was high fungal diversity during both samplings. Substantial differences in fungal abundance in both JPL and KSC cleanrooms might be attributed to the exploratory nature of the study and several other factors, like different geographical locations, climates (relative humidity) and in-house cleaning procedures used to maintain the cleanrooms. The analysis of the STs using CCA showed that samples collected during the same sampling event tend to cluster together and separately from samples collected during different sampling events in both cleanrooms.

This might be due to the biogeographical attributes (Moissl et al., 2007) corresponding to each facility since human traffic related to mission activities and the exchange of clean air into the facility were occurring. In addition, differences in spatial distribution of constituents of the mycobiome was documented. For instance, JPL-SAF location L1 had no viable fungi reported after the first sampling, while high fungal abundance were observed during the second sampling. This may be related to the fact that the location L1 is the only entry point for cleanroom personnel, while the L12 location, which showed extremely low fungal presence, was not heavily used during on-going M2020 activities. Similarly, increased human traffic was documented in locations L13–L16, which corresponds with high fungal presence. Inversely, other sampled locations (L7–L11) exhibited both lower human traffic and lower fungal abundance. This was of no surprise since fungal communities have been reported to be affected by human presence in a simulated closed habitat (Blachowicz et al., 2017) and aboard the ISS (Checinska Sielaff et al., 2019).

One of the aspects that is crucial to address is how the fungal diversity observed in cleanrooms corresponds to the fungal populations in other confined spaces. So far, the majority of the available reports were based on the cultivable approach, while culture-independent mycobiome analyses are scarce. When the mycobiome of the simulated closed habitat ILMAH was assessed during 30-day occupation by student astronauts, the dominant genera included *Epicoccum, Alternaria,* and family *Pleosporaceae* (Blachowicz et al., 2017), while in the spacecraft assembly cleanrooms the most dominant members were *Aspergillus, Cladosporium,* and *Didymella* (La Duc et al., 2012). Interestingly, the ecological association of fungi reported in JPL-SAF and KSC-PHSF revealed dominance of saprothrophs, endophytes, and animal pathogens. In contrast to fungal populations usually associated with the natural outdoor environment in ILMAH and examined cleanrooms, the surveillance of the ISS surfaces revealed the presence of fungi associated more with humans rather than with outdoor-associated environments (Checinska Sielaff et al., 2019). Similarly, fungi isolated from neonatal environmental surfaces of intensive care units composed of different kinds of fungi than observed in the cleanrooms with overwhelming presence of *Candida* and *Saccharomyces* (Heisel et al., 2019). The facility-specific fungal composition might be attributed to the cleaning management and system engineering of the controlled environment.

Metagenomic analysis of collected samples provided higher resolution for species identification. Eight identified species were present in both cleanrooms revealing that despite the geographical distance certain core species remain the same, including species of environmental origins (*Alternaria alternata, Aspergillus sydowii)* and human commensals (*Malassezia globosa, M. pachydermatis, M. sympodialis*). Noteworthy, the mycobiome associated with environmental surfaces in the crewed ISS was dominated by *Penicillium* species and *Aspergillus calidoustus* (Singh et al., 2018), while high abundance of *Cyberlindnera jadinii* was reported on the skin of astronauts (Sugita et al., 2016). Species identification via shotgun metagenomics presents an opportunity to employ these fungi as model organisms to develop technologies aiming at microbial reduction and eradication especially because ecological analysis showed that some of the reported fungi might be opportunistic human pathogens. The dominance of *C. jadinii* on the oligotrophic cleanroom surfaces requires further investigation. This biotechnologically interesting yeast was previously tested for the production of its biomass/protein on lignocellulosic sugars and nitrogen-compounds as an alternative to plant or animal fed (Lapeña et al., 2020). Considering that discovery, *C. jadinii* could potentially nourish growth of other microorganisms found in closed habitats. In addition, understanding the mechanisms of resistance to very high doses of UV-C (Valero et al., 2007) and identifying genes and secondary metabolites involved in the survival of fungi following exposure to simulated Mars conditions and UV-C (Blachowicz et al., 2019a, 2020) might foster developing appropriate cleaning regimens. Lastly, the analysis of metagenomic reads in this study showed high abundance of transporter and retrotransposon proteins as well as cytochrome P450. These proteins were reported to contribute to improving the fungal fitness and fostering fungal adaptation and survival when exposed to harsh environmental niches (Sandmeyer and Clemens, 2010; Kulikova-Borovikova et al., 2018; Shin et al., 2018). Only by conducting such in depth phenotypic analyses on isolates found in SAF environments can PP investigators tailor the countermeasures to improve prevention of forward contamination.

Despite the incidence of fungi in SAF environments (La Duc et al., 2012; Weinmaier et al., 2015), no NASA studies have assessed the extent to which these microorganisms are resistant to, or grow under, space conditions and might potentially survive and withstand existing PP microbial reduction and cleaning approaches (Rummel et al., 2002). Importantly, recent studies have shown that representatives of *Cladosporium* and *Aspergillus* survive exposure to simulated Mars conditions and space radiation (Blachowicz et al., 2019a; Cortesão et al., 2020). Hence, the results generated in this study might aid in controlling the potential for forward contamination of fungi through robotic and crewed missions. The results of this study also directly respond to the National Research Council (NRC) recommendations for the Prevention of the Forward Contamination of Mars, that suggest examining the survival of microorganisms under adverse space conditions to assess the microbial potential to survive on spacecraft *en route* to other planetary bodies (National Research Council, 2006). Furthermore, fungal persistence on spacecraft associated surfaces pose a health threat to astronauts, as some fungal strains are pathogenic. One example is, *Aspergillus fumigatus*, a species ubiquitous in the environment, but shown to be more virulent when exposed to space conditions on the ISS than known clinical isolates (Knox et al., 2016). More importantly, *A. fumigatus* conidia have survived exposure to simulated Mars conditions for 30 min, underscoring its potential to survive in environmental extremes, including space (Blachowicz et al., 2019a). Thus, fungi isolated in this study that survived exposure to UV-C 1,000 J/m^2

are of paramount importance for conducting further studies to assess their capability to resist and survive exposure to various space conditions, including X-rays or cosmic radiation. Such understanding would promote both development of appropriate countermeasures and containing fungal forward contamination during long-duration interplanetary exploration.

While endospore-forming bacteria and halophilic archaea are known to survive and even proliferate under extreme desiccation or hypersaline conditions, it is fungi that are capable of thriving and proliferating in low water activity (a_w) environments. The lowest a_w limit where fungal growth was reported is 0.62 (Beuchat and Scouten, 2002); hence, fungal spores and conidia may survive and germinate at somewhat lower water activity (Wheeler and Hocking, 1988). Investigators have also pointed out that manipulating the accumulated intracellular polyol content can extend the range of water availability over which fungal propagules can germinate (Hallsworth and Magan, 1995). Fungi have been traditionally thought of as mostly mesophilic, heterotrophic organisms; however, eukaryotes isolated from polar ice samples are mostly fungal spores and conidia (Vorobyova et al., 2001). More than that, several fungi exhibiting radiotropism (growth toward the radiation source) were isolated from Chernobyl Nuclear Power Plant accident sites (Zhdanova et al., 2004; Blachowicz et al., 2019a). In light of these reports and considering the current interest in astrobiology and PP implications, systematical investigation of fungal presence in NASA associated cleanrooms is of high value.

To summarize, presence of fungi in habitats, even when in low abundance in comparison to bacterial counts, should not be neglected. Fungi introduced to cleanrooms and other built environments with human presence might potentially pose a threat to the closed habitats (biocorrosion) as well as their immunocompromised occupants (pathogenicity), and more critically, to the pristine environments of other celestial bodies during interplanetary explorations. This study was a first step to decipher cultivable, viable and total fungal populations in spacecraft assembly facilities and to evaluate the "real" potential of fungi encountered as biological contaminants. Furthermore, the outcomes of such studies could be translated and implemented in other cleanrooms requiring reduced counts of problematic resilient fungal populations, like intensive care units, operating rooms, or cleanrooms in the semiconducting and pharmaceutical industries.

DATA AVAILABILITY STATEMENT

Amplicon and metagenome data are available under BioProject PRJNA641079, while ITS reads of 593 cultivable isolates are available in GeneBank under accession numbers MT704867–MT704931.

AUTHOR CONTRIBUTIONS

KV designed the study, interpreted the data, and co-drafted the manuscript. AB wrote the manuscript, generated figures, and contributed to data analysis and interpretation. SM collected, processed the samples, and generated all the presented results. NS processed and analyzed the metagenomic data. JW processed, analyzed the amplicon sequencing data, and performed statistical analyses. CP performed cultivable species identification. CL collected the samples, drafted corresponding parts of the manuscript, and provided sample location figure. DB and CM performed shotgun metagenomic sequencing. All authors read and approved the final manuscript.

FUNDING

This research was supported by the NASA Planetary Protection Office 15-PPR15-0001 awarded to KV which also supported the postdoctoral fellowship for AB, SM, and CP.

ACKNOWLEDGMENTS

Part of the research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. Researchers associated with Biotechnology and Planetary Protection Group at JPL are acknowledged for their facility support. 2022 California Institute of Technology. Government sponsorship acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.777133/full#supplementary-material

REFERENCES

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

Adams, R. I., Miletto, M., Taylor, J. W., and Bruns, T. D. (2013). Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *ISME J.* 7:1460. doi: 10.1038/ismej.2013.84

Alekhova, T. A., Aleksandrova, A. A., Novozhilova, T. I., Lysak, L. V., Zagustina, N. A., and Bezborodov, A. M. (2005). Monitoring of microbial degraders in manned space stations. *Prikl. Biokhim. Mikrobiol.* 41, 435–443.

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

Bashir, M., Ahmed, M., Weinmaier, T., Ciobanu, D., Ivanova, N., Pieber, T. R., et al. (2016). Functional metagenomics of spacecraft assembly cleanrooms: presence of virulence factors associated with human pathogens. *Front. Microbiol.* 7:1321. doi: 10.3389/fmicb.2016.01321

Benardini, J. N., La Duc, M. T., Beaudet, R. A., and Koukol, R. (2014). Implementing planetary protection measures on the mars science laboratory. *Astrobiology* 14, 27–32. doi: 10.1089/ast.2013.0989

Beuchat, L. R., and Scouten, A. J. (2002). Combined effects of water activity, temperature and chemical treatments on the survival of *Salmonella* and *Escherichia coli* O157:H7 on alfalfa seeds. *J. Appl. Microbiol.* 92, 382–395. doi: 10.1046/j.1365-2672.2002.01532.x

Blachowicz, A., Chiang, A. J., Elsaesser, A., Kalkum, M., Ehrenfreund, P., Stajich, J. E., et al. (2019a). Proteomic and metabolomic characteristics of extremophilic fungi under simulated mars conditions. *Front. Microbiol.* 10:1013. doi: 10.3389/fmicb.2019.01013

Blachowicz, A., Chiang, A. J., Romsdahl, J., Kalkum, M., Wang, C. C. C., and Venkateswaran, K. (2019b). Proteomic characterization of *Aspergillus fumigatus* isolated from air and surfaces of the International Space Station. *Fungal Genet. Biol.* 124, 39–46. doi: 10.1016/j.fgb.2019.01.001

Blachowicz, A., Mayer, T., Bashir, M., Pieber, T. R., De León, P., and Venkateswaran, K. (2017). Human presence impacts fungal diversity of inflated lunar/Mars analog habitat. *Microbiome* 5:62. doi: 10.1186/s40168-017-0280-8

Blachowicz, A., Raffa, N., Bok, J. W., Choera, T., Knox, B., Lim, F. Y., et al. (2020). Contributions of spore secondary metabolites to UV-C protection and virulence vary in different aspergillus fumigatus strains. *mBio* 11:e03415–19. doi: 10.1128/mBio.03415-19

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Braak, C. J. F. T. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67, 1167–1179. doi: 10.2307/1938672

Brockett, R. M., Ferguson, J. K., and Henney, M. R. (1978). Prevalence of fungi during Skylab missions. *Appl. Environ. Microbiol.* 36, 243–246. doi: 10.1128/aem.36.2.243-246.1978

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Checinska Sielaff, A., Urbaniak, C., Mohan, G. B. M., Stepanov, V. G., Tran, Q., Wood, J. M., et al. (2019). Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. *Microbiome* 7:50. doi: 10.1186/s40168-019-0666-x

Checinska, A., Probst, A. J., Vaishampayan, P., White, J. R., Kumar, D., Stepanov, V. G., et al. (2015). Microbiomes of the dust particles collected from the international space station and spacecraft assembly facilities. *Microbiome* 3:50. doi: 10.1186/s40168-015-0116-3

Cooke, W. B. (1954). The use of antibiotics in media for the isolation of fungi from polluted water. *Antibiot. Chemother. Northfield III* 4, 657–662.

Cortesão, M., de Haas, A., Unterbusch, R., Fujimori, A., Schütze, T., Meyer, V., et al. (2020). *Aspergillus niger* spores are highly resistant to space radiation. *Front. Microbiol.* 11:560. doi: 10.3389/fmicb.2020.00560

Dolan, C. T. (1971). Optimal combination and concentration of antibiotics in media for isolation of pathogenic fungi and nocardia asteroides. *Appl. Microbiol.* 21, 195–197. doi: 10.1128/am.21.2.195-197.1971

Gorbushina, A. (2003). Microcolonial fungi: survival potential of terrestrial vegetative structures. *Astrobiology* 3, 543–554. doi: 10.1089/153110703322610636

Hallsworth, J. E., and Magan, N. (1995). Manipulation of intracellular glycerol and erythritol enhances germination of conidia at low water availability. *Microbiol. Read. Engl.* 141(Pt 5), 1109–1115. doi: 10.1099/13500872-141-5-1109

Heisel, T., Nyaribo, L., Sadowsky, M. J., and Gale, C. A. (2019). Breastmilk and NICU surfaces are potential sources of fungi for infant mycobiomes. *Fungal Genet. Biol.* 128, 29–35. doi: 10.1016/j.fgb.2019.03.008

Hendrickson, R., Lundgren, P., Mohan, G. B. M., Urbaniak, C., Benardini, J., and Venkateswaran, K. (2017). *Comprehensive Measurement of Microbial Burden in Nutrient-Deprived Cleanrooms*. Available online at: https://ttu-ir.tdl.org/handle/2346/72984 [Accessed September 9, 2019].

Herring, C. M., Brandsberg, J. W., Oxborrow, G. S., and Puleo, J. R. (1974). Comparison of media for detection of fungi on spacecraft. *Appl. Microbiol.* 27, 566–569. doi: 10.1128/am.27.3.566-569.1974

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2018). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition – interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12:e1004957. doi: 10.1371/journal.pcbi.1004957

Keller, N. P. (2019). Fungal secondary metabolism: regulation, function and drug discovery. *Nat. Rev. Microbiol.* 17, 167–180. doi: 10.1038/s41579-018-0121-1

Knox, B. P., Blachowicz, A., Palmer, J. M., Romsdahl, J., Huttenlocher, A., Wang, C. C. C., et al. (2016). Characterization of *Aspergillus fumigatus* isolates from air and surfaces of the international space station. *mSphere* 1:e00227–16. doi: 10.1128/mSphere.00227-16

Koburger, J. A., and Rodgers, M. F. (1978). Single or multiple antibiotic-amended media to enumerate yeasts and molds. *J. Food Prot.* 41, 367–369. doi: 10.4315/0362-028X-41.5.367

Kulikova-Borovikova, D., Lisi, S., Dauss, E., Alamae, T., Buzzini, P., Hallsworth, J. E., et al. (2018). Activity of the α-glucoside transporter Agt1 in *Saccharomyces cerevisiae* cells during dehydration-rehydration events. *Fungal Biol.* 122, 613–620. doi: 10.1016/j.funbio.2018.03.006

Kwan, K., Cooper, M., La Duc, M. T., Vaishampayan, P., Stam, C., Benardini, J. N., et al. (2011). Evaluation of procedures for the collection, processing, and analysis of biomolecules from low-biomass surfaces. *Appl. Environ. Microbiol.* 77, 2943–2953. doi: 10.1128/AEM.02978-10

La Duc, M. T., Kern, R., and Venkateswaran, K. (2004). Microbial monitoring of spacecraft and associated environments. *Microb. Ecol.* 47, 150–158. doi: 10.1007/s00248-003-1012-0

La Duc, M. T., Nicholson, W., Kern, R., and Venkateswaran, K. (2003). Microbial characterization of the Mars Odyssey spacecraft and its encapsulation facility. *Environ. Microbiol.* 5, 977–985. doi: 10.1046/j.1462-2920.2003.00496.x

La Duc, M. T., Vaishampayan, P., Nilsson, H. R., Torok, T., and Venkateswaran, K. (2012). Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars. *Appl. Environ. Microbiol.* 78, 5912–5922. doi: 10.1128/AEM.01435-12

Lai, X., Cao, L., Tan, H., Fang, S., Huang, Y., and Zhou, S. (2007). Fungal communities from methane hydrate-bearing deep-sea marine sediments in South China Sea. *ISME J.* 1, 756–762. doi: 10.1038/ismej.2007.51

Lapeña, D., Olsen, P. M., Arntzen, M. Ø, Kosa, G., Passoth, V., Eijsink, V. G. H., et al. (2020). Spruce sugars and poultry hydrolysate as growth medium in repeated fed-batch fermentation processes for production of yeast biomass. *Bioprocess Biosyst. Eng.* 43, 723–736. doi: 10.1007/s00449-019-02271-x

Legendre, P., and Legendre, L. (2020). *Numerical Ecology, Volume 24 - 3rd Edn. Amsterdam: Elsevier*. Available online at: https://www.elsevier.com/books/numerical-ecology/legendre/978-0-444-53868-0 [Accessed May 26, 2020].

Mayer, T., Blachowicz, A., Probst, A. J., Vaishampayan, P., Checinska, A., Swarmer, T., et al. (2016). Microbial succession in an inflated lunar/Mars analog habitat during a 30-day human occupation. *Microbiome* 4:22. doi: 10.1186/s40168-016-0167-0

Meadow, J. F., Altrichter, A. E., Bateman, A. C., Stenson, J., Brown, G. Z., Green, J. L., et al. (2015). Humans differ in their personal microbial cloud. *PeerJ* 3:e1258. doi: 10.7717/peerj.1258

Moissl, C., Osman, S., La Duc, M. T., Dekas, A., Brodie, E., DeSantis, T., et al. (2007). Molecular bacterial community analysis of clean rooms where spacecraft are assembled. *FEMS Microbiol. Ecol.* 61, 509–521. doi: 10.1111/j.1574-6941.2007.00360.x

Mora, M., Perras, A., Alekhova, T. A., Wink, L., Krause, R., Aleksandrova, A., et al. (2016). Resilient microorganisms in dust samples of the International Space Station—survival of the adaptation specialists. *Microbiome* 4:65. doi: 10.1186/s40168-016-0217-7

National Research Council (2006). *Preventing the Forward Contamination of Mars*. Washington, DC: National Academies Press, doi: 10.17226/11381

Nayfach, S., and Pollard, K. S. (2016). Toward accurate and quantitative comparative metagenomics. *Cell* 166, 1103–1116. doi: 10.1016/j.cell.2016.08.007

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 20, 241–248. doi: 10.1016/j.funeco.2015.06.006

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., and Tedersoo, L. (2019a). Mycobiome diversity: high-throughput sequencing and

identification of fungi. *Nat. Rev. Microbiol.* 17, 95–109. doi: 10.1038/s41579-018-0116-y

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019b). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47, D259–D264. doi: 10.1093/nar/gky1022

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013). *Vegan: Community Ecology Package*.

Oong, G. C., and Tadi, P. (2021). *Chloramphenicol. StatPearls Publishing*. Available online at: https://www.ncbi.nlm.nih.gov/books/NBK555966/ [Accessed February 2, 2022].

Oong, G. C., and Tadi, P. (2022). *"Chloramphenicol," in StatPearls (Treasure Island (FL): StatPearls Publishing)*. Available online at: http://www.ncbi.nlm.nih.gov/books/NBK555966/ [Accessed February 2, 2022].

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., et al. (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40, D284–D289. doi: 10.1093/nar/gkr1060

Romsdahl, J., Blachowicz, A., Chiang, A. J., Singh, N., Stajich, J. E., Kalkum, M., et al. (2018a). Characterization of *Aspergillus niger* Isolated from the International Space Station. *mSystems* 3:e00112–18.

Romsdahl, J., Blachowicz, A., Chiang, A. J., Singh, N., Stajich, J. E., Kalkum, M., et al. (2018b). Characterization of *Aspergillus niger* isolated from the international space station. *mSystems* 3:e00112–18. doi: 10.1128/mSystems.00112-18

Romsdahl, J., Blachowicz, A., Chiang, Y.-M., Venkateswaran, K., and Wang, C. C. C. (2020). Metabolomic analysis of *Aspergillus niger* isolated from the international space station reveals enhanced production levels of the antioxidant pyranonigrin A. *Front. Microbiol.* 11:931.

Rummel, J. D., Stabekis, P. D., Devincenzi, D. L., and Barengoltz, J. B. (2002). COSPAR's planetary protection policy: a consolidated draft. *Adv. Space Res.* 30, 1567–1571. doi: 10.1016/S0273-1177(02)00479-9

Sandmeyer, S. B., and Clemens, K. A. (2010). Function of a retrotransposon nucleocapsid protein. *RNA Biol.* 7, 642–654. doi: 10.4161/rna.7.6.14117

Satoh, K., Yamazaki, T., Nakayama, T., Umeda, Y., Alshahni, M. M., Makimura, M., et al. (2016). Characterization of fungi isolated from the equipment used in the International Space Station or Space Shuttle. *Microbiol. Immunol.* 60, 295–302. doi: 10.1111/1348-0421.12375

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 37, D5–D15. doi: 10.1093/nar/gkn741

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* 14:e1002533. doi: 10.1371/journal.pbio.1002533

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shin, J., Kim, J.-E., Lee, Y.-W., and Son, H. (2018). Fungal cytochrome P450s and the P450 complement (CYPome) of *Fusarium graminearum*. *Toxins* 10:112. doi: 10.3390/toxins10030112

Singh, N. K., Wood, J. M., Karouia, F., and Venkateswaran, K. (2018). Succession and persistence of microbial communities and antimicrobial resistance genes associated with International Space Station environmental surfaces. *Microbiome* 6:3990. doi: 10.1186/s40168-018-0585-2

Sterflinger, K., Tesei, D., and Zakharova, K. (2012). Fungi in hot and cold deserts with particular reference to microcolonial fungi. *Fungal Ecol.* 5, 453–462. doi: 10.1016/j.funeco.2011.12.007

Sugita, T., Yamazaki, T., Makimura, K., Cho, O., Yamada, S., Ohshima, H., et al. (2016). Comprehensive analysis of the skin fungal microbiota of astronauts during a half-year stay at the International Space Station. *Med. Mycol.* 54, 232–239. doi: 10.1093/mmy/myv121

Taylor, D. L., and Bruns, T. D. (1999). Community structure of ectomycorrhizal fungi in a Pinus muricata forest: minimal overlap between the mature forest and resistant propagule communities. *Mol. Ecol.* 8, 1837–1850. doi: 10.1046/j.1365-294x.1999.00773.x

Tournas, V., Stack, M. E., Mislivec, P. B., Koch, H. A., and Bandler, R. (2020). *BAM Chapter 18: Yeasts, Molds and Mycotoxins. FDA*. Available online at: https://www.fda.gov/food/laboratory-methods-food/bam-chapter-18-yeasts-molds-and-mycotoxins [Accessed December 10, 2020].

Urbaniak, C., Sielaff, A. C., Frey, K. G., Allen, J. E., Singh, N., Jaing, C., et al. (2018). Detection of antimicrobial resistance genes associated with the international space station environmental surfaces. *Sci. Rep.* 8:814. doi: 10.1038/s41598-017-18506-4

Urbaniak, C., Van Dam, P., Zaborin, A., Zaborina, O., Gilbert, J. A., Torok, T., et al. (2019). Genomic characterization and virulence potential of two *Fusarium oxysporum* isolates cultured from the international space station. *mSystems* 4:e00345–18. doi: 10.1128/mSystems.00345-18

Valero, A., Begum, M., Leong, S. L., Hocking, A. D., Ramos, A. J., Sanchis, V., et al. (2007). Effect of germicidal UVC light on fungi isolated from grapes and raisins. *Lett. Appl. Microbiol.* 45, 238–243. doi: 10.1111/j.1472-765X.2007.02175.x

Venkateswaran, K., La Duc, M. T., and Vaishampayan, P. (2012). *Genetic Inventory Task: Final Report*. Pasadena, CA: Jet Propulsion Laboratory, California Institute of Technology.

Vorobyova, E., Minkovsky, N., Mamukelashvili, A., Zvyagintsev, D., Soina, V., Polanskaya, L., et al. (2001). "Micro-organisms and biomarkers in permafrost," in *Permafrost Response on Economic Development, Environmental Security and Natural Resources NATO Science Series*, eds R. Paepe, V. P. Melnikov, E. Van Overloop, and V. D. Gorokhov (Dordrecht: Springer Netherlands), 527–541. doi: 10.1007/978-94-010-0684-2_36

Weinmaier, T., Probst, A. J., La Duc, M. T., Ciobanu, D., Cheng, J.-F., Ivanova, N., et al. (2015). A viability-linked metagenomic analysis of cleanroom environments: eukarya, prokaryotes, and viruses. *Microbiome* 3:62. doi: 10.1186/s40168-015-0129-y

Wheeler, K. A., and Hocking, A. D. (1988). Water relations of *Paecilomyces variotii*, *Eurotium amstelodami*, *Aspergillus candidus* and *Aspergillus sydowii*, *Xerophilic fungi* isolated from Indonesian dried fish. *Int. J. Food Microbiol.* 7, 73–78. doi: 10.1016/0168-1605(88)90074-8

Zhdanova, N. N., Tugay, T., Dighton, J., Zheltonozhsky, V., and Mcdermott, P. (2004). Ionizing radiation attracts soil fungi. *Mycol. Res.* 108, 1089–1096. doi: 10.1017/S0953756204000966

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Blachowicz, Mhatre, Singh, Wood, Parker, Ly, Butler, Mason and Venkateswaran. 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 | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 May 2022
doi: 10.3389/fmicb.2022.738205

Establishment of Genome Based Criteria for Classification of the Family Desulfovibrionaceae and Proposal of Two Novel Genera, *Alkalidesulfovibrio* gen. nov. and *Salidesulfovibrio* gen. nov.

*Mi-Jeong Park[1,2], Yun Jae Kim[1], Myeongkyu Park[3], Jihyun Yu[1,2], Teddy Namirimu[1,2], Yoo-Rim Roh[1,2] and Kae Kyoung Kwon[1,2]**

[1] Marine Biotechnology Research Center, Korea Institute of Ocean Science & Technology, Busan, South Korea, [2] Department of Applied Ocean Science, University of Science and Technology, Daejeon, South Korea, [3] Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea

OPEN ACCESS

Edited by:
Rafael R. de la Haba, University of Sevilla, Spain

Reviewed by:
Jan Kuever, Bremen Institute for Materials Testing, Germany
Alexander Galushko, Agrophysical Research Institute, Russia

***Correspondence:**
Kae Kyoung Kwon
kkkwon@kiost.ac.kr

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

Received: *08 July 2021*
Accepted: *11 April 2022*
Published: *25 May 2022*

Citation:
Park M-J, Kim YJ, Park M, Yu J, Namirimu T, Roh Y-R and Kwon KK (2022) Establishment of Genome Based Criteria for Classification of the Family Desulfovibrionaceae and Proposal of Two Novel Genera, Alkalidesulfovibrio gen. nov. and Salidesulfovibrio gen. nov. Front. Microbiol. 13:738205. doi: 10.3389/fmicb.2022.738205

Bacteria in the Desulfovibrionaceae family, which contribute to S element turnover as sulfate-reducing bacteria (SRB) and disproportionation of partially oxidized sulfoxy anions, have been extensively investigated since the importance of the sulfur cycle emerged. Novel species belonging to this taxon are frequently reported, because they exist in various environments and are easy to culture using established methods. Due to the rapid expansion of the taxon, correction and reclassification have been conducted. The development of high-throughput sequencing facilitated rapid expansion of genome sequence database. Genome-based criteria, based on these databases, proved to be potential classification standard by overcoming the limitations of 16S rRNA-based phylogeny. Although standards methods for taxogenomics are being established, the addition of a novel genus requires extensive calculations with taxa, including many species, such as Desulfovibrionaceae. Thus, the genome-based criteria for classification of Desulfovibrionaceae were established and validated in this study. The average amino-acid identity (AAI) cut-off value, 63.43 ± 0.01, was calculated to be an appropriate criterion for genus delineation of the family Desulfovibrionaceae. By applying the AAI cut-off value, 88 genomes of the Desulfovibrionaceae were divided into 27 genera, which follows the core gene phylogeny results. In this process, two novel genera (*Alkalidesulfovibrio* and *Salidesulfovibrio*) and one former invalid genus ("*Psychrodesulfovibrio*") were officially proposed. Further, by applying the 95–96% average nucleotide identity (ANI) standard and the 70% digital DNA–DNA hybridization standard values for species delineation of strains that were classified as the same species, five strains have the potential to be newly classified. After verifying that the

classification was appropriately performed through relative synonymous codon usage analysis, common characteristics were listed by group. In addition, by detecting metal resistance related genes *via in silico* analysis, it was confirmed that most strains display metal tolerance.

Keywords: Desulfovibrionaceae, AAI, RSCU, genome, *Desulfovibrio*, taxogenomics, classification criteria

INTRODUCTION

Sulfur is an essential element of biomolecules and an important factor in climate change through direct and indirect effects in the H_2SO_4 form, which brought attention to the sulfur cycle (Kellogg et al., 1972). The Desulfovibrionaceae family has been identified as one of the major contributor to the sulfur cycle on Earth. Discoveries of novel species are frequently reported as these species exist in various environments and can be easily cultured using well-established methods (Postgate, 1984; Widdel and Bak, 1992). Members of the family Desulfovibrionaceae have been employed as model organisms for sulfate reducing bacteria (SRB). Studies regarding anaerobic respiration have been actively conducted using *Desulfovibrio* since the 1950s, when dissimilatory sulfate reduction and a sulfate reductase named desulfoviridin (Postgate, 1956) were first identified in *Desulfovibrio* (Peck, 1959, 1961; Vosjan, 1975; Kobayashi et al., 1982; Barton et al., 1983; Lie et al., 1996; Matias et al., 2005; Pereira, 2008; Keller and Wall, 2011; Keller et al., 2014). Recently, interesting research results have been published on the effect of Desulfovibrionaceae family bacteria on host health. Although the exact correlation and mechanism have not been established, the results showed that the relative abundance of Desulfovibrionaceae was significantly increased in obese and metabolically impaired mice (Just et al., 2018), and the amount of *Desulfovibrio* spp. detected in the feces of Parkinson's disease, patients showed a significant correlation with the severity of the disease (Murros et al., 2021). However, *Desulfovibrio* spp. do not always play negative roles. For example, *Desulfovibrio* spp. may live in sulfate depleted habitats in syntropy with methanogenic archaea (Scholten et al., 2007). SRB converts sulfate into sulfide, which reacts with heavy metals to form metal sulfide. Subsequently, spontaneous precipitation of metal sulfide achieves bioremediation (Ayangbenro et al., 2018). Aside from precipitation of sulfide, *Pseudodesulfovibrio hydrargyri* and *Pseudodesulfovibrio mercurii* are well-known mercury methylating bacteria, and activity against 10 ppb and 1 ppm inorganic Hg has been confirmed (Goñi-Urriza et al., 2020). *Desulfovibrio desulfuricans* showed resistance and high removal rates at concentrations of 50, 100, and 200 ppm in a mixed solution of cadmium nitrate tetrahydrate, nickel sulfate hexahydrate, and chromium oxide (Jeong et al., 2015). *Desulfovibrio magneticus* exhibited a sharp decrease in cadmium before the exponential phase when 1.3 ppm cadmium chloride was added to the medium (Arakaki et al., 2002). *Nitratidesulfovibrio vulgaris* was reported to tolerate a wide range of metal ions, such as 10 ppm Mn(II), 15 ppm Cr(III), 4 ppm Cu(II), 8.5 ppm Ni(II), and 20 ppm Zn(II) (Goulhen et al., 2006).

After the genus *Desulfovibrio* was first proposed in 1936 (Kluyver and Van Niel, 1936), the first attempt to classify this taxon, based on of DNA composition and physiological and biochemical properties, was crucial in establishing the classification criteria (Postgate and Campbell, 1966). In 2002, the genus *Desulfomonas* (Moore et al., 1976) was reclassified as a member of the genus *Desulfovibrio* based on molecular analysis (Loubinoux et al., 2002). Subsequently, the Desulfovibrionaceae family was officially recognized in 2006 (Kuever et al., 2005). The genera *Bilophila* (Baron et al., 1989), *Desulfovibrio* (Kluyver and Van Niel, 1936), and *Lawsonia* (McOrist et al., 1995) were reclassified into this family. Thereafter, *Desulfocurvus* (Klouche et al., 2009) and *Desulfobaculum* (Zhao et al., 2012) were added. The addition of several new genera, namely *Pseudodesulfovibrio* (Cao et al., 2016), *Halodesulfovibrio* (Shivani et al., 2017), "*Mailhella*" (Ndongo et al., 2017), *Desulfohalovibrio*, *Desulfocurvibacter* (Spring et al., 2019), "*Paradesulfovibrio*" (Kim et al., 2020), and *Desulfolutivibrio* (Thiel et al., 2020), finally led to a major reorganization based on the genome sequence data (Waite et al., 2020), resulting in Desulfovibrionaceae currently comprising nine validly published genera (*Fundidesulfovibrio, Humidesulfovibrio, Maridesulfovibrio, Megalodesulfovibrio, Nitratidesulfovibrio, Oleidesulfovibrio, Paradesulfovibrio, Paucidesulfovibrio,* and *Solidesulfovibrio*) and three published but not validated genera ("*Alteridesulfovibrio,*" "*Aminidesulfovibrio,*" and "*Frigididesulfovibrio*") besides the existing genera. Reorganizing the confusing taxonomy of the Desulfovibrionaceae family was completed by adding two valid [*Macrodesulfovibrio* (Galushko and Kuever, 2020b), *Oceanidesulfovibrio* (Galushko and Kuever, 2020c)] and one invalid genera ["*Psychrodesulfovibrio*" (Galushko and Kuever, 2020d)] at the end of 2020 through Bergey's manual (Galushko and Kuever, 2020a). As of August 2021, the Desulfovibrionaceae family comprised 21 validated genera, one synonym, and five non-validated genera.

The recently introduced classification method using the whole genome has been a powerful alternative to the two-step approach, which combined 16S rRNA gene sequence similarity and DNA–DNA hybridization (DDH) (Chun and Rainey, 2014). A genome sequence-based classification, such as average nucleotide identity (ANI) and digital DNA–DNA hybridization (*d*DDH), was introduced (Richter and Rosselló-Móra, 2009) to improve the resolution between highly similar interspecies sequences caused by the short length of the 16S rRNA sequence (Wambui et al., 2021). However, the resolution was insufficient for classification at the genus level, as it comprised only four nucleotide types. Therefore, as an alternative, the average amino acids identity (AAI)-based classification method was introduced to compare the genome

information composed of amino acids (Konstantinidis and Tiedje, 2005b; Chun and Rainey, 2014; Rodriguez-R and Konstantinidis, 2014; Barco et al., 2020). There are several successful cases of reclassification using the whole genome, for example, dividing 39 strains belonging to 27 species of *Arcobacter* spp. into seven genera (Pérez-Cataluña et al., 2018) and reorganizing 91 genomes belonging to the existing three genera of the order *Methylococcales* into four genera (Orata et al., 2018). These cases followed a similar methodology, wherein the housekeeping genes were extracted, and phylogeny was performed using the ANI and *d*DDH values for the delineation of species, and percentage of conserved proteins (POCP) and AAI values for the delineation of genera. As the POCP is calculated using an amino acid sequence, it allows a higher resolution comparison for distant groups than the ANI or *d*DDH values that use a nucleotide sequence. Further, genome sequence-based reclassification of *Epsilonproteobacteria* and *Deltaproteobacteria*, to which SRB are affiliated, had been conducted (Waite et al., 2017, 2020). Although this method is not based on comparing genomic indices, it became the foundation for SRB classification by applying an alternative taxogenomic method based on phylogeny to a vast range of taxa and reorganizing them into a new phylum.

Although the results published in Waite et al. (2020) and Bergey's manual (Galushko and Kuever, 2020a) almost corrected the confusing taxonomy of the family Desulfovibrionaceae, still issues require further investigation. First, the research did not provide a numerical value for genus delineation. Therefore, researchers proposing a new genus must perform extensive calculations and investigation. Comparing genomic indices becomes increasingly complex when studying large genera. Second, the classification was based solely on phylogeny results, indicating that the genus classification that introduced the numerical standard value comparison of genomic indices, which is considered a formal method of taxonomy, has not been made. Third, several mis-classified groups remain, particularly *Cupidesulfovibrio* [a genus newly proposed in early 2021 (Wan et al., 2021)] which requires further reclassification because it collides with the genus *Nitratidesulfovibrio*. Referring to the taxonomic comments on the Desulfovibrionaceae family in the 2020 version of Bergey's manual (Galushko and Kuever, 2020a), some suggested that several species of *Pseudodesulfovibrio*, which are related to '*Paradesulfovibrio onnuriensis*' IOR2^T, deserve to be considered as an independent genus.

In this study, we performed phylogeny investigations using core genes extracted from all genomes belonging to Desulfovibrionaceae and roughly divided them into smaller groups. To evaluate the categorization process, several genomic indices, such as G + C content, *d*DDH, ANI, AAI, and POCP, were explored. Moreover, we not only arranged mis-classified strains at the species level but also proposed new genera reflecting phenotypes of other well-divided groups. This research will provide a genus cut-off genomic index that can be referenced without comparing extensive genome indices with previously reported species. In addition, the metal tolerance enzyme prediction using genome data highlights that the family Desulfovibrionaceae displays tolerance to certain extreme environments.

MATERIALS AND METHODS

Bacterial Strains and Sequences

Information on all strains was collected from their related literature (**Supplementary Table 1**). All genomes of isolates, except for *Pseudodesulfovibrio tunisiensis* RB22^T, registered under the Desulfovibrionaceae family were obtained from two public databases, the National Center for Biotechnology Information (NCBI) and EzBioCloud (Yoon et al., 2017a). *Pseudodesulfovibrio tunisiensis* RB22^T strains were ordered from JCM RICKEN and cultured for two days in DSM143 medium injected with H_2/CO_2 gas. Afterward, their genomic DNAs were extracted using QIAGEN DNeasy tissue kits and sequenced using the Oxford Nanopore PromethION sequencer according to NICEM's commercial process. The raw data were *de novo* assembled using Flye v2.9-b1768 (Kolmogorov et al., 2019) and polished four times using Racon v1.5.0 (Vaser et al., 2017) and one time using Medaka v1.6.0.[1] As a result, a complete genome of 1 contig of 3.6Mb size showing coverage of 675X depth was obtained (GenBank Assembly accession: GCA_022809775). To maintain the same conditions, all genomes were annotated using Prokka version 1.14.6 (Seemann, 2014). Information regarding CDS, rRNA, CRISPR repeat region, tRNA, and tmRNA was obtained from the Prokka output. Genome assembly statistics, such as number of contigs, bases (size), N50, and G + C content, were obtained using the Genome Assembly Annotation Service (GAAS) tool kit. Small subunit (SSU) rRNA sequences of the type strains were obtained from EzBioCloud. The data for strains with unpublished SSU rRNA sequences were directly exported from the annotation file for each strain.

Phylogenetic Analyses

Core gene phylogeny was performed using PhyloPhlAn (Segata et al., 2013) with .faa result files obtained from Prokka. The phylogenomic tree was constructed with default options of 400 universal protein markers (**Supplementary Table 2**) and the following tools for the internal steps: USEARCH v5.2.32 (Edgar, 2010) for mapping into amino acid databases; MUSCLE v3.6 (Edgar, 2004) for multiple sequence alignment, and FastTree v2.1.10 (Price et al., 2009) with 1,000 bootstraps for phylogenic inference. The genomes of two strains belonging to the phylum *Rhodothermaeota* were downloaded from NCBI and used as an outgroup. In addition, the results were compared by phylogeny based on the 16S rRNA sequences. The 16S rRNA-based phylogenetic tree was multiple aligned using ClustalW (Chenna et al., 2003) in MEGA 6.0 (Tamura et al., 2013), and 1,000 bootstrap iterations were conducted to construct a tree using the neighbor-joining (Saitou and Nei, 1987) and the maximum-likelihood (White, 1982) methods with the Jukes–Cantor model (Jukes and Cantor, 1969) and the maximum-parsimony (Moore et al., 1973) method.

[1] https://github.com/nanoporetech/medaka

Calculation of Genomic Indices

Average nucleotide identity values were calculated using the OAU (OrthoANI-usearch tool) (Yoon et al., 2017b), and *d*DDH (*digital*DDH, known as *is*DDH) was calculated using the Genome-to-Genome Distance Calculator (GGDC) (Meier-Kolthoff et al., 2013) provided by the DSMZ (Braunschweig, Germany). The AAI calculator by Kostas lab (Luo et al., 2014) was employed to compare two genomes written in amino acids with 20% minimum identity and 50 minimum alignments as alignment options. POCP is a comparison of the genomes as an amino acid sequence of two strains using BLASTP. The matched proteins with an E-value less than 10^{-5}, a sequence identity over 40%, and a query cover of over 50% were regarded as conserved proteins. As strains belonging to the same genus share conserved proteins, with at least half of the whole protein, the strains showing a POCP value over 50% can be considered belonging to the same genus. POCP values were calculated as $[(C_1 + C_2)/(T_1 + T_2)] \times 100\%$, where C_1 and C_1 represent the conserved number of proteins in the two genomes, respectively; and T_1 and T_1 represent the total number of proteins in the two genomes being compared, respectively (Qin et al., 2014).

Relative Synonymous Codon Usage Analysis

The open reading frame (ORF) search was performed as a priority to compare codon usage. The ORF finder function of the Sequence Manipulation Suite tool (Stothard, 2000) was used with the bacterial genetic code beginning with only ATG. The obtained ORFs were calculated using the bacterial genetic code for the codon usage function of the same tool. Analysis was conducted only for codons, excluding Met (ATG) and Trp (TGG), which are encoded by only one codon, and stop codons (TAG, TGA, and TAA) that encode no amino acids. A global test between codon bias in each group was conducted using Analysis of similarity (ANOSIM) with R's vegan package (Oksanen et al., 2013; Gu et al., 2020). Principal component analysis (PCA) was performed using R's prcomp formula, and since 93% of the variables could be explained with three principal components, 3D plotting was performed using plot_ly packages (Sievert, 2020). The distance between plots was obtained by introducing the coordinates into Equation 1.

$$d = [(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2]^{1/2} \quad (1)$$

Comparison of Phenotype and Metal Resistance-Related Genes Between Groups

A literature study was conducted on all strains of the Desulfovibrionaceae family to investigate their phenotype. DIAMOND BLASTP was performed on the BacMet v2.0 database to identify metal resistance-related genes. Genes which satisfy the following conditions were searched: E-value lower than 10^{-5}, 70% or more query coverage, and 30% similarity. The relative abundance of metal resistance-related genes was shown as a heatmap for each genome using ggplot2.

RESULTS

Bacterial Strains and Genomes

All genomes (88 genomes) belonging to Desulfovibrionaceae were collected for subsequent analyses, and their basic information is summarized in **Supplementary Table 1**. A portion of the target genomes originate from host body isolates, including blood, gut, and feces (Moore et al., 1976; Baron et al., 1989; McOrist et al., 1995) but mainly from anaerobic environments such as sludge, mud, wastewater (sulfidic water), oil field, and marine sediment (Baena et al., 1998; Hernandez-Eugenia et al., 2000). Some members of this family have also been isolated from extreme environments, such as permafrost, ocean vent fields (hydrothermal chimneys), and heavy-metal-effected lake sediments (Ramsay et al., 2015; Ryzhmanova et al., 2019; Kim et al., 2020). Likewise, their genomic characteristics also showed large variations amongst isolation sources. Genome sizes ranged from 1.41 to 5.77 Mb, and the G + C content varied over a wide range (32.9–69.8 mol%). Genome statistics, such as contig number and N50, and characteristics, such as CDS, rRNA, tRNA, tmRNA, and CRISPR repeat regions, are summarized in **Supplementary Table 1** along with the GenBank assembly accession number.

Core Genes Phylogeny

Phylogeny was performed using core genes extracted from previously collected genomes (**Figure 1**). A quick skim of a classified group with a difference of less than 1.5 distance, referred to the scale bar on the tree, suggested that 88 genomes, excluding the outgroup, could be subdivided into 27 groups. The results of clustering corresponded well with the validly published genera in most group cases. Although the genera constituting groups 5, 15, 21, and 23 are not officially recognized genera, each cluster resulting from those groups matched an individual genus. Group 1 is a mixture of the genera "*Paradesulfovibrio*" ("*Paradesulfovibrio onnuriensis*"), *Desulfovibrio* ("*Desulfovibrio brasiliensis*" and *Desulfovibrio oxyclinae*), and *Pseudodesulfovibrio* (*Pseudodesulfovibrio senegalensis*, *Pseudodesulfovibrio halophilus*). Group 8 is a mix of *Paradesulfovibrio* (*Paradesulfovibrio bizertensis*) and *Desulfobaculum* (*Desulfobaculum xiamenense*). In group 18, *Cupidesulfovibrio* and *Nitratidesulfovibrio* genera coexist. *Desulfohalovibrio alkalitolerans* belonging to group 9, has been classified as belonging to the same genus as *Desulfohalovibrio*, and with *Desulfohalovibrio reitneri* belong in group 10. However, as they show sufficient evolutionary distance, it was determined that they should be classified into separate groups. In the same vein, *Fdv*, from group 23, classified as "*Frigididesulfovibrio*" can also be proposed as a new genus. Last, group 18 is a cluster of two genera, *Nitratidesulfovibrio* (*Nitratidesulfovibrio oxamicus, Ntd. vulgaris,* and *Nitratidesulfovibrio termitidis*) and *Cupidesulfovibrio* (*Cupidesulfovibrio liaohensis*).

Genome-Based Similarity Indices

Amino-acid identity and POCP are popular examples of genome-based similarity indices for genus delineation. The

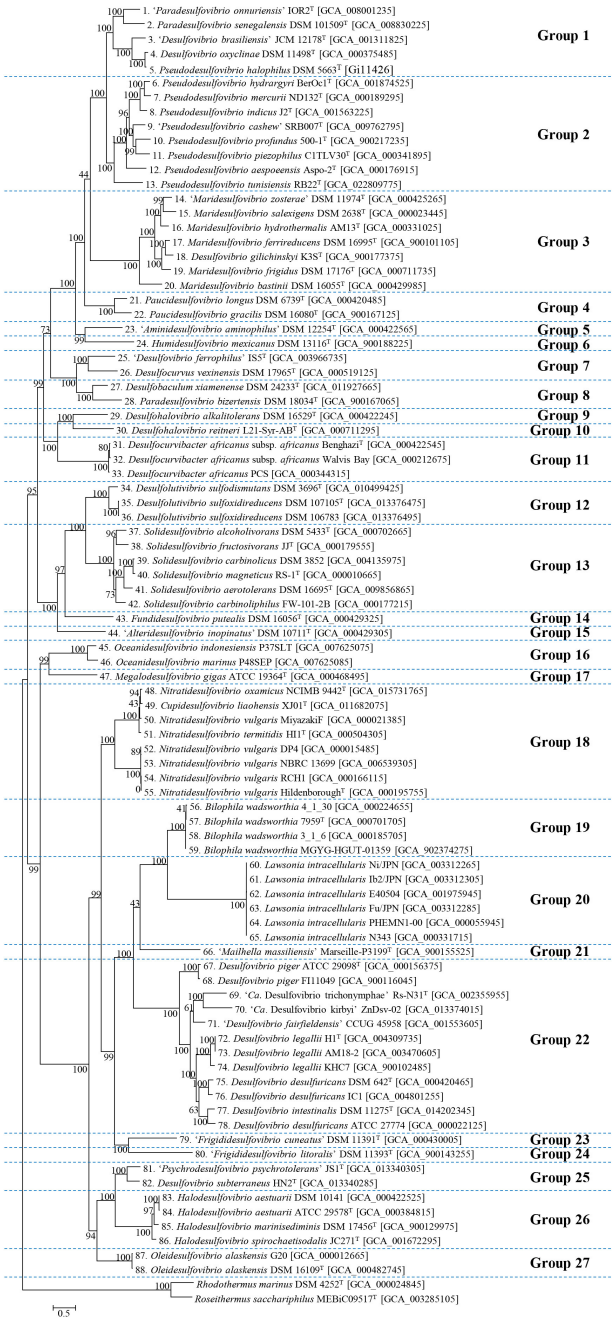


FIGURE 1 | Phylogenomic tree of strains belonging to Desulfovibrionacae. Using 400 core gene markers, 88 Desulfovibrionaceae family strains were subdivided into 27 groups. Two type strains belonging to Rhodothermaceae were selected as outgroups. Each number before the scientific name implies the identical organism with the corresponding number at the subsequent analysis. The scale bar represents 5% estimated sequence divergence.

AAI cut-off value applied for the delineation of genera has a wide range of 60–80% (Konstantinidis and Tiedje, 2005a; Rodriguez-R and Konstantinidis, 2014). In previous studies on the Desulfovibrionaceae family, the AAI value of 60% was adopted to distinguish the genus by checking whether the existing genus satisfies these values (Spring et al., 2019). However, in reality, many taxa do not meet the AAI value criteria above, so it is necessary to verify whether this is an "appropriate AAI criterion that does not violate the monophyly rule" (Orata et al., 2018). Therefore, we devised a method to determine the threshold through clustering and scoring after much consideration. Subjective intervention was avoided by repetitive clustering and evaluation with gradual increment of cut-off parameters to establish unbiased criteria, and not setting strict borders after dividing the genus first. As the AAI cut-off value applied for each taxon varied, a Python script was written to objectively evaluate various thresholds and establish the criteria. This Python script automatically iterates clustering work and evaluates its result to determine an optimum threshold between 60.0 and 80.0. In the clustering step, threshold-based clustering was performed using the given pairwise AAI matrix (**Supplementary Table 3**). In the scoring step, inclusion of a member in multiple clusters was heavily penalized, whereas the single member group received relatively small point deductions. When the scoring is complete, the threshold is raised by 0.01 and the steps are repeated until the threshold reaches 80.00 (**Supplementary Figure 1**). For example, if there are three genomes A, B, and C, and A shows similarity above the given threshold with B and C, respectively, but B and C show similarity below the threshold with each other, Three clusters are created: [A, B, C], [A, B], and [A,C]. At this time, since A belongs to these three clusters, a large deduction is given. In order to distinguish the penalty given to the singleton member and the penalty given for overlapping clusters, the large deduction was given more than the total number of genome used. An executable file for Windows is available from the following repository: https://github.com/PMKYU98/monophyly_cutoff. Subsequently, the threshold that finally obtained the maximum score was set as the optimal AAI value. Therefore, there were five AAI threshold ranges calculated as possible thresholds (63.43 $\pm$ 0.01, 76.33 $\pm$ 0.23, 76.99 $\pm$ 0.17, 78.34 $\pm$ 0.11, and 78.50 $\pm$ 0.01). The threshold 63.43 $\pm$ 0.01 was determined to be the best, and when these values were set as cut-off values, it was confirmed that each genome belonged to one of the 27 independent groups without re-occurrence in multiple groups. These 27 groups matched the 27 groups abstracted in the previous phylogenomic tree. The second-best range, 76.33 $\pm$ 0.23, formed 49 groups, but an excessive number of unary groups occurred. Therefore, we selected 63.43 $\pm$ 0.01 as a promising maximum score, which is in the range of the recommended AAI threshold to classify the genus (Rodriguez-R and Konstantinidis, 2014), but slightly higher than the value used by Spring et al. (2019).

POCP played the role of an auxiliary criterion besides the AAI value, and the value for genera delineation was set at 50% (Qin et al., 2014). However, when plotting the AAI and POCP values between genomes (**Supplementary Figure 2**), 50% POCP values did not correspond with the 63.43 $\pm$ 0.01 AAI cut-off values. Moreover, applying the 50% POCP cut-off violated the monophyly rule for taxon delineation. Previous studies have reported that the POCP cut-off cannot be an appropriate standard for genus delineation in several taxa (Aliyu et al., 2016;

Li et al., 2017; Lopes-Santos et al., 2017; Orata et al., 2018; Wirth and Whitman, 2018). Therefore, it was decided not to adapt the POCP value as a criterion for genus delineation. The AAI and POCP results between each genome are displayed in a heatmap shown in **Figure 2**, and detailed values can be found in **Supplementary Tables 3, 4**.

Average nucleotide identity and *d*DDH are widely used for species delineation because their standard values are defined. If either ANI or *d*DDH values from the comparison of two genomes are under the cut-off of 95% ANI and 70% *d*DDH (Colston et al., 2014), they are considered as different species. Among the *Nitratidesulfovibrio vulgaris* strains belonging to group 18, the Miyazaki F strain should be classified as a different species. Group 22, including *Desulfovibrio piger* FI11049, *Desulfovibrio legallii* KHC7, *Dsv. desulfuricans* IC1 and *Dsv. desulfuricans* ATCC 27774, should be separated into new species from their type strains. Excluding the strains mentioned above, every other strain met both the ANI and *d*DDH criteria, and therefore, did not require additional classification. The ANI and *d*DDH results between each genome are summarized in a heatmap and shown in **Supplementary Figure 3**. Detailed values are provided in **Supplementary Tables 5, 6**.

Comparison for Codon Bias

Codon usage can reflect evolutionary processes because it is influenced by the G + C content, replication strand skew, or gene expression (Sharp et al., 2005). Thus, we calculated the relative synonymous codon usage (RSCU) values and measured codon usage bias for each group to verify their evolutionary relationship and clustering (**Supplementary Table 7**). The RSCU values obtained for each genome were then calculated and plotted. (**Figure 3A** and **Supplementary Figure 4**). ANOSIM was used to observe the similarity in the codon usage bias of each group. The R statistic was 0.678 and the *p*-value was 0.0001 in the global test result. These figures indicate each genus group has a codon usage bias that differs significantly from other groups, suggesting that the group previously divided at the genus level was sufficiently discriminated. In addition, the PCA using the RSCU values also verified whether they were the same species. **Figure 3** shows the distance between them

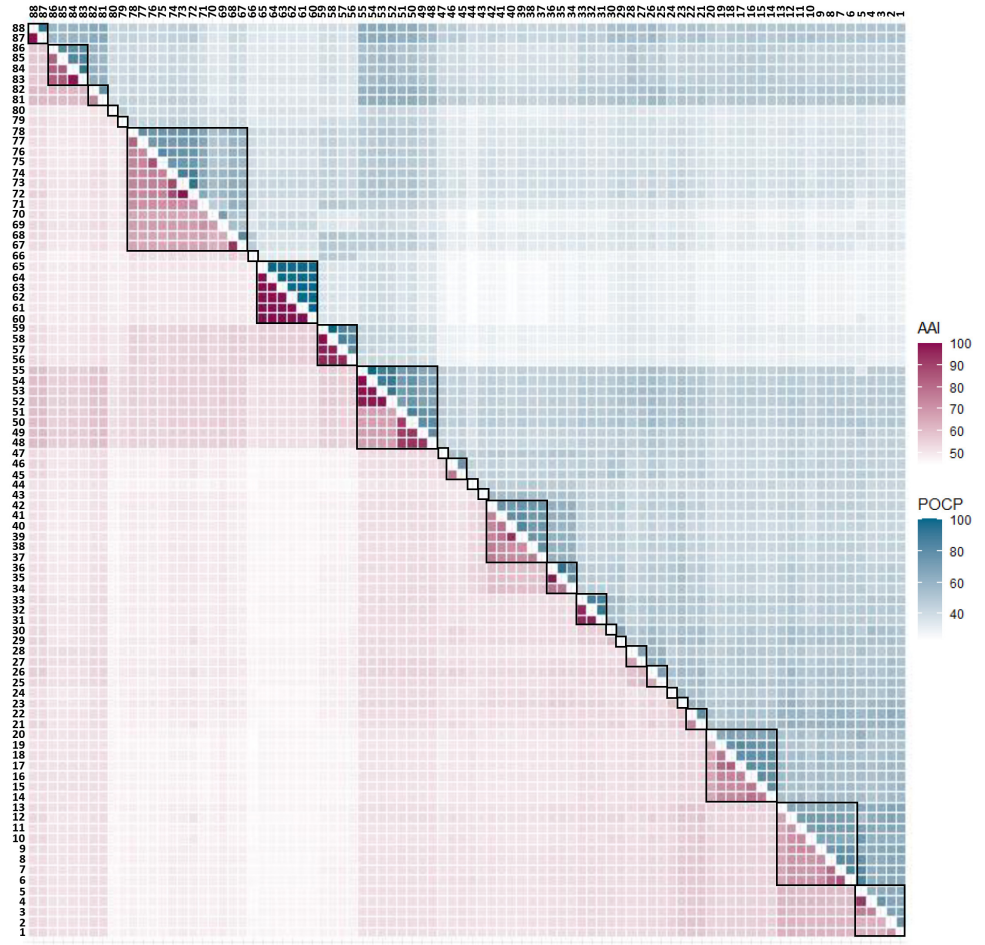
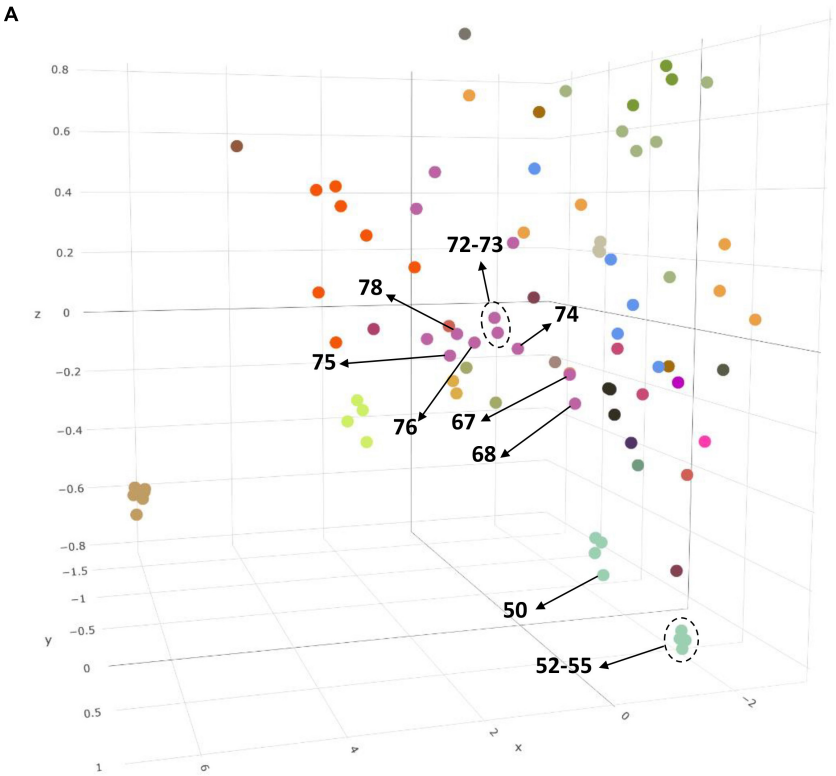


FIGURE 2 | AAI and POCP from pairwise whole-genome comparisons. The values of AAI and POCP were expressed as heatmap. AAI is the lower left triangle indicated in purple color, and POCP is the upper right triangle indicated in cyan color. When the group was divided based on the AAI value alone, it was subdivided into 27 groups, which are separately surrounded by a black square on the figure.



B

Pairs	Distances
50-52	1.59
50-53	1.64
50-54	1.55
50-55	1.58
52-53	0.06
52-54	0.07
52-55	0.03
53-54	0.10
53-55	0.07
54-55	0.06
67-68	0.19
72-73	0.09
72-74	0.45
73-74	0.53
75-76	0.53
75-78	0.14
76-78	0.42

FIGURE 3 | (A) 3D plot of the three major axes generated by principal component analysis (PCA) of the RSCU values. The RSCU of each strain except for the 5 codons of Met, Trp, and Stop codons were plotted after PCA analysis. In this scatter plot, the same color indicates the same genus and the label of each dot can be found in the accompanying.html file. **(B)** The distance on the PCA plot between the existing group and the strain to be reported as a novel species. 50, *Ntd. vulgaris* Miyazaki F; 52, *Ntd. vulgaris* DP4; 53, *Ntd. vulgaris* NBRC 13699; 54, *Ntd. vulgaris* RCH1; 55, *Ntd. vulgaris* Hildenborough^T; 67, *Dsv. piger* ATCC 29098^T; 68, *Dsv. piger* FI11049; 72, *Dsv. legallii* H1^T; 73, *Dsv. legallii* AM18-2; 74, *Dsv. legallii* KHC7; 75, *Dsv. desulfuricans* DSM 642^T; 76, *Dsv. desulfuricans* IC1; 78, *Dsv. desulfuricans* ATCC 27774.

more intuitively in the 3D plot. The numbers in parentheses next to the scientific name in the following statements indicate the label in the figure. Four strains of *Ntd. vulgaris,* DP4, NBRC 13699, RCH1 and Hildenborough^T (52–55), were plotted closely (distance: 0.065 ± 0.035), whereas the other *Nitratidesulfovibrio vulgaris* strain, Miyazaki F (50), was plotted slightly further away (distance:1.59 ± 0.04). Similarly, *Dsv. piger* FI11049 (68), *Dsv. legallii* KHC7 (74), *Dsv. desulfuricans* IC1 (76), and *Dsv. desulfuricans* ATCC 27774 (78), which were separated into new species, showed sufficient distance to be separated from strains belonging to the existing group (**Figure 3B**).

Phylogenetic Analysis Based on 16S rRNA Sequence

Based on the previous analysis, 88 genomes were divided into 27 genera. During the process, two new genera were added and the possibility of six new species was confirmed. However, since there are many species in the Desulfovibrionaceae family without identified genome sequences, extended phylogenetic analysis using 16S rRNA sequence had to be performed. For this, type strains corresponding to all species and subspecies of taxa registered in the Ez-taxon DB were listed together. To compare with the previous core gene phylogenetic tree, 16S rRNA sequences of strains used for genome analysis were also added. With strains not registered with the 16S rRNA sequence, the sequence was directly extracted from the annotation results. The results are summarized in **Figure 4** (illustrated as a mirror image; the left side indicates previously published taxa names, and the right side indicates newly reorganized taxa names), which implies that most of the type strains belonging to Desulfovibrionaceae are well grouped according to new classification groups, as determined by our analysis. *Psd. tunisiensis* RB22^T belonged to the *Pseudodesulfovibrio* group in the phylogenetic analysis using the genome sequence. Still, it was classified to the *Salidesulfovibrio* group in analysis using the 16S rRNA sequence. *Dsv. cavernae* H1M^T, which was grouped with *Dhv. reitneri* L21-Syr-AB^T, appears to be far enough to be classified as a different genus. However, as reclassification using the genome sequence was impossible, the status of this species remains to be investigated.

Defining Phenotype and Genotype Characteristics for Each Group

To define the characteristics of the newly classified groups, a literature search was conducted on the characteristics of the phenotype for each strain. The information obtained is

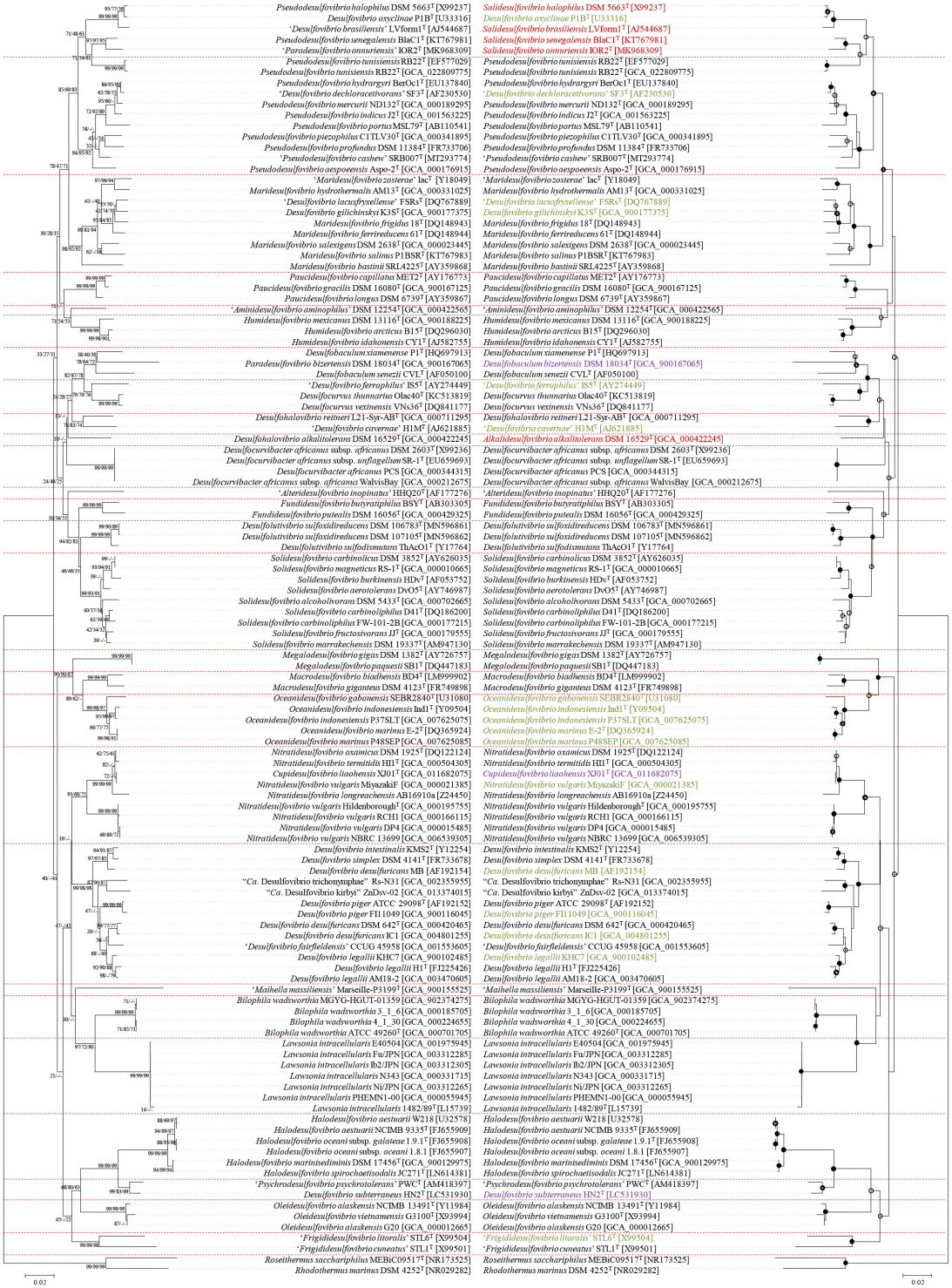


FIGURE 4 | Reclassification of the Desulfovibrionaceae by 16S rRNA gene sequence similarity inferred by genomic tree. This figure includes mirror image. Left sided tree is before reclassifications, Right sided tree is after reclassifications. Purple colored label, rearrangement through this study; Red colored label, proposed as a new genus; Green colored label, which has the potential to be newly classified, but requires further researches because it cannot be achived at this moment: lack of detailed phenotypic characterization or lack of available genome sequence or lack of available culture in two public culture collections. The tree is based on the Jukes & Cantor distances model and the neighbor-joining method with 1,000 bootstraps. Nodes with branch support > = 70% recovered by the three algorithms (the neighbor joining, the maximum-likelihood, and the maximum-parsimony algorithm) were indicated with ●; Nodes recovered by the three methods but with < 70% bootstrap values were indicated with ◎; Nodes recovered by two of the above methods were indicated with ○.

TABLE 1 | Phenotype and genotype information of the genera of the family Desulfovibrionaceae.

Genus		*Alkalidesulfovibrio*	*"Alteridesulfovibrio"*	*"Aminidesulfovibrio"*	*Bilophila*	*Desulfobaculum*	*Desulfocurvibacter*	*Desulfocurvus*	*Desulfohalovibrio*	*Desulfolutivibrio*
Morphology		Vibrio	Vibrio	Vibrio	Pleomorphic rod with swollen ends	v (Rod or Vibrio)	Vibrio	v (Rod or Vibrio)	v (Sigmoid or Vibrio)	Vibrio
Flagellation		Single polar	Single polar	Single polar	–	Single polar	Lophotrichous	Single polar	Single polar or Monopolar bitrichous	Single polar
Motility		+	+	+	–	+	+	+	+	+
G + C content (genome)		64.5	49.1	66.2	59.2–59.3	52.3–63.6	61.1–61.4	56.0–69.7	65.5	63.5–64.1
Respiration Quinone		MK-7	Nr	Nr	Nr	MK-7[a]	MK-7[b]	Nr	MK-7/MK-6[c]	MK-7
pH range (Opt.)		6.9–9.9 (9.0–9.4)	5.0–8.0 (6.9–7.2)	6.7–8.0 (7.5)	Nr	6.0–8.1 (7.0–7.6)	6.3–7.7 (7.0)	Nr (6.9–7.1)	5.5–8.5 (7.5)	6.5–8.5 (6.5–7.3)
Temp. range (Opt.) °C		16–47 (43)	15–42 (30)	25–40 (35)	Nr (35–37)	15–42 (35–40)	20–40 (37)	Nr (37–40)	20–48 (37–40)	15–45 (35–37)
Salinity. range (Opt.)%		0.085–0.7 (0.13)	0.5–2.0 (1.0)	0–2.0 (0.05–0.75)	Nr	0–12.5 (0.5–2.5)	0–4.25 (0–1.0)	0–5.0 (0.2)	0.7–18.0 (4.0–6.0)	0.1–2.0 (0.1–0.7)
Fermentative growth		+	–	+	Nr	v+	–	+	+	Nr
Major electron donors	All	H_2†, Formate†, Lactate, Pyruvate	H_2‡, Formate‡, Ethanol, Lactate, Pyruvate, Fumarate, Malate, Sugars	Formate‡, Ethanol‡, Lactate, Pyruvate, Amino acids	Pyruvate	H_2†, Lactate, Pyruvate	H_2†, Formate†, Ethanol, Lactate, Pyruvate	Formate‡, Lactate, Pyruvate	Formate†, Lactate, Pyruvate, Ethanol	H_2‡, Ethanol, Lactate
	Most					Fumarate, Succinate, Malate, Cysteine		H_2†		
	Some					C2–C4 alcohols		Iron	H_2†, C1 and C3-C5 alcohols, Fumarate, Succinate, Malate, Glycerol	Formate‡
Major electron acceptors	All	Sulfate, Sulfite, Thiosulfate	Sulfate, Sulfite, Thiosulfate	Sulfate, Sulfite, Thiosulfate	Nitrate	Sulfate, Sulfite	Sulfate, Sulfite, Thiosulfate, Sulfur(w)	Sulfate, Sulfite, Thiosulfate	Sulfate, Sulfite, Thiosulfate	Sulfate, Sulfite, Thiosulfate
	Most					Thiosulfate, Fumarate				
	Some					Sulfur			Nitrate	DMSO
Isolation source		Metal coupon in corroison monitoring reactor in District heating plant	Marine sediment	Wastewater	Intra-abdominal specimen	Marine sediment, solar saltern	Marine sediment, well water	Deep ground water, marine sediment, wastewater treatment reactor	Microbial mat of hypersaline lake, deep subsurface	Sewage sludge, freshwater mud
References		Abildgaard et al., 2006	Reichenbecher and Schink, 1997	Baena et al., 1998	Baron et al., 1989	Haouari et al., 2006; Zhao et al., 2012; Tsu et al., 1998	Brown et al., 2011; Campbell et al., 1966; Spring et al., 2019; Castañeda-Carrión et al., 2010; Brown et al., 2011	Klouche et al., 2009; Hamdi et al., 2013; Dinh et al., 2004	Spring et al., 2019; Sass and Cypionka, 2004	Thiel et al., 2020; Bak and Pfennig, 1987

Genus	*Desulfovibrio*	*"Frigididesulfovibrio"*	*Fundidesulfovibrio*	*Halodesulfovibrio*	*Humidesulfovibrio*	*Lawsonia*	*Macrodesulfovibrio*	*"Mailhella"*	*Maridesulfovibrio*
Morphology	v (Rod or sprillum or **Vibrio**)	Vibrio	Vibrio	Vibrio	Vibrio	Vibrio	v (Rod or vibrio)	Rod	v (Rod or **Vibrio**)
Flagellation	Single polar	Single polar	Single polar	Single polar	Single polar	–	Single polar	Nr	**Single polar** or Monopolar bitrichous
Motility	v+	+	+	+	v+	–	+	–	+
G + C content (genome)	53.6–64.2	36.9–53.5	62.8	45.0–46.2	65.0–65.6	32.9–33.1	54.6–59.5[v]	59.1	41.8–47.1
Respiration Quinone	MK-6[d]	Nr	MK-6(H_2)[e]	MK-6[f], MK-6(H2)[g]	MK-6(H2)	Nr	Nr	Nr	MK-6(H2)[h]
pH range (Opt.)	4.9–9.0 (7.0–7.5)	Nr	6.2–8.0 (7.0–7.1)	6.5–8.5 (7.0–8.0)	6.0–8.2 (6.5–7.2)	Nr	5.5–8.5 (7.5–7.9)	Nr	5.2–8.5 (5.8–7.8)
Temp. range (Opt.) °C	10–45 (28–37)	0–33 (28)	20–40 (30–35)	10–40 (20–37)	2–40 (24–37)	Nr (35–37)	15–45 (35–37)	Nr	0–50 (19–40)
Salinity. range (Opt.)%	0–3.0 (0–0.5)	0.25 (Nr)	0–6.0 (0–1.0)	0.05–6.0 (2.0–3.5)	0–2.0 (0–0.2)	Nr	0.2–8.0 (0.2–3.0)	Nr	0–12.0 (1.0–4.0)
Fermentative growth	v+	+	+	+	v+	Nr	+	Nr	v+

(Continued)

TABLE 1 | (Continued)

Genus		*Desulfovibrio*	*"Frigididesulfovibrio"*	*Fundidesulfovibrio*	*Halodesulfovibrio*	*Humidesulfovibrio*	*Lawsonia*	*Macrodesulfovibrio*	*"Mailhella"*	*Maridesulfovibrio*
Major electron donors	All	H_2‡, Ethanol, Lactate, Pyruvate	H_2‡, Formate‡, Lactate, Pyruvate, Fumarate	H_2†, Ethanol, C4 alcohols, Lactate, Pyruvate	Formate‡, Lactate, Pyruvate, Fumarate, Malate	H_2†, Formate†, Lactate, Pyruvate	Nr	H_2‡ Lactate, Pyruvate	Nr	Formate†, Lactate, Pyruvate,
	Most	Fumarate, Succinate			H_2‡, Glycerol, Succinate	Ethanol				H_2†, Ethanol, Fumarate
	Some	Formate‡, C3-C4 alcohols, Malate, Aromatic aldehydes, Alanine	Ethanol, Malate	C3 alcohols, C4–C5 fatty acids, Fumarate, Malate, Butyrate	Ethanol, C3 alcohols, Amino acids, Sugars	Fumarate, Succinate, Malate, Amino acids,		Formate‡, Ethanol, C3–C4 alcohols, Glycerol, Fumarate, Cysteine		Glycerol, C3–C4 alcohols, Succinate, Malate†, Amino acids, Sugars
Major electron acceptors	All	Sulfate, Thiosulfate	Sulfate, Sulfite, Thiosulfate, Sulfur	Sulfate, Thiosulfate	Sulfate, Sulfite, Thiosulfate	Sulfate, Sulfite, Thiosulfate, Sulfur, DMSO	Nr	Sulfate, Sulfite, Thiosulfate	Sulfate	Sulfate, Sulfite, Thiosulfate
	Most	Sulfite			Taurine, Fumarate					
	Some	Sulfur, Nitrate, Fumarate, Metal ions[#]		Sulfite, Fumarate		Fumarate		Sulfur		Sulfur, Fumarate
Isolation source		Hindgut of termite, rumen of a sheep, human faeces and blood, tar and sand mix, anaerobic sludge	Littoral sediment	Well water, sewage sludge	Anoxic sea sediment, sea water, soil	Anaerobic sludge of cheese wastewater, permafrost, sediment	Intestines of animals, intracellular parasite	Thermal spring, lagoon sediment	Human faeces	Saline and freshwater lake, sediment, hydrothermal chimney, seagrass, permafrost, deep subsurface oil well
References		Loubinoux et al., 2002; Zellner et al., 1989; Jyothsna et al., 2008; Thabet et al., 2010	Sass et al., 1998	Basso et al., 2005; Suzuki et al., 2010	Stüven, 1960; Postgate and Campbell, 1966; Finster and Kjeldsen, 2010; Shivani et al., 2017; Takii et al., 2008	Hernandez-Eugenia et al., 2000; Pecheritsyna et al., 2012; Sass et al., 2009	McOrist et al., 1995	Fadhlaoui et al., 2015; Esnault et al., 1988	Ndongo et al., 2017	Magot et al., 2004; Nielsen et al., 1999; Alazard et al., 2003; Postgate and Campbell, 1966; Ben Ali Gam et al., 2018; Ryzhmanova et al., 2019; Vandieken et al., 2006; Sattley and Madigan, 2010

Genus		*Megalodesulfovibrio*	*Nitratidesulfovibrio*	*Oceanidesulfovibrio*	*Oleidesulfovibrio*	*Paucidesulfovibrio*	*Pseudodesulfovibrio*	*Psychrodesulfovibrio*	*Salidesulfovibrio*	*Solidesulfovibrio*
Morphology		Vibrio	v (Sigmoid or **Vibrio**)	v (Sigmoid or **Vibrio**)	Vibrio	v (Rod or vibrio or spirillum)	Vibrio	Vibrio	v (Rod or **Vibrio**)	v (Rod or **Vibrio**)
Flagellation		Single polar	**Single polar** or Monotrichous polar	Single polar	Single polar	Single polar	Single polar or peritrichous	Single polar	v (Negative or **Single polar**)	Single polar
Motility		+	+	+	+	+	+	+	+	v+
G + C content (genome)		63.0–63.3	63.2–67.1	60.4–62.4	57.8–64.1	58.4–63.6	49.9–65.2	56.7–59.3	58.1–61.0	61.6–66.5
Respiration Quinone		MK-6[l]	Nr	Nr	Nr	Nr	MK-6(H2)[j]	MK-6(H2)[k]	Nr	MK-6[l], MK-7(H2)[m]
pH range (Opt.)		6.5–8.5 (7.2–7.4)	5.5–9.0 (6.6–7.4)	6.4–8.5 (6.9–7.3)	5.0–10.0 (7.0–7.5)	5.4–8.8 (6.8–7.4)	4.5–9.9 (6.5–7.5)	6.0–10.0 (7.2–9.0)	4.5–9.0 (6.5–7.6)	5.3–8.7 (6.5–7.3)
Temp. range (Opt.) °C		10–45 (30–36)	16–50 (35–37)	10–50 (30–37)	10–45 (37)	10–50 (35–40)	4–65 (25–35)	10–50 (28–35)	15–45 (30–40)	3–50 (29–38)
Salinity. range (Opt.)%		0–10 (0)	0–5.0 (0–0.1)	0–17.0 (5.0–6.0)	0–10.0 (2.5–5.0)	0–12.0 (1.0–6.0)	0–10.0 (0.6–8.0)	0–4.0 (0–0.5)	0.5–22.5 (3.0–10.0)	0–5.0 (0–1.0)
Fermentative growth		+	+	+	v	v–	+	+	v+	+
Major electron donors	All	H_2†, Formate†, Lactate, Pyruvate, Fumarate, Succinate, Malate	Lactate, Pyruvate	H_2‡, Formate‡, Ethanol, Lactate, Pyruvate, Fumarate, Succinate, Malate, Fructose	Formate‡, Ethanol, Glycerol, Lactate, Pyruvate, Fumarate, Malate	H_2‡, Formate‡, Lactate, Pyruvate	Lactate	Ethanol, Lactate, Pyruvate, Succinate, Malate	Lactate, Pyruvate	Ethanol, Lactate, Pyruvate, Fumarate
	Most		H_2†, Formate†, Ethanol, Fumarate, Malate, Organic acids				H_2†, Pyruvate, Fumarate, Malate		H_2†, Formate†	H_2‡, Formate‡, Glycerol, C3–C5 alcohols, Malate

(Continued)

TABLE 1 | (Continued)

Genus		*Megalodesulfovibrio*	*Nitratidesulfovibrio*	*Oceanidesulfovibrio*	*Oleidesulfovibrio*	*Paucidesulfovibrio*	*Pseudodesulfovibrio*	*Psychrodesulfovibrio*	*Salidesulfovibrio*	*Solidesulfovibrio*
	Some	Ethanol, C3–C5 alcohols, Glycerol	C1 and C4 alcohols, Sugars	C4 alcohols	Succinate, Ethanol, C4 alcohols		Formate†, Fatty acids^, Ethanol^, C1 and C3–C4 alcohols, Succinate, Amino acid	H_2†, Formate†, C4 alcohols, Fumarate, Cysteine	Ethanol, Glycerol, C1 and C3–C4 alcohols, Fumarate, Succinate, Malate, Amino acids	C1 alcohols‡, Succinate
Major electron acceptors	All	Sulfate, Sulfite, Thiosulfate, Sulfur	Sulfate, Sulfite	Sulfate, Sulfite, Thiosulfate, Sulfur	Sulfate, Sulfite Thiosulfate	Sulfate, Sulfite, Thiosulfate, Sulfur	Sulfate, Sulfite, Thiosulfate	Sulfate	Sulfate, Sulfite, Thiosulfate	Sulfate, Thiosulfate
	Most		Thiosulfate			Fumarate			Sulfur	Sulfite, Sulfur
	Some	Fumarate, Haloaromatic compounds	Sulfur, Iron(III), Nitrate, Nitrite, Oxygen, Fumarate	Fumarate	Fumarate		Sulfur, Fe(III), Nitrate, Nitrite, DMSO, Fumarate	Sulfite, Thiosulfate, Fe(III), Mn(IV), DMSO, Fumarate, AQDS	Fe(III), DMSO, Oxygen, Fumarate	Fumarate, Malate
Isolation source		Water, wastewater from a zinc smelter	Heavy metal impacted sediment, uranium mining waste piles, hindgut of a termite, degraded paddy field, oilfield fluids, chicken feed	Sea water from oilfield, water from oil pipeline	Oil field	Oil field	Marine sediment, deep subsurface, ground water	Mud and ice, deep sea surface sediment	Marine sediment, microbial mat in saline lake, oil refinery plant wastewater, dolomite	Waste water, sulfide-rich sediment, ricefield soil, contaminated ground water
References		Le Gall and Dragoni, 1966; Morais-Silva et al., 2014; Van Houten et al., 2009	Trinkerl et al., 1990; Aketagawa et al., 1985; Yagi, 1969; Ogata and Yagi, 1986; Ozawa et al., 2000; López-Cortés et al., 2006; Wan et al., 2021; Redburn and Patel, 1994; Voordouw, 2002; Heidelberg et al., 2004	Feio et al., 1998; Thabet et al., 2007; Tardy-Jacquenod et al., 1996	Dang et al., 1996; Feio et al., 2004	Magot et al., 2004; Miranda-Tello et al., 2003; Magot et al., 1992	Ranchou-Peyruse et al., 2018; Sun et al., 2000; Cao et al., 2016; Suzuki et al., 2010; Motamedi and Pedersen, 1998; Bale et al., 1997; Khelaifia et al., 2011 Ben Ali Gam et al., 2009	Jyothsna et al., 2008	Kim et al., 2020; Thioye et al., 2017; Warthmann et al., 2005; Krekeler et al., 1997; Caumette et al., 1991	Nanninga and Gottschal, 1987; Sakaguchi et al., 2002; Ouattara et al., 1999; Mogensen et al., 2005; Ramsay et al., 2015; Allen et al., 2008; Qatibi et al., 1991; Vainshtein et al., 1992; Ollivier et al., 1988; Chamkh et al., 2009

†: Require acetate or yeast extract.
Nr: Not reported.
‡: Require acetate.
Bold type: predominant characteristic.
^: Require 2-chlorophenol.
v: variable.
∨: G + C content measured based on HPLC.
v+: variable but positive is predominant.
#: only in sulfate-free conditions.
v-: variable but negative is predominant.
+ : positive.
(w): weak activity.
-: negative.
a: Desulfobaculum xiamenense.
b: Desulfocurvibacter africanus benghazi.
c: Desulfohalovibrio reitneri.
d: Desulfovibrio piger, Desulfovibrio legallii.
e: Fundidesulfovibrio butyratiphilus.
f: Halodesulfovibrio aestuarii, Halodesulfovibrio oceani, Halodesulfovibrio spirochaetisodalis.
g: Halodesulfovibrio marinisediminis.
h: Maridiesulfovibrio gilichinskyi.
i: Megalodesulfovibrio gigas.
j: Pseudodesulfovibrio portus.
k: Psychrodesulfovibrio subterraneus.
l: Solidesulfovibrio alcoholivorans, Solidesulfovibrio fructosivorans.
m: Solidesulfovibrio magneticus.

summarized in **Table 1** by group (genus). Most show morphology in the form of curved rods (vibrio) and have motility with a single polar flagellum. *Paucidesulfovibrio* and *Desulfovibrio* include members of the spirillum type and *Desulfohalovibrio*, *Oceanidesulfovibrio*, and *Nitratidesulfovibrio* include sigmoid types, inferring that one genus can show various morphologies, as well as vibrio. Regarding growth conditions, the characteristics of each group were more pronounced. Particularly, depending on whether the taxa comprised terrestrial-derived species or marine-derived species, they either did not require salt or demanded

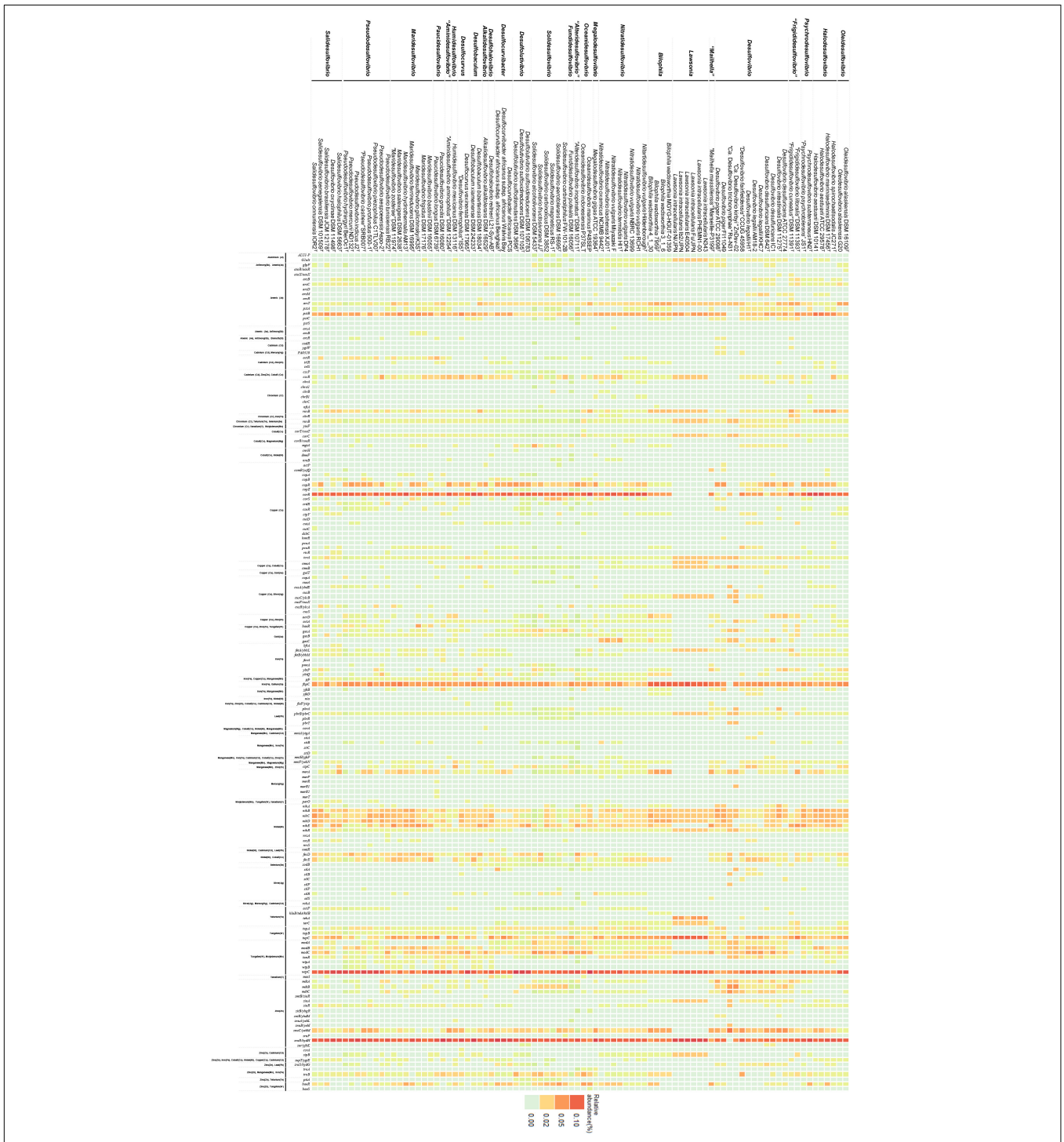


FIGURE 5 | Heatmap of metal related genes. This figure verifies that there is a distinguishable difference in pattern of metal resistance related genes between each group.

a NaCl concentration of 2% or more for optimal growth. In addition, the newly classified genera, *Alkalidesulfovibrio,* require an optimal pH of higher than 8, which is higher than that required by other genera (generally pH 7). Regarding the G + C content, the difference between the maximum and minimum values in each group of *Frigididesulfovibrio*, *Pseudodesulfovibrio, Desulfocurvus, Desulfobaculum,* and *Desulfovibrio* were 16.6, 13.9, 13.7, 11.3, and 10.6, respectively, showing a difference of over 10 mol%, whereas in other groups, the gap was only around 5 mol%.

Comparison of Metal Resistance-Related Gene Abundance Between Groups

Most of the Desulfovibrionaceae are isolated from habitats with high metal concentrations (**Supplementary Table 1**). This inspires interest in the search for metal resistance genes in the Desulfovibrionaceae family. Most species belonging to the family exhibit metal resistance by forming metal-sulfide precipitates, resulting from sulfate reduction. Besides forming metal-sulfide precipitates, several mechanisms are required for metal resistance, including import system regulation, efflux system, extracellular barrier, and reduction. Because many previous studies explored the genes related to sulfate reduction in Desulfovibrionaceae taxa, we focused on other metal resistance genes. Through this analysis, we aimed to explore species with the potential to show metal resistance, and to determine differences between species showing resistance potential and species that do not. The result of PCA of relative abundance data indicates significantly different patterns between each group (**Supplementary Figure 5**). Although not all species were investigated, obvious trends can be observed within the data. While most Desulfovibrionaceae family members display many metal resistance-related genes (**Figure 5**), several strains of the genus *Desulfovibrio*, *Lawsonia*, *Bilophila*, and "*Mailhella*" have low level gene abundances of metal import system regulators, such as *copR* and *corR*. They were commonly isolated from the biotic environment (intestine, feces, and blood). Considering that other strains were isolated from environments where metal elements are easily accessible (hydrothermal vent field, heavy metal affected sediment, aquifer, and mud), we estimate that the environmental condition of the habitat is reflected. Conversely, groups under the monophyly are adjacent in a PCA 3D plot (**Supplementary Figure 5**), which implies that gene abundance patterns of evolutionarily related species show a relative resemblance. Although the analysis was limited to species with identified genome sequences, we observed a distinction between each genus. Regarding *Solidesulfovibrio* and *Desulfovibrio*, the abundance of the multidrug efflux system is greater than other genera. The gene abundance related to arsenate methylation was relatively high in the genera *Solidesulfovibrio*, *Pseudodesulfovibrio*, *Desulfolutivibrio*, and *Desulfovibrio*. The mercury reductase gene, *merA*, displayed great abundance in the genera *Maridesulfovibrio*, *Desulfocurvibacter*, "*Mailhella*", *Desulfovibrio*, and *Bilophila* (refer to **Supplementary Table 8** for details).

DISCUSSION AND CONCLUSION

The literature shows several gaps in knowledge regarding the current taxonomy of the family Desulfovibrionaceae. Therefore, this study reclassified the mis-classified taxa. In this process, the entire species was reclassified and expanded to the type strain of the species without genome sequencing. According to previous analysis, a significant number of taxa require classification. However, to propose a new taxon or to change an existing taxon, Rule 30-3b of the International Code of Nomenclature of Prokaryotes (ICNP) must be followed (Parker et al., 2019), it indicates that the type strain must be deposited into two publicly recognized public culture collections. The classification cannot be ideally completed because some taxa of Desulfovibrionaceae do not meet the mentioned condition due to deposits remaining in only one culture collection or loss of resources (*Dsv. legallii* KHC7, *Dsv. desulfuricans* IC1, *Dsv. piger* FI11049, *Dsv. litorals* DSM 11393, "*Dsv. cavernae*" H1M, "*Dsv. ferrophilus*" IS5, "*Dsv. lacusfryxellense*" FSRs, "*Dsv. dechloracetivorans*" SF3, "*Dsv. oxyclinae*" P1B), while some others lack the complete phenotypic description and chemotaxonomic characterization required to be proposed as a new species (*Dsv. desulfuricans* ATCC 27774 and *Ntd. vulgaris* Miyazaki F), which are the common limitations shared with previous research (Galushko and Kuever, 2020a; Waite et al., 2020). Despite the unclear distinction between *Oceanidesulfovibrio* and *Macrodesulfovibrio* in phylogeny based on the 16S rRNA sequence, this study with AAI-based criteria could not confirm that *Oceanidesulfovibrio* and *Macrodesulfovibrio* should be reorganized into the same genera (**Figure 1**, **4**) due to the lack of reported genomes of strains belonging to those groups. Therefore, it is necessary to conduct a classification study applying the AAI classification standard value suggested when the genome of any type strain belonging to *Macrodesulfovibrio* is reported. In the case of *Pseudodesulfovibrio tunisiensis*, 16S rRNA phylogeny showed that the bacteria clearly belong to the *Salidesulfovibrio* group, but genome phylogeny showed that it belongs to the *Pseudodesulfovibrio* group. *Marixanthomonas spongiae* HN-E44^T has an AAI value of 85.2% with *Marixanthomonas ophiurae* JCM 14121^T. Similarly, AAI and 16S rRNA sequence similarity analysis results conflict also in the case of *Marixanthomonas spongiae* HN-E44^T and *Marixanthomonas ophiurae* JCM 14121^T. Two bacteria have an AAI value of 85.2%, and this exceeds the AAI criteria for genus delineation of Flavobacteriaceae. However, they showed only a 16S rRNA sequence similarity of 93.6%, which is lower than the minimum identity value of 94.9 $\pm$ 0.4% to guarantee the circumscription of a single genus (Yarza et al., 2008). This inconsistency is unaccountable yet, but several possibilities can be discussed. An introduction of an external 16S rRNA sequence might be charge of this discrepancy. However, the three 16S rRNA sequences from *Pseudodesulfovibrio tunisiensis* were highly similar, so this possibility should be excluded. Another possibility is that the evolutionary rate of the rRNA sequence and the genome did not match. It can happen when a large amount of external functional genes are introduced into the genome through various mechanisms, including horizontal gene transfer. This hypothesis is supported by the

vast G + C content range of the *Pseudodesulfovibrio* group from 49.9 to 65.2.

Finally, two novel genera were proposed and the possibility of five independent species was confirmed. In addition, nine species were reclassified into different genera. According to rule 23a of the ICNP, "Each taxon above species, up to and including order, with a given circumscription, position, and rank can bear only one correct name, that is, the earliest that is in accordance with the rules of this code." Therefore, for group 18, where *Nitratidesulfovibrio* and *Cupidesulfovibrio* collided, *Nitratidesulfovibrio* was established as the genus name of the group. In the same context, *Paradesulfovibrio bizertensis* of group 8 belongs to the same genus as *Desulfobaculum xiamenense* and is corrected as *Desulfobaculum bizertensis*. Despite the vacancy in the genus *Paradesulfovibrio*, group 1 deserves proposal as a new genus, and because they are of marine origin, we propose a new genus called *Salidesulfovibrio*.

From this study, objectivity and accuracy were obtained using indices, such as ANI, *d*DDH, and AAI, to compare genomic similarity and the classification was not based on phylogenetic analysis alone. As there are no clear standards for classifying the genus in the family Desulfovibrionaceae, an AAI cut-off value that did not incur a member included in multiple clusters for taxon delineation was defined to establish a clear standard. These accurately presented values will serve as a criterion to facilitate classification when new genera are added to this taxon.

Not all strains of the Desulfovibrionaceae family are extremophiles. However, strains belonging to this taxon have been reported to tolerate high heavy metal concentrations, and many are isolated from extreme marine environments such as hypersaline environments, deep sea sediment, and hydrothermal vent fields (with high metal concentration), and extreme terrestrial environments, such as regions contaminated with uranium, oil, and heavy metals. Because we noticed the frequent detection of this group in metal rich conditions, we explored the distribution of metal resistance-related genes through macroscopic *in silico* analysis. The analysis based on the genomes of Desulfovibrionaceae from the public database alone does not define characteristic of each group, yet it reveals a clear tendency within the given data. Based on taxa isolated from the biotic environment showed relatively low gene abundance and that evolutionarily close taxa were located closely on the PCA plot, it could be inferred that each genus showed a uniquely distinct gene abundance pattern under the influence of both environmental and evolutionary factors.

TAXONOMIC RECLASSIFICATIONS

Description of *Alkalidesulfovibrio* gen. nov.

Alkalidesulfovibrio [Al.ka.li.de.sul.fo.vi'.bri.o. N.L. n. *alkali* (from Arabic article *al* the; Arabic n. *qaliy* ashes of saltwort) *alkali*; N.L. masc. n. *Desulfovibrio* a bacterial genus; N.L. masc. n. *Alkalidesulfovibrio* a *Desulfovibrio* living in alkaline environment].

Cells are vibrio-shaped, 0.5–0.8 × 1.4–1.9 μm. DNA G + C content is 64.5 mol%. Cells are motile by a single polar flagellum. The member of this genus shows anaerobic respiration, but tolerates short exposure to oxygen. Fermentative growth is observed. Sulfate, thiosulfate, and sulfite serve as electron acceptors and are reduced to sulfide. H_2/CO_2 and formate can serve as an electron donor in the presence of yeast extract or acetate. Thermotolerant, the optimum temperature for growth is 43°C and alkaliphilic, the optimum pH for growth is 9.0–9.4. The major menaquinone is MK-7. Desulfoviridin is present. The type species is *Alkalidesulfovibrio alkalitolerans.*

Description of *Alkalidesulfovibrio alkalitolerans* comb. nov.

Basonym: *Desulfovibr*io *alkalitolerans* (Abildgaard et al., 2006).

Other synonym: *Desulfohalovibrio alkalitolerans* (Spring et al., 2019).

The description is the same given by Abildgaard et al. (2006). The type strain is DSM 16529^T (= $RT2^T$ = JCM 12612^T). Genome and 16S rRNA sequence accession number: NZ_ATHI00000000 and AY649785.

Description of *Salidesulfovibrio* gen. nov.

Salidesulfovibrio [Sa.li.de.sul.fo.vi'.bri.o. L. masc. n. *sal* (*gen. salis*), salt; N.L. masc. n. *Desulfovibrio* a bacterial genus; N.L. masc. n. *Salidesulfovibrio* a *Desulfovibrio* living in saline environment].

Cells are rod or vibrio shaped, 0.3–0.5 × 1.0–4.0 μm with motility. Lactate and pyruvate are used as electron donors and some members show fermentative growth. When yeast extract and cysteine or acetate are present, H_2/CO_2 and formate can be used. Sulfate can be reduced to sulfide, and some members can use sulfur and Fe(III) as electron acceptors. Mesophilic, the optimum temperature for growth is 30–40°C and neutrophilic, the optimum pH for growth is 6.5-7.6. The DNA G + C content is 55.8–61.0 mol%. The type species is *Salidesulfovibrio onnuriiensis*.

Description of *Salidesulfovibrio onnuriiensis* comb. nov. nom. rev.

Basonym: *'Paradesulfovibrio onnuriensis'* (Kim et al., 2020).

The description is the same given by Kim et al. (2020). The type strain is $IOR2^T$ (= KCTC 15845^T = MCCC $1K04559^T$). Genome and 16S rRNA sequence accession number: CP040751.1 and MK968309.

Description of *Salidesulfovibrio brasiliensis* comb. nov.

Basonym: *'Desulfovibrio brasiliensis'* (Warthmann et al., 2005).

The description is the same given by Warthmann et al. (2005). The type strain is JCM 12178^T (= DSM 15816^T = $LVform1^T$). Genome and 16S rRNA sequence accession number: NZ_BBCB01000000 and AJ544687.

Description of *Salidesulfovibrio halophilus* comb. nov.

Basonym: *Desulfovibrio halophilus* (Caumette et al., 1991).

Other synonym: *Pseudodesulfovibrio halophilus* (Waite et al., 2020).

The description is the same given by Caumette et al. (1991). The type strain is DSM 5663^T (= ATCC 51179^T = $SL8903^T$). Genome and 16S rRNA sequence accession number: SRX1760576 and X99237.

Description of *Salidesulfovibrio senegalensis* comb. nov.

Basonym: *Desulfovibrio senegalensis* (Thioye et al., 2017).

Other synonym: *Pseudodesulfovibrio senegalensis* (Galushko and Kuever, 2019), '*Paradesulfovibrio senegalensis*' (Kim et al., 2020).

The description is the same given by Thioye et al. (2017), Galushko and Kuever (2019), and Kim et al. (2020). The type strain is DSM 101509^T (= BLaC1^T = JCM 31063^T). Genome and 16S rRNA sequence accession number: NZ_WAIE00000000 and KT767981.

Emended Description of the Genus *Psychrodesulfovibrio* Galushko and Kuever, 2020d

Psychrodesulfovibrio [Psy.chro.de.sul.fo.vi'bri.o. Gr. masc. adj. *psychros*, cold; N.L. masc. n. *Desulfovibrio*, a bacterial genus; N.L. masc. n. *Psychrodesulfovibrio*, a *Desulfovibrio* living in the cold].

Cells are vibrio-shaped. DNA G + C content is 56.7-59.3 mol%. Cells are motile by a single polar flagellum. Strictly anaerobic with respiratory metabolism type. Sulfate serves as an electron acceptor. Succinate, malate, fumarate, pyruvate, lactate, and ethanol serve as electron donors with sulfate. Mesophilic, the optimum temperature for growth is 30-36^{O}C and neutrophilic, the optimum pH for growth is 7.2-9.0. The type species is *Psychrodesulfovibrio psychrotolerans*.

Description of *Psychrodesulfovibrio subterraneus* comb. nov.

Basonym: *Desulfovibrio subterraneus* (Ueno et al., 2021).

The description is the same given by Ueno et al. (2021). The type strain is HN2^T (= DSM 101010^T = NBRC 112213^T). Genome and 16S rRNA sequence accession number: NZ_BLVO00000000 and LC531930.

Description of *Desulfobaculum bizertensis* comb. nov.

Basonym: *Desulfovibrio bizertensis* (Haouari et al., 2006).

Other synonym: *Paradesulfovibrio bizentensis* (Waite et al., 2020).

The description is the same given by Haouari et al. (2006). The type strain is DSM 18034^T (= MB3^T = NCIMB 14199^T). Genome and 16S rRNA sequence accession number: SRX1760595 and DQ422859.

Description of *Nitratidesulfovibrio liaohensis* comb. nov.

Basonym: *Cupidesulfovibrio liaohensis* (Wan et al., 2021).

The description is the same given by Wan et al. (2021). The type strain is XJ01^T (= DSM 107637^T = CGMCC 1.5227^T). Genome and 16S rRNA sequence accession number: NZ_VSMK00000000.1 and MK260014.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

M-JP conducted every computational analysis and wrote the manuscript. YK and KK designed experiment and instructed manuscript. MP helped programming for data treatment and repetitive calculation process. JY, TN, and Y-RR conducted literature study and organized data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Korea Institute of Ocean Science & Technology (PE99922) and Korea Ministry of Oceans & Fisheries (20170411).

ACKNOWLEDGMENTS

We would like to gratefully thank editor and the two reviewers who shared their insights into Desulfovibrionaceae taxa and led constructive discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.738205/full#supplementary-material

Supplementary Figure 1 | Repetitive clustering and evaluation. Repetitive clustering and evaluation designed to find an appropriate AAI classification value without inclusion of a member in multiple cluster. Subjective intervention was avoided by repetitive clustering and evaluation with gradual increment of cut-off parameter to establish unbiased criteria, not setting strict borders after dividing the genus first.

Supplementary Figure 2 | Correlation between AAI and POCP. Red dots represent values between inter-genus strains, and blue dots represent values between intra-genus strains. The color of regression equation in the upper left corresponds to the color of each dot.

Supplementary Figure 3 | ANI and *d*DDH from pairwise whole-genome comparisons. The values of ANI and *d*DDH were expressed as heatmap. ANI is the lower left triangle indicated in violet color, and *d*DDH is the upper right triangle indicated in red color. Groups identified as the same species based on ANI and *d*DDH values are surrounded by black squares.

Supplementary Figure 4 | Comparison of the RSCU data of the 27 groups. This figure verifies that there is a distinguishable difference in codon usage between each group. A line plot showing the average value of RSCU for each group.

Supplementary Figure 5 | PCA results for metal resistance related gene abundance using z-score normalization. This figure verifies that there is a distinguishable difference in pattern of metal resistance related genes between each group.

Supplementary Table 1 | Isolation information and genomic indices by bacterial strains.

Supplementary Table 2 | The universal protein markers using for multi locus sequence alignments.

Supplementary Table 3 | AAI values between genomes.

Supplementary Table 4 | POCP values between genomes.

Supplementary Table 5 | ANI values between genomes.

Supplementary Table 6 | *d*DDH values between genomes.

Supplementary Table 7 | RSCU values of each genome.

Supplementary Table 8 | Relative abundance of metal related genes.

REFERENCES

Abildgaard, L., Nielsen, M. B., Kjeldsen, K. U., and Ingvorsen, K. (2006). *Desulfovibrio alkalitolerans* sp. nov., a novel alkalitolerant, sulphate-reducing bacterium isolated from district heating water. *Int. J. Syst. Evol. Microbiol.* 56, 1019–1024. doi: 10.1099/ijs.0.63909-0

Aketagawa, J., Kobayashi, K., and Ishimoto, M. (1985). Purification and properties of thiosulfate reductase from *Desulfovibrio vulgaris, Miyazaki F. J. Biochem.* 97, 1025–1032. doi: 10.1093/oxfordjournals.jbchem.a135144

Alazard, D., Dukan, S., Urios, A., Verhé, F., Bouabida, N., Morel, F., et al. (2003). *Desulfovibrio hydrothermalis* sp. nov., a novel sulfate-reducing bacterium isolated from hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* 53, 173–178. doi: 10.1099/ijs.0.02323-0

Aliyu, H., Lebre, P., Blom, J., Cowan, D., and De Maayer, P. (2016). Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. *Syst. Appl. Microbiol.* 39, 527–533. doi: 10.1016/j.syapm.2016.09.004

Allen, T. D., Kraus, P. F., Lawson, P. A., Drake, G. R., Balkwill, D. L., and Tanner, R. S. (2008). *Desulfovibrio carbinoliphilus* sp. nov., a benzyl alcohol-oxidizing, sulfate-reducing bacterium isolated from a gas condensate-contaminated aquifer. *Int. J. Syst. Evol. Microbiol.* 58, 1313–1317. doi: 10.1099/ijs.0.65524-0

Arakaki, A., Takeyama, H., Tanaka, T., and Matsunaga, T. (2002). "Cadmium recovery by a sulfate-reducing magnetotactic bacterium, *Desulfovibrio magneticus* RS-1, using magnetic separation," in *Biotechnology for Fuels and Chemicals*, eds B. H. Davison, C. E. Wyman, and M. Finkelstein (Berlin: Springer), 833–840. doi: 10.1385/abab:98-100:1-9:833

Ayangbenro, A. S., Olanrewaju, O. S., and Babalola, O. O. (2018). Sulfate-reducing bacteria as an effective tool for sustainable acid mine bioremediation. *Front. Microbiol.* 9:1986. doi: 10.3389/fmicb.2018.01986

Baena, S., Fardeau, M.-L., Labat, M., Ollivier, B., Garcia, J.-L., and Patel, B. (1998). *Desulfovibrio aminophilus* sp. nov., a novel amino acid degrading and sulfate reducing bacterium from an anaerobic dairy wastewater lagoon. *Syst. Appl. Microbiol.* 21, 498–504. doi: 10.1016/s0723-2020(98)80061-1

Bak, F., and Pfennig, N. (1987). Chemolithotrophic growth of *Desulfovibrio sulfodismutans* sp. nov. by disproportionation of inorganic sulfur compounds. *Arch. Microbiol.* 147, 184–189. doi: 10.1007/bf00415282

Bale, S., Goodman, K., Rochelle, P. A., Marchesi, J. R., Fry, J. C., Weightman, A. J., et al. (1997). *Desulfovibrio profundus* sp. nov., a novel barophilic sulfate-reducing bacterium from deep sediment layers in the Japan sea. *Int. J. Syst. Evol. Microbiol.* 47, 515–521. doi: 10.1099/00207713-47-2-515

Barco, R., Garrity, G., Scott, J., Amend, J., Nealson, K., and Emerson, D. (2020). A genus definition for bacteria and archaea based on a standard genome relatedness index. *MBio* 11:e02475-19. doi: 10.1101/392480

Baron, E. J., Summanen, P., Downes, J., Roberts, M. C., Wexler, H., and Finegold, S. M. (1989). *Bilophila wadsworthia*, gen. nov. and sp. nov., a unique gram-negative anaerobic rod recovered from appendicitis specimens and human faeces. *Microbiology* 135, 3405–3411. doi: 10.1099/00221287-135-12-3405

Barton, L., Legall, J., Odom, J., and Peck, H. Jr. (1983). Energy coupling to nitrite respiration in the sulfate-reducing bacterium *Desulfovibrio gigas*. *J. Bacteriol.* 153, 867–871. doi: 10.1128/jb.153.2.867-871.1983

Basso, O., Caumette, P., and Magot, M. (2005). *Desulfovibrio putealis* sp. nov., a novel sulfate-reducing bacterium isolated from a deep subsurface aquifer. *Int. J. Syst. Evol. Microbiol.* 55, 101–104. doi: 10.1099/ijs.0.63303-0

Ben Ali Gam, Z., Thioye, A., Cayol, J.-L., Joseph, M., Fauque, G., and Labat, M. (2018). Characterization of *Desulfovibrio salinus* sp. nov., a slightly halophilic sulfate-reducing bacterium isolated from a saline lake in Tunisia. *Int. J. Syst. Evol. Microbiol.* 68, 715–720. doi: 10.1099/ijsem.0.002567

Ben Ali Gam, Z., Oueslati, R., Abdelkafi, S., Casalot, L., Tholozan, J. L., and Labat, M. (2009). *Desulfovibrio tunisiensis* sp. nov., a novel weakly halotolerant, sulfate-reducing bacterium isolated from exhaust water of a Tunisian oil refinery. *Int. J. Syst. Evol. Microbiol.* 59, 1059–1063. doi: 10.1099/ijs.0.000943-0

Brown, S. D., Wall, J. D., Kucken, A. M., Gilmour, C. C., Podar, M., Brandt, C. C., et al. (2011). Genome sequence of the mercury-methylating and pleomorphic *Desulfovibrio africanus* strain Walvis Bay. *J. Bacteriol.* 193, 4037–4038. doi: 10.1128/jb.05223-11

Campbell, L. L., Kasprzycki, M. A., and Postgate, J. R. (1966). *Desulfovibrio africanus* sp. n., a new dissimilatory sulfate-reducing bacterium. *J. Bacteriol.* 92, 1122–1127. doi: 10.1128/jb.92.4.1122-1127.1966

Cao, J., Gayet, N., Zeng, X., Shao, Z., Jebbar, M., and Alain, K. (2016). Pseudodesulfovibrio indicus gen. nov., sp. nov., a piezophilic sulfate-reducing bacterium from the Indian Ocean and reclassification of four species of the genus *Desulfovibrio*. *Int. J. Syst. Evol. Microbiol.* 66, 3904–3911. doi: 10.1099/ijsem.0.001286

Castañeda-Carrión, I. N., Sheik, C. S., and Krumholz, L. R. (2010). *Desulfovibrio africanus* subsp. *uniflagellum* subsp. nov., a sulfate-reducing bacterium from a uranium-contaminated subsurface aquifer. *Int. J. Syst. Evol. Microbiol.* 60, 880–886. doi: 10.1099/ijs.0.006668-0

Caumette, P., Cohen, Y., and Matheron, R. (1991). Isolation and characterization of *Desulfovibrio halophilus* sp. nov., a halophilic sulfate-reducing bacterium isolated from Solar Lake (Sinai). *Syst. Appl. Microbiol.* 14, 33–38. doi: 10.1016/s0723-2020(11)80358-9

Chamkh, F., Spröer, C., Lemos, P. C., Besson, S., El Asli, A.-G., Bennisse, R., et al. (2009). *Desulfovibrio marrakechensis* sp. nov., a 1, 4-tyrosol-oxidizing, sulfate-reducing bacterium isolated from olive mill wastewater. *Int. J. Syst. Evol. Microbiol.* 59, 936–942. doi: 10.1099/ijs.0.003822-0

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 31, 3497–3500. doi: 10.1093/nar/gkg500

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

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

Dang, P. N., Dang, T. C. H., Lai, T. H., and Stan-Lotter, H. (1996). *Desulfovibrio vietnamensis* sp. nov., a halophilic sulfate-reducing bacterium from Vietnamese oil fields. *Anaerobe* 2, 385–392.

Dinh, H. T., Kuever, J., Mußmann, M., Hassel, A. W., Stratmann, M., and Widdel, F. (2004). Iron corrosion by novel anaerobic microorganisms. *Nature* 427, 829–832. doi: 10.1038/nature02321

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

Esnault, G., Caumette, P., and Garcia, J.-L. (1988). Characterization of *Desulfovibrio giganteus* sp. nov., a sulfate-reducing bacterium isolated from a brackish coastal lagoon. *Syst. Appl. Microbiol.* 10, 147–151. doi: 10.1016/s0723-2020(88)80029-8

Fadhlaoui, K., Hania, W. B., Postec, A., Fauque, G., Hamdi, M., Ollivier, B., et al. (2015). Characterization of *Desulfovibrio biadhensis* sp. nov., isolated from a thermal spring. *Int. J. Syst. Evol. Microbiol.* 65, 1256–1261. doi: 10.1099/ijs.0.000088

Feio, M. J., Beech, I. B., Carepo, M., Lopes, J. M., Cheung, C. W. S., Franco, R., et al. (1998). Isolation and characterization of a novel sulphate-reducing bacterium of the *Desulfovibrio* genus. *Anaerobe* 4, 117–130. doi: 10.1006/anae.1997.0142

Feio, M. J., Zinkevich, V., Beech, I. B., Llobet-Brossa, E., Eaton, P., Schmitt, J., et al. (2004). *Desulfovibrio alaskensis* sp. nov., a sulphate-reducing bacterium from a soured oil reservoir. *Int. J. Syst. Evol. Microbiol.* 54, 1747–1752. doi: 10.1099/ijs.0.63118-0

Finster, K. W., and Kjeldsen, K. U. (2010). *Desulfovibrio oceani* subsp. *oceani* sp. nov., subsp. nov. and *Desulfovibrio oceani* subsp. *galateae* subsp. nov., novel sulfate-reducing bacteria isolated from the oxygen minimum zone off the coast of Peru. *Antonie Van Leeuwenhoek* 97, 221–229. doi: 10.1007/s10482-009-9403-y

Galushko, A., and Kuever, J. (2019). "Pseudodesulfovibrio," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Hoboken, NJ: Wiley), 1–11. doi: 10.1002/9781118960608.gbm01574

Galushko, A., and Kuever, J. (2020b). "*Macrodesulfovibrio* gen. nov," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons, Inc.), 1–5. doi: 10.1002/9781118960608.gbm01893

Galushko, A., and Kuever, J. (2020c). "*Oceanidesulfovibrio* gen. nov," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons, Inc.), 1–5. doi: 10.1002/9781118960608.gbm01894

Galushko, A., and Kuever, J. (2020d). "*Psychrodesulfovibrio* gen. nov," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman

(Hoboken, NJ: John Wiley & Sons, Inc.), 1–5. doi: 10.1002/9781118960608.gbm01892

Galushko, A., and Kuever, J. (2020a). "Desulfovibrionaceae," in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. W. B. Whitman (Hoboken, NJ: John Wiley & Sons, Inc.), 1–13. doi: 10.1002/9781118960608.fbm00199.pub2

Goñi-Urriza, M., Klopp, C., Ranchou-Peyruse, M., Ranchou-Peyruse, A., Monperrus, M., Khalfaoui-Hassani, B., et al. (2020). Genome insights of mercury methylation among *Desulfovibrio* and *Pseudodesulfovibrio* strains. *Res. Microbiol.* 171, 3–12. doi: 10.1016/j.resmic.2019.10.003

Goulhen, F., Gloter, A., Guyot, F., and Bruschi, M. (2006). Cr (VI) detoxification by *Desulfovibrio vulgaris* strain Hildenborough: microbe–metal interactions studies. *Appl. Microbiol. Biotecnhol.* 71, 892–897. doi: 10.1007/s00253-005-0211-7

Gu, H., Chu, D. K., Peiris, M., and Poon, L. L. (2020). Multivariate analyses of codon usage of SARS-CoV-2 and other *betacoronaviruses*. *Virus Evol.* 6:veaa032. doi: 10.1093/ve/veaa032

Hamdi, O., Hania, W. B., Postec, A., Bartoli, M., Hamdi, M., Bouallagui, H., et al. (2013). Isolation and characterization of *Desulfocurvus thunnarius* sp. nov., a sulfate-reducing bacterium isolated from an anaerobic sequencing batch reactor treating cooking wastewater. *Int. J. Syst. Evol. Microbiol.* 63, 4237–4242. doi: 10.1099/ijs.0.051664-0

Haouari, O., Fardeau, M.-L., Casalot, L., Tholozan, J.-L., Hamdi, M., and Ollivier, B. (2006). Isolation of sulfate-reducing bacteria from Tunisian marine sediments and description of *Desulfovibrio bizertensis* sp. nov. *Int. J. Syst. Evol. Microbiol.* 56, 2909–2913. doi: 10.1099/ijs.0.64530-0

Heidelberg, J. F., Seshadri, R., Haveman, S. A., Hemme, C. L., Paulsen, I. T., Kolonay, J. F., et al. (2004). The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat. Biotechnol.* 22, 554–559. doi: 10.1038/nbt959

Hernandez-Eugenia, G., Fardeau, M.-L., Patel, B. K., Macarie, H., Garcia, J.-L., and Ollivier, B. (2000). *Desulfovibrio mexicanus* sp. nov., a Sulfate-reducing bacterium isolated from an upflow anaerobic sludge blanket (UASB) reactor treating cheese wastewaters. *Anaerobe* 6, 305–312. doi: 10.1006/anae.2000.0354

Jeong, O. J., Choi, J.-H., Kim, I. H., Kim, Y.-K., and Oh, B.-K. (2015). Effective bioremediation of Cadmium (II), nickel (II), and chromium (VI) in a marine environment by using *Desulfovibrio desulfuricans*. *Biotechnol. Bioprocess Eng.* 20, 937–941. doi: 10.1007/s12257-015-0287-6

Jukes, T. H., and Cantor, C. R. (1969). "Evolution of protein molecules," in *Mammalian Protein Metabolism*, ed. H. N. Munro (New York, NY: Academic press), 21–132. doi: 10.1016/b978-1-4832-3211-9.50009-7

Just, S., Mondot, S., Ecker, J., Wegner, K., Rath, E., Gau, L., et al. (2018). The gut microbiota drives the impact of bile acids and fat source in diet on mouse metabolism. *Microbiome* 6, 1–18. doi: 10.1186/s40168-018-0510-8

Jyothsna, T. S., Sasikala, C., and Ramana, C. V. (2008). *Desulfovibrio psychrotolerans* sp. nov., a psychrotolerant and moderately alkaliphilic sulfate-reducing deltaproteobacterium from the Himalayas. *Int. J. Syst. Evol. Microbiol.* 58, 821–825. doi: 10.1099/ijs.0.65402-0

Keller, K. L., and Wall, J. D. (2011). Genetics and molecular biology of the electron flow for sulfate respiration in *Desulfovibrio*. *Front. Microbiol.* 2:135.

Keller, K. L., Rapp-Giles, B. J., Semkiw, E. S., Porat, I., Brown, S. D., and Wall, J. D. (2014). New model for electron flow for sulfate reduction in *Desulfovibrio alaskensis* G20. *Appl. Environ. Microb.* 80, 855–868. doi: 10.1128/aem.02963-13

Kellogg, W. W., Cadle, R., Allen, E., Lazrus, A., and Martell, E. (1972). The sulfur cycle. *Science* 175, 587–596. doi: 10.1126/science.175.4022.587

Kim, Y. J., Yang, J.-A., Lim, J. K., Park, M.-J., Yang, S.-H., Lee, H. S., et al. (2020). *Paradesulfovibrio onnuriensis* gen. nov., sp. nov., a chemolithoautotrophic sulfate-reducing bacterium isolated from the Onnuri vent field of the Indian Ocean and reclassification of *Desulfovibrio senegalensis* as *Paradesulfovibrio senegalensis* comb. nov. *J. Microbiol.* 58, 252–259. doi: 10.3389/fmicb.2011.00135

Klouche, N., Basso, O., Lascourreges, J.-F., Cayol, J.-L., Thomas, P., Fauque, G., et al. (2009). *Desulfocurvus vexinensis* gen. nov., sp. nov., a sulfate-reducing bacterium isolated from a deep subsurface aquifer. *Int. J. Syst. Evol. Microbiol.* 59, 3100–3104. doi: 10.1099/ijs.0.010363-0

Kluyver, A. J., and Van Niel, C. B. (1936). Prospects for a natural system of classification of bacteria. *Zentralbl. Bakter. Par.* 94, 369–403. doi: 10.2307/j.ctv176g0.4

Kobayashi, K., Hasegawa, H., Takagi, M., and Ishimoto, M. (1982). Proton translocation associated with sulfite reduction in a sulfate-reducing bacterium, *Desulfovibrio vulgaris*. *FEBS Lett.* 142, 235–237. doi: 10.1016/0014-5793(82)80142-7

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

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

Krekeler, D., Sigalevich, P., Teske, A., Cypionka, H., and Cohen, Y. (1997). A sulfate-reducing bacterium from the oxic layer of a microbial mat from solar lake (Sinai), *Desulfovibrio oxyclinae* sp. nov. *Arch. Microbiol.* 167, 369–375. doi: 10.1007/s002030050457

Kuever, J., Rainey, F., and Widdel, F. (2005). "Family I. *Desulfovibrionaceae* fam. nov," in *Bergey's Manual of Systematic Bacteriology, The Proteobacteria*, 2nd Edn, eds D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (New York, NY: Springer-Verlag), 926–938.

Le Gall, J., and Dragoni, N. (1966). Dependance of sulfite reduction on a crystallized ferredoxin from *Desulfovibrio gigas*. *Biochem. Biophys. Res. Commun.* 23, 145–149. doi: 10.1016/0006-291x(66)90519-5

Li, Y., Xue, H., Sang, S.-Q., Lin, C.-L., and Wang, X.-Z. (2017). Phylogenetic analysis of family *Neisseriaceae* based on genome sequences and description of *Populibacter corticis* gen. nov., sp. nov., a member of the family *Neisseriaceae*, isolated from symptomatic bark of *Populus*× *euramericana* canker. *PLoS One* 12:e0174506. doi: 10.1371/journal.pone.0174506

Lie, T. J., Pitta, T., Leadbetter, E. R., Godchaux Iii, W., and Leadbetter, J. R. (1996). Sulfonates: novel electron acceptors in anaerobic respiration. *Arch. Microbiol.* 166, 204–210. doi: 10.1007/s002030050376

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

López-Cortés, A., Fardeau, M.-L., Fauque, G., Joulian, C., and Ollivier, B. (2006). Reclassification of the sulfate-and nitrate-reducing bacterium *Desulfovibrio vulgaris* subsp. *oxamicus* as *Desulfovibrio oxamicus* sp. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 56, 1495–1499. doi: 10.1099/ijs.0.64074-0

Loubinoux, J., Valente, F. M., Pereira, I. A., Costa, A., Grimont, P. A., and Le Faou, A. E. (2002). Reclassification of the only species of the genus *Desulfomonas*, *Desulfomonas pigra*, as *Desulfovibrio piger* comb. nov. *Int. J. Syst. Evol. Microbiol.* 52, 1305–1308. doi: 10.1099/00207713-52-4-1305

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

Magot, M., Basso, O., Tardy-Jacquenod, C., and Caumette, P. (2004). *Desulfovibrio bastinii* sp. nov. and *Desulfovibrio gracilis* sp. nov., moderately halophilic, sulfate-reducing bacteria isolated from deep subsurface oilfield water. *Int. J. Syst. Evol. Microbiol.* 54, 1693–1697. doi: 10.1099/ijs.0.02977-0

Magot, M., Caumette, P., Desperrier, J., Matheron, R., Dauga, C., Grimont, F., et al. (1992). *Desulfovibrio longus* sp. nov., a sulfate-reducing bacterium isolated from an oil-producing well. *Int. J. Syst. Evol. Microbiol.* 42, 398–402. doi: 10.1099/00207713-42-3-398

Matias, P. M., Pereira, I. S. A., Soares, C. M., and Carrondo, M. A. (2005). Sulphate respiration from hydrogen in *Desulfovibrio* bacteria: a structural biology overview. *Prog. Biophys. Mol. Bio.* 89, 292–329. doi: 10.1016/j.pbiomolbio.2004.11.003

McOrist, S., Gebhart, C. J., Boid, R., and Barns, S. M. (1995). Characterization of *Lawsonia intracellularis* gen. nov., sp. nov., the obligately intracellular bacterium of porcine proliferative enteropathy. *Int. J. Syst. Evol. Microbiol.* 45, 820–825. doi: 10.1099/00207713-45-4-820

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

Miranda-Tello, E., Fardeau, M.-L., Fernández, L., RamıRez, F., Cayol, J.-L., Thomas, P., et al. (2003). *Desulfovibrio capillatus* sp. nov., a novel

sulfate-reducing bacterium isolated from an oil field separator located in the Gulf of Mexico. *Anaerobe* 9, 97–103. doi: 10.1016/s1075-9964(03)00064-7

Mogensen, G. L., Kjeldsen, K. U., and Ingvorsen, K. (2005). *Desulfovibrio aerotolerans* sp. nov., an oxygen tolerant sulphate-reducing bacterium isolated from activated sludge. *Anaerobe* 11, 339–349. doi: 10.1016/j.anaerobe.2005.04.002

Moore, W. C., Johnson, J., and Holdeman, L. (1976). Emendation of *Bacteroidaceae* and *Butyrivibrio* and descriptions of *Desulfomonas* gen. nov. and ten new species in the genera *Desulfomonas*, *Butyrivibrio*, *Eubacterium*, *Clostridium*, and *Ruminococcus*. *Int. J. Syst. Evol. Microbiol.* 26, 238–252. doi: 10.1099/00207713-26-2-238

Moore, W. G., Barnabas, J., and Goodman, M. (1973). A method for constructing maximum parsimony ancestral amino acid sequences on a given network. *J. Theor. Biol.* 38, 459–485. doi: 10.1016/0022-5193(73)90252-x

Morais-Silva, F. O., Rezende, A. M., Pimentel, C., Santos, C. I., Clemente, C., Varela-Raposo, A., et al. (2014). Genome sequence of the model sulfate reducer *Desulfovibrio gigas*: a comparative analysis within the *Desulfovibrio* genus. *Microbiologyopen* 3, 513–530. doi: 10.1002/mbo3.184

Motamedi, M., and Pedersen, K. (1998). Note: *Desulfovibrio aespoeensis* sp. nov., a mesophilic sulfate-reducing bacterium from deep groundwater at Aspö hard rock laboratory, Sweden. *Int. J. Syst. Evol. Microbiol.* 48, 311–315. doi: 10.1099/00207713-48-1-311

Murros, K. E., Huynh, V. A., Takala, T. M., and Saris, P. E. (2021). *Desulfovibrio* bacteria are associated with Parkinson's disease. *Front. Cell. Infect. Microbiol.* 11:378. doi: 10.3389/fcimb.2021.652617

Nanninga, H. J., and Gottschal, J. C. (1987). Properties of *Desulfovibrio carbinolicus* sp. nov. and other sulfate-reducing bacteria isolated from an anaerobic-purification plant. *Appl. Environ. Microbiol.* 53, 802–809. doi: 10.1128/aem.53.4.802-809.1987

Ndongo, S., Cadoret, F., Dubourg, G., Delerce, J., Fournier, P.-E., Raoult, D., et al. (2017). '*Collinsella phocaeensis*' sp. nov.,'*Clostridium merdae*'sp. nov.,'*Sutterella massiliensis*' sp. nov.,'*Sutturella timonensis*' sp. nov.,'*Enorma phocaeensis*' sp. nov.,'*Mailhella massiliensis*' gen. nov., sp. nov.,'*Mordavella massiliensis*' gen. nov., sp. nov. and '*Massiliprevotella massiliensis*' gen. nov., sp. nov., 9 new species isolated from fresh stool samples of healthy French patients. *New Microbes New Infect.* 17, 89–95. doi: 10.1016/j.nmni.2017.02.005

Nielsen, J. T., Liesack, W., and Finster, K. (1999). *Desulfovibrio zosterae* sp. nov., a new sulfate reducer isolated from surface-sterilized roots of the seagrass *Zostera marina*. *Int. J. Syst. Evol. Microbiol.* 49, 859–865. doi: 10.1099/00207713-49-2-859

Ogata, M., and Yagi, T. (1986). Pyruvate dehydrogenase and the path of lactate degradation in *Desulfovibrio vulgaris* Miyazaki F. *J. Biochem.* 100, 311–318. doi: 10.1093/oxfordjournals.jbchem.a121717

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., et al. (2013). *Package 'vegan'*. Available online at: https://CRAN.R-project.org/package=vegan (accessed September 2, 2021).

Ollivier, B., Cord-Ruwisch, R., Hatchikian, E., and Garcia, J.-L. (1988). Characterization of *Desulfovibrio fructosovorans* sp. nov. *Arch. Microbiol.* 149, 447–450. doi: 10.1007/bf00425586

Orata, F. D., Meier-Kolthoff, J. P., Sauvageau, D., and Stein, L. Y. (2018). Phylogenomic analysis of the gamma*proteobacteria*l methanotrophs (Order *Methylococcales*) calls for the reclassification of members at the genus and species levels. *Front. Microbiol.* 9:3162. doi: 10.3389/fmicb.2018.03162

Ouattara, A., Patel, B., Cayol, J.-L., Cuzin, N., Traore, A., and Garcia, J.-L. (1999). Note: Isolation and characterization of *Desulfovibrio burkinensis* sp. nov. from an African ricefield, and phylogeny of *Desulfovibrio alcoholivorans*. *Int. J. Syst. Evol. Microbiol.* 49, 639–643. doi: 10.1099/00207713-49-2-639

Ozawa, K., Meikari, T., Motohashi, K., Yoshida, M., and Akutsu, H. (2000). Evidence for the presence of an F-type ATP synthase involved in sulfate respiration in *Desulfovibrio vulgaris*. *J. Bacteriol.* 182, 2200–2206. doi: 10.1128/jb.182.8.2200-2206.2000

Parker, C. T., Tindall, B. J., and Garrity, G. M. (2019). International code of nomenclature of prokaryotes: prokaryotic code (2008 revision). *Int. J. Syst. Evol. Microbiol.* 69, S1–S111. doi: 10.1099/ijsem.0.000778

Pecheritsyna, S. A., Rivkina, E. M., Akimov, V. N., and Shcherbakova, V. A. (2012). *Desulfovibrio arcticus* sp. nov., a psychrotolerant sulfate-reducing bacterium from a cryopeg. *Int. J. Syst. Evol. Microbiol.* 62, 33–37. doi: 10.1099/ijs.0.021451-0

Peck, H. D. Jr. (1959). The ATP-dependent reduction of sulfate with hydrogen in extracts of *Desulfovibrio desulfuricans*. *Proc. Natl. Acad. Sci. U.S.A.* 45, 701.

Peck, H. Jr. (1961). Enzymatic basis for assimilatory and dissimilatory sulfate reduction. *J. Bacteriol.* 82, 933–939. doi: 10.1128/jb.82.6.933-939.1961

Pereira, I. A. C. (2008). "Respiratory membrane complexes of *Desulfovibrio*," in *Microbial Sulfur Metabolism*, eds C. Dahl and C. G. Friedrich (Berlin: Springer), 24–35. doi: 10.1007/978-3-540-72682-1_3

Pérez-Cataluña, A., Salas-Massó, N., Diéguez, A. L., Balboa, S., Lema, A., Romalde, J. L., et al. (2018). Revisiting the taxonomy of the genus *Arcobacter*: getting order from the chaos. *Front. Microbiol.* 9:2077. doi: 10.3389/fmicb.2018.02077

Postgate, J. (1956). Cytochrome c3 and desulphoviridin; pigments of the anaerobe *Desulphovibrio desulphuricans*. *Microbiology* 14, 545–572. doi: 10.1099/00221287-14-3-545

Postgate, J. (1984). *The Sulphate Reducing Bacteria*. Cambridge: Cambridge University Press.

Postgate, J. R., and Campbell, L. L. (1966). Classification of Desulfovibrio species, the nonsporulating sulfate-reducing bacteria. *Bacteriol. Rev.* 30, 732–738. doi: 10.1128/mmbr.30.4.732-738.1966

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Qatibi, A., Niviere, V., and Garcia, J.-L. (1991). *Desulfovibrio alcoholovorans* sp. nov., a sulfate-reducing bacterium able to grow on glycerol, 1, 2-and 1, 3-propanediol. *Arch. Microbiol.* 155, 143–148. doi: 10.1007/bf00248608

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

Ramsay, B. D., Hwang, C., Woo, H. L., Carroll, S. L., Lucas, S., Han, J., et al. (2015). High-quality draft genome sequence of *Desulfovibrio carbinoliphilus* FW-101-2B, an organic acid-oxidizing sulfate-reducing bacterium isolated from uranium (VI)-contaminated groundwater. *Genome Announc.* 3:e00092-15. doi: 10.1128/genomea.00092-15

Ranchou-Peyruse, M., Goñi-Urriza, M., Guignard, M., Goas, M., Ranchou-Peyruse, A., and Guyoneaud, R. (2018). *Pseudodesulfovibrio hydrargyri* sp. nov., a mercury-methylating bacterium isolated from a brackish sediment. *Int. J. Syst. Evol. Microbiol.* 68, 1461–1466. doi: 10.1099/ijsem.0.002173

Redburn, A., and Patel, B. (1994). *Desulfovibrio longreachii* sp. nov., a sulfate-reducing bacterium isolated from the Great Artesian Basin of Australia. *FEMS Microbiol. Lett.* 115, 33–38. doi: 10.1111/j.1574-6968.1994.tb06610.x

Reichenbecher, W., and Schink, B. (1997). *Desulfovibrio inopinatus*, sp. nov., a new sulfate-reducing bacterium that degrades hydroxyhydroquinone (1, 2, 4-trihydroxybenzene). *Arch. Microbiol.* 168, 338–344. doi: 10.1007/s002030050546

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

Rodriguez-R, L. M., and Konstantinidis, K. T. (2014). Bypassing cultivation to identify bacterial species. *Microbe* 9, 111–118.

Ryzhmanova, Y., Abashina, T., Petrova, D., and Shcherbakova, V. (2019). *Desulfovibrio gilichinskyi* sp. nov., a cold-adapted sulfate-reducing bacterium from a Yamal Peninsula cryopeg. *Int. J. Syst. Evol. Microbiol.* 69, 1081–1086. doi: 10.1099/ijsem.0.003272

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

Sakaguchi, T., Arakaki, A., and Matsunaga, T. (2002). *Desulfovibrio magneticus* sp. nov., a novel sulfate-reducing bacterium that produces intracellular single-domain-sized magnetite particles. *Int. J. Syst. Evol. Microbiol.* 52, 215–221. doi: 10.1099/00207713-52-1-215

Sass, H., and Cypionka, H. (2004). Isolation of sulfate-reducing bacteria from the terrestrial deep subsurface and description of *Desulfovibrio cavernae* sp. nov. *Syst. Appl. Microbiol.* 27, 541–548. doi: 10.1078/0723202041748181

Sass, H., Berchtold, M., Branke, J., König, H., Cypionka, H., and Babenzien, H.-D. (1998). Psychrotolerant sulfate-reducing bacteria from an oxic freshwater sediment, description of *Desulfovibrio cuneatus* sp. nov. and *Desulfovibrio*

litoralis sp. nov. *Syst. Appl. Microbiol.* 21, 212–219. doi: 10.1016/s0723-2020(98)80025-8

Sass, H., Ramamoorthy, S., Yarwood, C., Langner, H., Schumann, P., Kroppenstedt, R., et al. (2009). *Desulfovibrio idahonensis* sp. nov., sulfate-reducing bacteria isolated from a metal (loid)-contaminated freshwater sediment. *Int. J. Syst. Evol. Microbiol.* 59, 2208–2214. doi: 10.1099/ijs.0.016709-0

Sattley, W. M., and Madigan, M. T. (2010). Temperature and nutrient induced responses of Lake Fryxell sulfate-reducing prokaryotes and description of *Desulfovibrio lacusfryxellense*, sp. nov., a pervasive, cold-active, sulfate-reducing bacterium from Lake Fryxell, Antarctica. *Extremophiles* 14, 357–366. doi: 10.1007/s00792-010-0315-6

Scholten, J. C., Culley, D. E., Brockman, F. J., Wu, G., and Zhang, W. (2007). Evolution of the syntrophic interaction between *Desulfovibrio vulgaris* and *Methanosarcina barkeri*: involvement of an ancient horizontal gene transfer. *Biochem. Biophys. Res. Commun.* 352, 48–54. doi: 10.1016/j.bbrc.2006.10.164

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Segata, N., Börnigen, 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

Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F., and Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33, 1141–1153. doi: 10.1093/nar/gki242

Shivani, Y., Subhash, Y., Sasikala, C., and Ramana, C. V. (2017). *Halodesulfovibrio spirochaetisodalis* gen. nov. sp. nov. and reclassification of four *Desulfovibrio* spp. *Int. J. Syst. Evol. Microbiol.* 67, 87–93. doi: 10.1099/ijsem.0.001574

Sievert, C. (2020). *Interactive Web-Based Data Visualization With R, Plotly, and Shiny*. Boca Raton, FL: CRC Press.

Spring, S., Sorokin, D. Y., Verbarg, S., Rohde, M., Woyke, T., and Kyrpides, N. C. (2019). Sulfate-reducing bacteria that produce exopolymers thrive in the calcifying zone of a hypersaline cyanobacterial mat. *Front. Microbiol.* 10:862. doi: 10.3389/fmicb.2019.00862

Stothard, P. (2000). The sequence manipulation suite: java script programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102–1104. doi: 10.2144/00286ir01

Stüven, K. (1960). Contribution to the physiology and systematics of sulphate-reducing bacteria. *Arch. Microbiol.* 35, 152–180.

Sun, B., Cole, J. R., Sanford, R. A., and Tiedje, J. M. (2000). Isolation and characterization of *Desulfovibrio dechloracetivorans* sp. nov., a marine dechlorinating bacterium growing by coupling the oxidation of acetate to the reductive dechlorination of 2-chlorophenol. *Appl. Environ. Microbiol.* 66, 2408–2413. doi: 10.1128/aem.66.6.2408-2413.2000

Suzuki, D., Ueki, A., Shizuku, T., Ohtaki, Y., and Ueki, K. (2010). *Desulfovibrio butyratiphilus* sp. nov., a Gram-negative, butyrate-oxidizing, sulfate-reducing bacterium isolated from an anaerobic municipal sewage sludge digester. *Int. J. Syst. Evol. Microbiol.* 60, 595–602. doi: 10.1099/ijs.0.013771-0

Takii, S., Hanada, S., Hase, Y., Tamaki, H., Uyeno, Y., Sekiguchi, Y., et al. (2008). *Desulfovibrio marinisediminis* sp. nov., a novel sulfate-reducing bacterium isolated from coastal marine sediment via enrichment with casamino acids. *Int. J. Syst. Evol. Microbiol.* 58, 2433–2438. doi: 10.1099/00207713-58-11-2673-b

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

Tardy-Jacquenod, C., Magot, M., Laigret, F., Kaghad, M., Patel, B., Guezennec, J., et al. (1996). *Desulfovibrio gabonensis* sp. nov., a new moderately halophilic sulfate-reducing bacterium isolated from an oil pipeline. *Int. J. Syst. Evol. Microbiol.* 46, 710–715. doi: 10.1099/00207713-46-3-710

Thabet, O. B. D., Fardeau, M.-L., Suarez-Nuñez, C., Hamdi, M., Thomas, P., Ollivier, B., et al. (2007). *Desulfovibrio marinus* sp. nov., a moderately halophilic sulfate-reducing bacterium isolated from marine sediments in Tunisia. *Int. J. Syst. Evol. Microbiol.* 57, 2167–2170. doi: 10.1099/ijs.0.64790-0

Thabet, O. B. D., Wafa, T., Eltaief, K., Cayol, J.-L., Hamdi, M., Fauque, G., et al. (2010). *Desulfovibrio legallis* sp. nov.: a moderately halophilic, sulfate-reducing bacterium isolated from a wastewater digestor in Tunisia. *Curr. Microbiol.* 62, 486–491. doi: 10.1007/s00284-010-9733-z

Thiel, J., Spring, S., Tindall, B. J., Spröer, C., Bunk, B., Koeksoy, E., et al. (2020). *Desulfolutivibrio sulfoxidireducens* gen. nov., sp. nov., isolated from a pyrite-forming enrichment culture and reclassification of *Desulfovibrio sulfodismutans* as *Desulfolutivibrio sulfodismutans* comb. nov. *Syst. Appl. Microbiol.* 43:126105. doi: 10.1016/j.syapm.2020.126105

Thioye, A., Gam, Z. B. A., Mbengue, M., Cayol, J.-L., Joseph-Bartoli, M., Touré-Kane, C., et al. (2017). *Desulfovibrio senegalensis* sp. nov., a mesophilic sulfate reducer isolated from marine sediment. *Int. J. Syst. Evol. Microbiol.* 67, 3162–3166. doi: 10.1099/ijsem.0.001997

Trinkerl, M., Breunig, A., Schauder, R., and König, H. (1990). *Desulfovibrio termitidis* sp. nov., a carbohydrate-degrading sulfate-reducing bacterium from the hindgut of a termite. *Syst. Appl. Microbiol.* 13, 372–377. doi: 10.1016/s0723-2020(11)80235-3

Tsu, I.-H., Huang, C.-Y., Garcia, J.-L., Patel, B. K., Cayol, J.-L., Baresi, L., et al. (1998). Isolation and characterization of *Desulfovibrio senezii* sp. nov., a halotolerant sulfate reducer from a solar saltern and phylogenetic confirmation of *Desulfovibrio fructosovorans* as a new species. *Arch. Microbiol.* 170, 313–317. doi: 10.1007/s002030050648

Ueno, A., Tamazawa, S., Tamamura, S., Murakami, T., Kiyama, T., Inomata, H., et al. (2021). *Desulfovibrio subterraneus* sp. nov., a mesophilic sulfate-reducing deltaproteobacterium isolated from a deep siliceous mudstone formation. *Int. J. Syst. Evol. Microbiol.* 71:004683. doi: 10.1099/ijsem.0.004683

Vainshtein, M., Hippe, H., and Kroppenstedt, R. M. (1992). Cellular fatty acid composition of *Desulfovibrio* species and its use in classification of sulfate-reducing bacteria. *Syst. Appl. Microbiol.* 15, 554–566. doi: 10.1016/s0723-2020(11)80115-3

Van Houten, B. H., Meulepas, R. J., Van Doesburg, W., Smidt, H., Muyzer, G., and Stams, A. J. (2009). *Desulfovibrio paquesii* sp. nov., a hydrogenotrophic sulfate-reducing bacterium isolated from a synthesis-gas-fed bioreactor treating zinc-and sulfate-rich wastewater. *Int. J. Syst. Evol. Microbiol.* 59, 229–233. doi: 10.1099/ijs.0.65616-0

Vandieken, V., Knoblauch, C., and Jørgensen, B. B. (2006). *Desulfovibrio frigidus* sp. nov. and *Desulfovibrio ferrireducens* sp. nov., psychrotolerant bacteria isolated from Arctic fjord sediments (Svalbard) with the ability to reduce Fe (III). *Int. J. Syst. Evol. Microbiol.* 56, 681–685. doi: 10.1099/ijs.0.64057-0

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Voordouw, G. (2002). Carbon monoxide cycling by *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* 184, 5903–5911. doi: 10.1128/jb.184.21.5903-5911.2002

Vosjan, J. (1975). Respiration and fermentation of the sulphate-reducing bacterium *Desulfovibrio desulfuricans* in a continuous culture. *Plant Soil* 43, 141–152. doi: 10.1007/bf01928482

Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., et al. (2020). Proposal to reclassify the *proteobacteria*l classes *Deltaproteobacteria* and *Oligoflexia*, and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* 70, 5972–6016. doi: 10.1099/ijsem.0.004213

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

Wambui, J., Cernela, N., Stevens, M. J., and Stephan, R. (2021). Whole Genome Sequence-based identification of *Clostridium estertheticum* complex strains supports the need for taxonomic reclassification within the species *Clostridium estertheticum*. *Front. Microbiol.* 12:727022. doi: 10.3389/fmicb.2021.727022

Wan, Y. Y., Luo, N., Liu, X.-L., Lai, Q.-L., and Goodfellow, M. (2021). *Cupidesulfovibrio liaohensis* gen. nov., sp. nov., a novel sulphate-reducing bacterium isolated from an oil reservoir and reclassification of *Desulfovibrio oxamicus* and *Desulfovibrio termitidis* as *Cupidesulfovibrio oxamicus* comb. nov. and *Cupidesulfovibrio termitidis* comb. nov. *Int. J. Syst. Evol. Microbiol* 71:004618. doi: 10.1099/ijsem.0.004618

Warthmann, R., Vasconcelos, C., Sass, H., and Mckenzie, J. A. (2005). *Desulfovibrio brasiliensis* sp. nov., a moderate halophilic sulfate-reducing bacterium from Lagoa Vermelha (Brazil) mediating dolomite formation. *Extremophiles* 9, 255–261. doi: 10.1007/s00792-005-0441-8

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25. doi: 10.2307/1912004

Widdel, F., and Bak, F. (1992). "Gram-negative mesophilic sulfate-reducing bacteria," in *The Prokaryotes*, eds A. Balows, H. G. Trüper, M. Dworkin, W.

Harder, and K. H. Schleifer (New York, NY: Springer), 3352–3378. doi: 10.1007/978-1-4757-2191-1_21

Wirth, J. S., and Whitman, W. B. (2018). Phylogenomic analyses of a clade within the *Roseobacter* group suggest taxonomic reassignments of species of the genera *Aestuariivita*, *Citreicella*, *Loktanella*, *Nautella*, *Pelagibaca*, *Ruegeria*, *Thalassobius*, *Thiobacimonas* and *Tropicibacter*, and the proposal of six novel genera. *Int. J. Syst. Evol. Microbiol.* 68, 2393–2411. doi: 10.1099/ijsem.0.002833

Yagi, T. (1969). Formate: cytochrome oxidoreductase of *Desulfovibrio vulgaris*. *J. Biochem.* 66, 473–478. doi: 10.1093/oxfordjournals.jbchem.a129171

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

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017a). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67:1613. doi: 10.1099/ijsem.0.001755

Yoon, S.-H., Ha, S.-M., Lim, J., Kwon, S., and Chun, J. (2017b). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Zellner, G., Messner, P., Kneifel, H., and Winter, J. (1989). *Desulfovibrio simplex* sp. nov., a new sulfate-reducing bacterium from a sour whey digester. *Arch. Microbiol.* 152, 329–334. doi: 10.1007/bf00425169

Zhao, C., Gao, Z., Qin, Q., Li, F., and Ruan, L. (2012). *Desulfobaculum xiamenensis* gen. nov., sp. nov., a member of the family *Desulfovibrionaceae* isolated from marine mangrove sediment. *Int. J. Syst. Evol. Microbiol.* 62, 1570–1575. doi: 10.1099/ijs.0.036632-0

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Park, Kim, Park, Yu, Namirimu, Roh and Kwon. 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.

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: frontiersin.org/about/contact

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